A Partially Inexact Proximal Alternating Direction Method of Multipliers and Its Iteration-Complexity Analysis

Vando A. Adona¹ · Max L. N. Gonçalves¹ · Jefferson G. Melo¹

Received: 13 April 2018 / Accepted: 30 March 2019 / Published online: 5 April 2019 © Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
This paper proposes a partially inexact proximal alternating direction method of multipliers for computing approximate solutions of a linearly constrained convex optimization problem. This method allows its first subproblem to be solved inexactly using a relative approximate criterion, whereas a proximal term is added to its second subproblem in order to simplify it. A stepsize parameter is included in the updating rule of the Lagrangian multiplier to improve its computational performance. Pointwise and ergodic iteration-complexity bounds for the proposed method are established. To the best of our knowledge, this is the first time that complexity results for an inexact alternating direction method of multipliers with relative error criteria have been analyzed. Some preliminary numerical experiments are reported to illustrate the advantages of the new method.

Keywords Alternating direction method of multipliers · Relative error criterion · Hybrid extragradient method · Convex program · Pointwise iteration-complexity · Ergodic iteration-complexity

Mathematics Subject Classification 47H05 · 49M27 · 90C25 · 90C60 · 65K10

Communicated by Alfredo N. Iusem.

Max L. N. Gonçalves
maxlng@ufg.br

Vando A. Adona
vandoadona@gmail.com

Jefferson G. Melo
jefferson@ufg.br

¹ IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil
1 Introduction

In this paper, we propose and analyze a partially inexact proximal alternating direction method of multipliers (ADMM) for computing approximate solutions of a linearly constrained convex optimization problem. Recently, there has been some growing interests in the ADMM \cite{1,2}, due to its efficiency for solving the aforementioned class of problems, see, for instance, \cite{3} for a complete review.

Many variants of the ADMM have been studied in the literature. Some of these variants included proximal terms in the subproblems of the ADMM, in order to make them easier to solve or even to have closed-form solutions. Others added a stepsize parameter in the Lagrangian multiplier updating to improve the performance of the method, see, for example, \cite{4–12}, for papers in which one or both of the above two strategies are used. Other works focused on studying inexact versions of the ADMM with different error conditions; for instance, \cite{13–15} analyzed variants, whose subproblems are solved inexactly using relative error criteria. Summable error conditions were also considered in \cite{13,16}; however, it was observed in \cite{13} that, in general, relative error conditions are more interesting from a computational viewpoint. The aforementioned relative error criteria were derived from the one considered in \cite{17} to study an inexact augmented Lagrangian method. The latter work, on the other hand, was motivated by \cite{18,19}, where the authors proposed inexact proximal point-type methods based on relative error criteria.

The contributions of this paper are threefold:

1. to propose an ADMM variant, which combines three of the aforementioned strategies. Namely, (i) the first subproblem of the method is allowed to be solved inexactly in such a way that a relative approximate criterion is satisfied; (ii) a general proximal term is added into the second subproblem; (iii) a stepsize parameter is included in the updating rule of the Lagrangian multiplier;

2. to provide pointwise and ergodic iteration-complexity bounds for the proposed method;

3. to illustrate, by means of numerical experiments, the efficiency of the new method for solving some real-life applications.

Previous most related works Paper \cite{20} was the first one to establish iteration-complexity bounds for the ADMM. Subsequently, \cite{6,7} analyzed ergodic and pointwise iteration-complexities of a proximal ADMM, respectively. In \cite{21}, the authors established iteration-complexity bounds for a variable metric proximal ADMM. Paper \cite{22} analyzed, in Hilbert spaces, the convergence of a variable metric proximal alternating minimization algorithm for solving a linearly constrained optimization problem, whose objective function can be decomposed as a sum of other functions with special structures such as strong convexity and Fréchet differentiability with Lipschitz continuous gradient. We refer the reader to \cite{23–29}, where iteration-complexities of other exact ADMM variants have been considered. Most recently, \cite{13–15} considered some inexact ADMMs, whose subproblems are solved based on absolute and/or relative error criteria. The latter references only considered inexact versions of the standard ADMM, i.e., there is neither proximal term in the subproblems nor stepsize parameter in the Lagrangian multiplier updating rule. Moreover, no iteration-complexity
results were presented. Finally, the complexity analysis of the present paper is based on showing that the proposed method falls within the setting of a hybrid proximal extragradient framework.

**Organization of the paper** Section 2 contains some preliminary results and it is divided into two subsections. The first subsection presents our notation and basic definitions, while the second one recalls a modified HPE framework and its basic iteration-complexity results. Section 3 introduces the partially inexact proximal ADMM and establishes its iteration-complexity bounds. Section 4 is devoted to the numerical experiments. Section 5 contains some conclusions.

**2 Preliminary Results**

This section is divided into two subsections. The first one presents our notation and basic results. The second subsection recalls a modified HPE framework and its iteration-complexity bounds.

**2.1 Notation and Basic Definitions**

This section presents some definitions, notation and basic results used in this paper.

The $p$-norm $(p \geq 1)$ and maximum norm of $z \in \mathbb{R}^n$ are denoted, respectively, by $\|z\|_p = \left( \sum_{i=1}^{n} |z_i|^p \right)^{1/p}$ and $\|z\|_\infty = \max\{|z_1|, \ldots, |z_n|\}$. The index $p$ is omitted when $p = 2$. Let $\mathcal{V}$ be a finite-dimensional real vector space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. For a given self-adjoint positive semidefinite linear operator $Q : \mathcal{V} \to \mathcal{V}$, the seminorm induced by $Q$ on $\mathcal{V}$ is defined by $\|\cdot\|_Q = \langle Q(\cdot), \cdot \rangle^{1/2}$. Since $\langle Q(v), \tilde{v} \rangle$ is symmetric and bilinear, for all $v, \tilde{v} \in \mathcal{V}$, we have

$$2 \langle Qv, \tilde{v} \rangle \leq \|v\|^2_Q + \|\tilde{v}\|^2_Q, \quad \|v + v'\|^2_Q \leq 2 \left( \|v\|^2_Q + \|v'\|^2_Q \right). \quad (1)$$

Given a set-valued operator $T : \mathcal{V} \rightrightarrows \mathcal{V}$, its domain and graph are defined, respectively, as

$$\text{Dom} T = \{v \in \mathcal{V} : T(v) \neq \emptyset\} \quad \text{and} \quad \text{Gr}(T) = \{(v, \tilde{v}) \in \mathcal{V} \times \mathcal{V} \mid \tilde{v} \in T(v)\}.$$  

The operator $T$ is said to be monotone iff

$$\langle u - v, \tilde{u} - \tilde{v} \rangle \geq 0 \quad \forall \,(u, \tilde{u}), \,(v, \tilde{v}) \in \text{Gr}(T).$$

Moreover, $T$ is maximal monotone iff it is monotone and there is no other monotone operator $S$ such that $\text{Gr}(T) \subset \text{Gr}(S)$. Given a scalar $\varepsilon \geq 0$, the $\varepsilon$-enlargement $T[\varepsilon] : \mathcal{V} \rightrightarrows \mathcal{V}$ of the operator $T$ is defined as

$$T[\varepsilon](v) = \{\tilde{v} \in \mathcal{V} : \langle \tilde{v} - \tilde{u}, v - u \rangle \geq -\varepsilon, \,(u, \tilde{u}) \in \text{Gr}(T)\} \quad \forall \, v \in \mathcal{V}. \quad (2)$$
The $\varepsilon$-subdifferential of a proper closed convex function $f : \mathcal{Y} \to [-\infty, \infty]$ is defined by

$$\partial_\varepsilon f(v) = \{ u \in \mathcal{Y} : f(\tilde{v}) \geq f(v) + \langle u, \tilde{v} - v \rangle - \varepsilon, \quad \forall \tilde{v} \in \mathcal{Y} \} \quad \forall v \in \mathcal{Y}.$$ 

When $\varepsilon = 0$, $\partial_0 f(v)$ is denoted by $\partial f(v)$ and is called the subdifferential of $f$ at $v$.

It is well-known that the subdifferential operator of a proper closed convex function is maximal monotone [30].

The next result is a consequence of the transportation formula in [31, Theorem 2.3] combined with [32, Proposition 2(i)].

**Theorem 2.1** Suppose $T : \mathcal{Y} \rightharpoonup \mathcal{Y}$ is maximal monotone and let $\tilde{v}_i, v_i \in \mathcal{Y}$, for $i = 1, \ldots, k$, be such that $v_i \in T(\tilde{v}_i)$. Define

$$\tilde{v}_a^a = \frac{1}{k} \sum_{i=1}^k \tilde{v}_i, \quad v_a^a = \frac{1}{k} \sum_{i=1}^k v_i, \quad \varepsilon_a^a = \frac{1}{k} \sum_{i=1}^k (v_i, \tilde{v}_i - \tilde{v}_a^a).$$

Then, the following hold:

(a) $\varepsilon_a^a \geq 0$ and $v_a^a \in T(v_a^a(\tilde{v}_a^a));$

(b) if, in addition, $T = \partial f$ for a proper closed and convex function $f$, then $v_a^a \in \partial_\varepsilon f(\tilde{v}_a^a)$.

### 2.2 A Modified HPE Framework

Our problem of interest in this section is the monotone inclusion problem

$$0 \in T(z), \quad (3)$$

where $T : \mathcal{Z} \rightharpoonup \mathcal{Z}$ is a maximal monotone operator and $\mathcal{Z}$ is a finite-dimensional real vector space. We assume that the solution set of (3), denoted by $T^{-1}(0)$, is nonempty.

Next, we formally describe a modified HPE framework for computing approximate solutions of (3).

**Modified HPE framework**

Step 0. Let $z_0 \in \mathcal{Z}$, $\eta_0 \in \mathbb{R}_+$, $\sigma \in [0, 1]$ and a self-adjoint positive semidefinite linear operator $M : \mathcal{Z} \to \mathcal{Z}$ be given, and set $k = 1$.

Step 1. Obtain $(z_k, \tilde{z}_k, \eta_k) \in \mathcal{Z} \times \mathcal{Z} \times \mathbb{R}_+$ such that

$$M(z_{k+1} - z_k) \subseteq T(\tilde{z}_k), \quad \| \tilde{z}_k - z_k \|^2_M + \eta_k \leq \sigma \| \tilde{z}_k - z_{k-1} \|^2_M + \eta_{k-1}. \quad (4)$$

Step 2. Set $k \leftarrow k + 1$ and go to step 1.

**Remark 2.1** (i) The modified HPE framework is a generalization of the proximal point method. Indeed, if $M = I$ and $\sigma = \eta_0 = 0$, then (4) implies that $\eta_k = 0$, $z_k = \tilde{z}_k$ and $0 \in z_k - z_{k-1} + T(z_k)$ for every $k \geq 1$, which corresponds to the proximal
point method to solve problem \((3)\). (ii) In Sect. 3, we propose a partially inexact proximal ADMM and show that it falls within the modified HPE framework setting. In particular, it is specified how the triple \((z_k, \tilde{z}_k, \eta_k)\) can be computed in this context. It is worth mentioning that the use of a positive semidefinite operator \(M\) instead of a positive definite one is essential in the analysis of Sect. 3 (see (15)). More examples of algorithms, which can be seen as special cases of HPE-type frameworks, can be found in [18,20,33].

We first present a pointwise iteration-complexity bound for the modified HPE framework derived in [23, Theorem 2.1] (see also [9, Theorem 3.3] for a more general result).

**Theorem 2.2** Let \(\{(z_k, \tilde{z}_k, \eta_k)\}\) be generated by the modified HPE framework. Then, for every \(k \geq 1\), we have \(M(z_{k-1} - z_k) \in T(\tilde{z}_k)\) and there exists \(i \leq k\) such that

\[
\|z_i - z_{i-1}\|_M \leq \frac{1}{\sqrt{k}} \sqrt{\frac{2(1 + \sigma)d_0 + 8\eta_0}{1 - \sigma}},
\]

where \(d_0 = \inf\{\|z^* - z_0\|_M^2 : z^* \in T^{-1}(0)\}\).

**Remark 2.2** For a given tolerance \(\bar{\rho} > 0\), it follows from Theorem 2.2 that in at most \(O(1/\bar{\rho}^2)\) iterations, the modified HPE framework computes an approximate solution \(\tilde{z}\) of \((3)\) and a residual \(r\), in the sense that \(Mr \in T(\tilde{z})\) and \(\|r\|_M \leq \bar{\rho}\). Although \(M\) is assumed to be only positive semidefinite, if \(\|r\|_M = 0\), then \(M^{1/2}r = 0\), which implies that \(Mr = 0\). Hence, the latter inclusion implies that \(\tilde{z}\) is a solution of problem \((3)\). Therefore, the aforementioned concept of approximate solutions makes sense.

We now state an ergodic iteration-complexity bound for the modified HPE framework, whose proof can be derived from [23, Theorem 2.2 with \(\sigma < 1\)] (see also [9, Theorem 3.4] for a more general result).

**Theorem 2.3** Let \(\{(z_k, \tilde{z}_k, \eta_k)\}\) be generated by the modified HPE framework. Consider the ergodic sequence \(\{(\tilde{z}_k^a, r_k^a, \varepsilon_k^a)\}\) defined by

\[
\tilde{z}_k^a = \frac{1}{k} \sum_{i=1}^k \tilde{z}_i, \quad r_k^a = \frac{1}{k} \sum_{i=1}^k (z_{i-1} - z_i), \quad \varepsilon_k^a = \frac{1}{k} \sum_{i=1}^k (M(z_{i-1} - z_i), \tilde{z}_i - \tilde{z}_k^a), \quad \forall k \geq 1.
\]

Then, for every \(k \geq 1\), there hold \(\varepsilon_k^a \geq 0\), \(Mr_k^a \in T(\tilde{z}_k^a)\) and

\[
\|r_k^a\|_M \leq \frac{2\sqrt{d_0 + 2\eta_0}}{k}, \quad \varepsilon_k^a \leq \frac{3(3 - 2\sigma)(d_0 + 2\eta_0)}{2(1 - \sigma)k},
\]

where \(d_0\) is as defined in Theorem 2.2.

**Remark 2.3** For a given tolerance \(\bar{\rho} > 0\), Theorem 2.3 ensures that in at most \(O(1/\bar{\rho})\) iterations of the modified HPE framework, the triple \((\tilde{z}, r, \varepsilon) := (\tilde{z}_k^a, r_k^a, \varepsilon_k^a)\) satisfies
\( M \mathbf{r} \in T^e(\tilde{z}) \) and \( \max\{\|\mathbf{r}\|_M, \varepsilon\} \leq \bar{\rho} \). Similarly to Remark 2.2, the point \( \tilde{z} \) can be interpreted as an approximate solution of (3). Note that the above ergodic-complexity bound is better than the pointwise one by a factor of \( O(1/\bar{\rho}) \); however, the above inclusion is, in general, weaker than that of the pointwise case.

## 3 A Partially Inexact Proximal ADMM and Its Iteration-Complexity Analysis

Consider the following linearly constrained problem

\[
\min \{ f(x) + g(y) : Ax + By = b \},
\]

where \( \mathcal{X} \), \( \mathcal{Y} \) and \( \Gamma \) are finite-dimensional real inner product vector spaces, \( f : \mathcal{X} \to \bar{\mathbb{R}} \) and \( g : \mathcal{Y} \to \bar{\mathbb{R}} \) are proper, closed and convex functions, \( A : \mathcal{X} \to \Gamma \) and \( B : \mathcal{Y} \to \Gamma \) are linear operators, and \( b \in \Gamma \). Although we assume that \( \mathcal{X} \), \( \mathcal{Y} \) and \( \Gamma \) are finite-dimensional spaces, all the results presented here can be naturally extended to infinite-dimensional Hilbert spaces. An important class of problems that can be fit into the above setting is the following composite convex optimization problem

\[
\min \{ f(x) + g(Qx) : x \in \mathcal{X} \},
\]

where \( Q : \mathcal{X} \to \Gamma \) is a linear operator. Indeed, this can be done by considering an artificial variable \( y = Qx \) and setting \( A = -Q \), \( B = I \), and \( b = 0 \). Special instances of (6) include: (i) LASSO \([34,35]\) and \( l_1 \)-regularized logistic regression \([36]\), where \( Q = I \); (ii) least absolute deviations \([3, \text{Sect. 6.1}]\) and total variation denoising \([37]\), where \( Q \) is associated to the least squares fitting model for the former application and the first-order finite difference for the latter.

In this section, we propose a partially inexact proximal ADMM for computing approximate solutions of (5) and establish pointwise and ergodic iteration-complexity bounds for it. Recall that ADMM and its variants are based on alternately solving the augmented Lagrangian subproblems. It is then important to note that the inexact ADMM proposed here is more suitable for applications in which the subproblem associated to the function \( f \) is significantly more challenging to solve than the one involving \( g \).

We begin by formally stating the method.

### Partially Inexact Proximal ADMM

Step 0. Let an initial point \((x_0, y_0, \gamma_0) \in \mathcal{X} \times \mathcal{Y} \times \Gamma\), a penalty parameter \( \beta > 0 \), error tolerance parameters \( \tau_1, \tau_2 \in [0, 1] \), and a self-adjoint positive semidefinite linear operator \( H : \mathcal{Y} \to \mathcal{Y} \) be given. Choose a stepsize parameter

\[
\theta \in \left[ 0, \frac{1 - 2 \tau_1 + \sqrt{(1 - 2 \tau_1)^2 + 4(1 - \tau_1)}}{2(1 - \tau_1)} \right],
\]

and set \( k = 1 \).
Step 1. Compute \((v_k, \tilde{x}_k) \in \mathcal{X} \times \mathcal{X}\) such that

\[
v_k \in \bar{\partial} f(\tilde{x}_k) - A^* \tilde{\gamma}_k, \quad \|\tilde{x}_k - x_{k-1} + \beta v_k\|^2 \\
\leq \tau_1 \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \tau_2 \|\tilde{x}_k - x_{k-1}\|^2,
\]

(8)

where

\[
\tilde{\gamma}_k = \gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b),
\]

(9)

and compute an optimal solution \(y_k \in \mathcal{Y}\) of the subproblem

\[
\min_{y \in \mathcal{Y}} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|A\tilde{x}_k + By - b\|^2 + \frac{1}{2}\|y - y_{k-1}\|_H^2 \right\}.
\]

(10)

Step 2. Set

\[
x_k = x_{k-1} - \beta v_k, \quad \gamma_k = \gamma_{k-1} - \theta \beta (A\tilde{x}_k + By_k - b)
\]

(11)

and \(k \leftarrow k + 1\), and go to step 1.

**Remark 3.1**

(i) If \(\tau_1 = \tau_2 = 0\), then \(\tilde{x}_k = x_k\) due to the inequality in (8) and the first relation in (11). Hence, since \(v_k = (x_{k-1} - x_k)/\beta\), the first subproblem of Step 1 is equivalent to compute an exact solution \(x_k \in \mathcal{X}\) of the following subproblem

\[
\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \gamma_{k-1}, Ax \rangle + \frac{\beta}{2} \|Ax + By_{k-1} - b\|^2 + \frac{1}{2\beta}\|x - x_{k-1}\|_H^2 \right\},
\]

(12)

and then, the partially inexact proximal ADMM becomes the proximal ADMM with stepsize parameter \(\theta \in ]0, (1 + \sqrt{5})/2[\) and proximal terms given by \((1/\beta)I\) and \(H\). Therefore, the proposed method can be seen as an extension of the proximal ADMM in which subproblem (12) is solved inexactly, using a relative approximate criterion. (ii) Subproblem (10) contains a proximal term defined by a self-adjoint positive semidefinite linear operator \(H\), which, appropriately chosen, makes the subproblem easier to solve, or even to have closed-form solution. For instance, if \(H = sI - \beta B^*B\) with \(s > \beta\|B\|^2\), subproblem (10) is equivalent to

\[
\min_{y \in \mathcal{Y}} \left\{ g(y) + \frac{s}{2}\|y - \bar{y}\|^2 \right\},
\]

for some \(\bar{y} \in \mathcal{Y}\), which has a closed-form solution in many applications. For example, if \(g(\cdot) = \|\cdot\|_1\), then to solve the above problem corresponds to evaluating the well-known (explicitly computed) thresholding operator, see (47); we refer the reader to [38,39] for other examples in which the solution of the above proximal subproblem can be explicitly computed. (iii) The use of a relative approximate criterion in (10) requires, as far as we know, the stepsize parameter \(\theta \in ]0, 1[\). However, since, in many applications, the second subproblem (10) is solved exactly and a stepsize parameter
\( \theta > 1 \) accelerates the method, here only the first subproblem is assumed to be solved inexactly. (iv) The partially inexact proximal ADMM is closely related to [13, Algorithm 2]. Indeed, the latter method corresponds to the former one with \( H = 0, \theta = 1 \) and the following condition

\[
2\beta |\langle \tilde{x}_k - x_{k-1}, v_k \rangle| + \beta^2 \| v_k \|^2 \leq \tau_1 \| \tilde{y}_k - y_{k-1} \|^2
\]

(13)

instead of the inequality in (8). Numerical comparisons between the partially inexact proximal ADMM and Algorithm 2 in [13] will be provided in Sect. 4.

In the following, we proceed to provide iteration-complexity bounds for the partially inexact proximal ADMM. Our analysis is done by showing that it is an instance of the modified HPE framework for computing approximate solutions of the monotone inclusion problem

\[
0 \in T(x, y, \gamma) = \begin{bmatrix}
\partial f(x) - A^*\gamma \\
\partial g(y) - B^*\gamma \\
Ax + By - b
\end{bmatrix},
\]

(14)

which corresponds to the Lagrangian system associated to problem (5). As is well-known, \((x^*, y^*)\) is a solution of problem (5) with an associated Lagrange multiplier \(\gamma^*\), iff \((x^*, y^*, \gamma^*)\) is a solution of (14). Throughout this section, we assume that the solution set of (14), denoted by \(\Omega^*\), is nonempty.

Let us now introduce the elements required by the setting of Sect. 2.2. Namely, consider the vector space \(\mathcal{Z} = \mathcal{X} \times \mathcal{Y} \times \Gamma\) and the self-adjoint positive semidefinite linear operator

\[
M = \begin{bmatrix}
I/\beta & 0 & 0 \\
0 & (H + \beta B^*B) & 0 \\
0 & 0 & I/(\theta \beta)
\end{bmatrix}.
\]

(15)

In this setting, the quantity \(d_0\) defined in Theorem 2.2 becomes

\[
d_0 = \inf \left\{ \| (x - x_0, y - y_0, \gamma - \gamma_0) \|_M^2 : (x, y, \gamma) \in T^{-1}(0) \right\}.
\]

(16)

Moreover, the maximal monotonicity of the operator \(T\) defined in (14) follows from the facts that \(\partial f\) and \(\partial g\) are maximal monotone operators and \(T\) can be decomposed as \(T = T_1 + T_2\), where \(T_1 : \mathcal{Z} \rightrightarrows \mathcal{Z}\) is the multi-valued map given by \(T_1(x, y, \gamma) = \partial f(x) \times \partial g(y) \times \{-b\}\) and \(T_2 : \mathcal{Z} \rightarrow \mathcal{Z}\) is the linear operator given by \(T_2(x, y, \gamma) = (-A^*\gamma, -B^*\gamma, Ax + By)\) (note that \(T_2\) is skew-symmetric, i.e., \(\langle T_2 z, \tilde{z} \rangle = -\langle z, T_2 \tilde{z} \rangle\) for all \(z, \tilde{z} \in \mathcal{Z}\)).

We start by presenting a preliminary technical result, which basically shows that a certain sequence generated by the partially inexact proximal ADMM satisfies the inclusion in (4), with \(T\) and \(M\) as above.

**Lemma 3.1** Consider \((x_k, y_k, \gamma_k)\) and \((\tilde{x}_k, \tilde{\gamma}_k)\) generated at the \(k\)-iteration of the partially inexact proximal ADMM. Then,
\[
\frac{1}{\beta} (x_{k-1} - x_k) \in \partial f(\tilde{x}_k) - A^*\tilde{y}_k, \quad (17)
\]

\[
(H + \beta B^* B)(y_{k-1} - y_k) \in \partial g(y_k) - B^*\tilde{y}_k, \quad (18)
\]

\[
\frac{1}{\theta\beta} (y_{k-1} - y_k) = A\tilde{x}_k + By_k - b. \quad (19)
\]

As a consequence, \( z_k = (x_k, y_k, \gamma_k) \) and \( \tilde{z}_k = (\tilde{x}_k, y_k, \tilde{\gamma}_k) \) satisfy inclusion (4) with \( T \) and \( M \) as in (14) and (15), respectively.

**Proof** Inclusion (17) follows trivially from the inclusion in (8) and the first relation in (11). Now, from the optimality condition for (10) and the definition of \( \tilde{\gamma}_k \) in (9), we obtain

\[
0 \in \partial g(y_k) - B^*\gamma_{k-1} + \beta B^*(A\tilde{x}_k + By_k - b) + H(y_k - y_{k-1}) = \partial g(y_k) - B^*[\gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b)] + \beta B^*B(y_k - y_{k-1}) + H(y_k - y_{k-1}),
\]

which proves (18). The relation (19) follows immediately from the second relation in (11). To end the proof, note that the last statement of the lemma follows directly from (17)–(19) and definitions of \( T \) and \( M \) in (14) and (15), respectively.  \( \square \)

The following result presents some relations satisfied by the sequences generated by the partially inexact proximal ADMM. These relations are essential to show that the latter method is an instance of the modified HPE framework.

**Lemma 3.2** Let \( \{(x_k, y_k, \gamma_k)\} \) and \( \{\tilde{(x}_k, \tilde{y}_k)\} \) be generated by the partially inexact proximal ADMM. Then, the following hold:

(a) for any \( k \geq 1 \), we have

\[
\tilde{y}_k - y_{k-1} = \frac{1}{\theta}(y_k - y_{k-1}) + \beta B(y_k - y_{k-1}), \quad \tilde{y}_k - y_k = \frac{1 - \theta}{\theta}(y_k - y_{k-1}) + \beta B(y_k - y_{k-1});
\]

(b) we have

\[
\frac{1}{2} \|y_1 - y_0\|^2_H - \frac{1}{\sqrt{\theta}} \langle B(y_1 - y_0), y_1 - y_0 \rangle \leq 2 \max \left\{ 1, \frac{\theta}{2 - \theta} \right\} d_0,
\]

where \( d_0 \) is as in (16);

(c) for every \( k \geq 2 \), we have

\[
\frac{1}{\theta} \langle y_k - y_{k-1}, B(y_k - y_{k-1}) \rangle \geq \frac{1 - \theta}{\theta} \langle y_{k-1} - y_{k-2}, B(y_k - y_{k-1}) \rangle + \frac{1}{2} \|y_k - y_{k-1}\|^2_H - \frac{1}{2} \|y_{k-1} - y_{k-2}\|^2_H.
\]
Proof (a) The first relation follows by noting that the definitions of \( \tilde{y}_k \) and \( y_k \) in (9) and (11), respectively, yield

\[
\tilde{y}_k - y_{k-1} = -\beta (A\tilde{x}_k + B\tilde{y}_{k-1} - b) = \frac{1}{\theta} (y_k - y_{k-1}) + \beta B (y_k - y_{k-1}).
\]

The second relation in (a) follows trivially from the first one.

(b) First, note that

\[
0 \leq \frac{1}{2\beta} \left\| \frac{1}{\sqrt{\theta}} \left( (y_1 - y_0) + \beta B(y_1 - y_0) \right) \right\|^2 = \frac{1}{2\theta\beta} \| y_1 - y_0 \|^2
\]

\[
+ \frac{1}{\sqrt{\theta}} \langle B(y_1 - y_0), y_1 - y_0 \rangle + \frac{\beta}{2} \| B(y_1 - y_0) \|^2,
\]

which, for every \( z^* = (x^*, y^*, \gamma^*) \in \Omega^* \), yields

\[
\frac{1}{2} \| y_1 - y_0 \|^2_H - \frac{1}{\sqrt{\theta}} \langle B(y_1 - y_0), y_1 - y_0 \rangle
\]

\[
\leq \frac{1}{2} \left( \| y_1 - y_0 \|^2_H + \frac{1}{\theta\beta} \| y_1 - y_0 \|^2_H + \beta \| B(y_1 - y_0) \|^2 \right)
\]

\[
\leq \| y_1 - y^* \|^2_H + \| y_0 - y^* \|^2_H + \frac{1}{\theta\beta} \| y_1 - y^* \|^2
\]

\[
+ \frac{1}{\theta\beta} \| y_0 - y^* \|^2 + \beta \| B(y_1 - y^*) \|^2 + \beta \| B(y_0 - y^*) \|^2,
\]

where the last inequality is due to the second property in (1). Hence, using (15), we obtain

\[
\frac{1}{2} \| y_1 - y_0 \|^2_H - \frac{1}{\sqrt{\theta}} \langle B(y_1 - y_0), y_1 - y_0 \rangle \leq \| z_1 - z^* \|^2_M + \| z_0 - z^* \|^2_M,
\]

where \( z_0 = (x_0, y_0, \gamma_0) \) and \( z_1 = (x_1, y_1, \gamma_1) \). On the other hand, from Lemma 3.1 with \( k = 1 \), we have \( M(z_0 - z_1) \in T(\tilde{z}_1) \), where \( \tilde{z}_1 = (\tilde{x}_1, \gamma_1, \gamma_1) \) and \( T \) is as in (14). Using this fact and the monotonicity of \( T \), we obtain \( \langle x_1 - z^*, M(z_0 - z_1) \rangle \geq 0 \) for all \( z^* = (x^*, y^*, \gamma^*) \in \Omega^* \). Hence,

\[
\| z^* - z_0 \|^2_M - \| z^* - z_1 \|^2_M = \| \tilde{z}_1 - z_0 \|^2_M - \| \tilde{z}_1 - z_1 \|^2_M + 2 \langle \tilde{z}_1 - z^*, M(z_0 - z_1) \rangle
\]

\[
\geq \| \tilde{z}_1 - z_0 \|^2_M - \| \tilde{z}_1 - z_1 \|^2_M.
\]

It follows from (15), item (a), and some direct calculations that

\[
\| \tilde{z}_1 - z_1 \|^2_M = \frac{1}{\beta} \| \tilde{x}_1 - x_1 \|^2 + \frac{1}{\theta\beta} \| \tilde{y}_1 - y_1 \|^2 = \frac{1}{\beta} \| \tilde{x}_1 - x_1 \|^2
\]

\[
+ \frac{1}{\theta\beta} \left\| 1 - \frac{\theta}{\beta} (y_1 - y_0) + \beta B(y_1 - y_0) \right\|^2
\]

\[
\leq \frac{1}{\beta} \| x_1 - x_0 \|^2 + \frac{1}{\theta\beta} \left\| 1 - \frac{\theta}{\beta} (y_1 - y_0) \right\|^2 H^2.
\]
Moreover, (15) and item (a) also yield

\[
\|\tilde{z}_1 - z_0\|_M^2 = \frac{1}{\beta} \|\tilde{x}_1 - x_0\|^2 + \|y_1 - y_0\|_{(\beta B^* + B + H)}^2 + \frac{1}{\theta \beta^3} \|\tilde{y}_1 - y_0\|^2
\]

\[
\geq \frac{1}{\beta} \|\tilde{x}_1 - x_0\|^2 + \beta \|B(y_1 - y_0)\|^2 + \frac{\tau_1}{\beta} \|\tilde{y}_1 - y_0\|^2
\]

\[
+ \frac{1 - \tau_1 \theta}{\theta \beta} \left( \frac{1}{\theta} \|y_1 - y_0\| + \beta \|B(y_1 - y_0)\| \right)^2
\]

\[
= \frac{1}{\beta} \|\tilde{x}_1 - x_0\|^2 + \frac{\tau_1}{\beta} \|\tilde{y}_1 - y_0\|^2 + \frac{[1 + (1 - \tau_1) \theta]}{\theta} \beta \|B(y_1 - y_0)\|^2
\]

\[
+ \frac{1 - \tau_1 \theta}{\theta \beta^3} \|y_1 - y_0\|^2 + \frac{2(1 - \tau_1)}{\theta^2} \langle B(y_1 - y_0), y_1 - y_0 \rangle.
\]  \tag{23}

Combining the above two conclusions, we obtain

\[
\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \geq \frac{1}{\beta} \left( \|\tilde{x}_1 - x_0\|^2 - \|\tilde{x}_1 - x_1\|^2 + \tau_1 \|\tilde{y}_1 - y_0\|^2 \right)
\]

\[
+ (1 - \tau_1) \beta \|B(y_1 - y_0)\|^2 + \frac{2 - \theta - \tau_1}{\beta \theta^2} \|y_1 - y_0\|^2
\]

\[
+ \frac{2(1 - \tau_1)}{\theta} \langle B(y_1 - y_0), y_1 - y_0 \rangle.
\]  \tag{24}

Now, note that the inequality in (8) with \( k = 1 \) and the definition of \( x_1 \) in (11) imply that

\[ 0 \leq \tau_2 \|\tilde{x}_1 - x_0\|^2 - \|\tilde{x}_1 - x_1\|^2 + \tau_1 \|\tilde{y}_1 - y_0\|^2, \]

which, combined with (24) and \( \tau_2 \in [0, 1[ \), yields

\[
\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \geq (1 - \tau_1) \beta \|B(y_1 - y_0)\|^2 + \frac{2 - \theta - \tau_1}{\beta \theta^2} \|y_1 - y_0\|^2
\]

\[
+ \frac{2(1 - \tau_1)}{\theta} \langle B(y_1 - y_0), y_1 - y_0 \rangle = \frac{1 - \theta}{\beta \theta^2} \|y_1 - y_0\|^2
\]

\[
+ (1 - \tau_1) \left\| \sqrt{\beta B(y_1 - y_0)} + \frac{1}{\theta \sqrt{\beta}} (y_1 - y_0) \right\|^2
\]

\[
\geq \frac{1 - \theta}{\beta \theta^2} \|y_1 - y_0\|^2.
\]
Hence, if $\theta \in ]0, 1]$, we have

$$\|\tilde{z}_1 - z_0\|_M^2 - \|\tilde{z}_1 - z_1\|_M^2 \geq 0,$$

which, combined with (21), yields

$$\|z_1 - z^*\|_M^2 \leq \|z_0 - z^*\|_M^2. \tag{25}$$

Now, if $\theta > 1$, we have

$$\|\tilde{z}_1 - z_1\|_M^2 - \|\tilde{z}_1 - z_0\|_M^2 \leq \frac{\theta - 1}{\theta^2} \gamma_1 - \gamma_0^2 \leq \frac{2(\theta - 1)}{\theta} \left( \frac{1}{\beta\theta} \gamma_1 - \gamma^* \|^2 + \frac{1}{\beta\theta} \gamma_0 - \gamma^* \|^2 \right) \leq \frac{2(\theta - 1)}{\theta} \left[ \|z_0 - z^*\|_M^2 + \|z_1 - z^*\|_M^2 \right], \tag{26}$$

where the second inequality is due to the second property in (1), and the last inequality is due to (15) and definitions of $z_0$, $z_1$ and $z^*$. It follows from (7) that $\theta < (1 + \sqrt{5})/2$, in particular, $\theta < 2$. Hence, adding (21) and (26), we obtain

$$\|z_1 - z^*\|_M^2 \leq \frac{3\theta - 2}{2 - \theta} \|z_0 - z^*\|_M^2. \tag{27}$$

Thus, it follows from (25) and the last inequality that

$$\|z_1 - z^*\|_M^2 \leq \max \left\{ 1, \frac{3\theta - 2}{2 - \theta} \right\} \|z_0 - z^*\|_M^2. \tag{27}$$

Therefore, the desired inequality follows from (20), (27) and the definition of $d_0$ in (16).

(c) From the optimality condition for (10), the definition of $\tilde{\gamma}_k$ in (9) and item (a), we have, for every $k \geq 1$,

$$\partial g(y_k) \ni B^*(\tilde{\gamma}_k - \beta B(y_k - y_{k-1})) - H(y_k - y_{k-1}) = \frac{1}{\theta} B^*(y_k - (1 - \theta)y_{k-1}) - H(y_k - y_{k-1}).$$

For any $k \geq 2$, using the above inclusion with $k \leftarrow k$ and $k \leftarrow k - 1$ and the monotonicity of $\partial g$, we obtain

$$\frac{1}{\theta} \langle B^*(y_k - y_{k-1}) - (1 - \theta) B^*(y_{k-1} - y_{k-2}), y_k - y_{k-1} \rangle \geq \langle H(y_k - y_{k-1}), y_k - y_{k-1} \rangle - \langle H(y_{k-1} - y_{k-2}), y_k - y_{k-1} \rangle \geq \frac{1}{2} \|y_k - y_{k-1}\|_H^2 - \frac{1}{2} \|y_{k-1} - y_{k-2}\|_H^2,$$
where the last inequality is due to the first property in (1), and so the proof of the lemma follows.

We next consider a technical result.

**Lemma 3.3** Let scalars $\tau_1, \tau_2$ and $\theta$ be as in step 0 of the partially inexact proximal ADMM. Then, there exists a scalar $\sigma \in [\tau_2, 1]$ such that the matrix

$$
G = \begin{bmatrix}
\sigma - 1 + (\sigma - \tau_1)\theta & -(1 - \theta)[\sigma - 1 + (1 - \tau_1)\theta] \\
-\lvert(1 - \theta)[\sigma - 1 + (1 - \tau_1)\theta]\rvert & \sigma - 1 + (2 - \theta - \tau_1)\theta
\end{bmatrix}
$$

(28)

is positive definite.

**Proof** Since $\tau_1$ and $\theta$ are fixed scalars given in step 0 of the partially inexact proximal ADMM, the determinant and trace of $G$ are polynomial functions of $\sigma$, denoted here by $\Phi(\sigma)$ and $\tilde{\Phi}(\sigma)$, respectively. It is easy to see that

$$
\Phi(1) = \theta^2(1 - \tau_1)\left[-(1 - \tau_1)^2 + (1 - 2\tau_1)\theta + 1\right], \quad \tilde{\Phi}(1) = [3 - 2\tau_1 - \theta]\theta.
$$

Note that the upper bound on $\theta$ given in (7), namely

$$
r := \frac{1 - 2\tau_1 + \sqrt{(1 - 2\tau_1)^2 + 4(1 - \tau_1)}}{2(1 - \tau_1)}
$$

corresponds to the positive root of the quadratic function $q(\theta) = -(1 - \tau_1)\theta^2 + (1 - 2\tau_1)\theta + 1$, which appears in the expression of $\Phi(1)$. Hence, since $\tau_1 \in [0, 1]$ and $\theta \in [0, r]$, we can conclude that $\Phi(1) > 0$. Now, by using $\tau_1 \in [0, 1]$ and some simple algebraic manipulations, it can be verified that $r < 3 - 2\tau_1$, which, combined with the fact that $\theta \in [0, r]$, yields $\tilde{\Phi}(1) > 0$. Therefore, there exists $\hat{\sigma} \in [0, 1]$ such that $\Phi(\sigma) > 0$ and $\tilde{\Phi}(\sigma) > 0$ for all $\sigma \in [\hat{\sigma}, 1]$, which, in turn implies, that $G := G(\sigma)$ is positive definite for all $\sigma \in [\hat{\sigma}, 1]$. The statement of the lemma follows now by choosing $\sigma = \max\{\tau_2, \hat{\sigma}\}$. \qed

In the following, we show that the partially inexact proximal ADMM can be regarded as an instance of the modified HPE framework.

**Proposition 3.1** Let $\{(x_k, y_k, \gamma_k)\}$ and $\{\tilde{(x_k, y_k)}\}$ be generated by the partially inexact proximal ADMM. Let also $T, M$ and $d_0$ be as in (14), (15) and (16), respectively. Define

$$
z_0 = (x_0, y_0, \gamma_0), \quad \mu = \frac{4[\sigma - 1 + (1 - \tau_1)\theta]}{\theta^{3/2}} \max\left\{1, \frac{\theta}{2 - \theta}\right\}, \quad \eta_0 = \mu d_0
$$

(29)
and, for all $k \geq 1$,

$$
\begin{align*}
z_k &= (x_k, y_k, \gamma_k), \quad \tilde{z}_k = (\tilde{x}_k, y_k, \tilde{\gamma}_k), \\
\eta_k &= \left[\frac{\sigma - 1 + (2 - \theta - \tau_1)\theta}{\beta \theta^3}\right] \|y_k - y_{k-1}\|^2 + \left[\frac{\sigma - 1 + (1 - \tau_1)\theta}{\theta}\right] \|y_k - y_{k-1}\|^2_H,
\end{align*}
$$

(30)

where $\sigma \in [\tau_2, 1]$ is given by Lemma 3.3. Then, $(z_k, \tilde{z}_k, \eta_k)$ satisfies the error condition in (4) for every $k \geq 1$. As a consequence, the partially inexact proximal ADMM is an instance of the modified HPE framework.

**Proof** First of all, since $\sigma < 1$ and the matrix $G$ in (28) is positive definite (in particular, $g_{11}$ is positive), we have

$$
[\sigma - 1 + (1 - \tau_1)\theta] \geq [\sigma - 1 + (\sigma - \tau_1)\theta] = g_{11} > 0.
$$

(32)

Now, using (15) and definitions of $\{z_k\}$ and $\{\tilde{z}_k\}$ in (30), we obtain

$$
\begin{align*}
\|\tilde{z}_k - z_{k-1}\|^2_M &= \frac{1}{\beta} \|\tilde{x}_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|^2_H \\
&\quad + \beta \|B(y_k - y_{k-1})\|^2 + \frac{1}{\beta \theta} \|\tilde{y}_k - y_{k-1}\|^2, \\
\|\tilde{z}_k - z_k\|^2_M &= \frac{1}{\beta} \|\tilde{x}_k - x_k\|^2 + \frac{1}{\beta \theta} \|\tilde{y}_k - y_k\|^2.
\end{align*}
$$

Hence,

$$
\begin{align*}
\sigma \|\tilde{z}_k - z_{k-1}\|^2_M - 2\|\tilde{z}_k - z_k\|^2_M &= \frac{1}{\beta} \left(\sigma \|\tilde{x}_k - x_{k-1}\|^2 - \|\tilde{x}_k - x_k\|^2 + \tau_1 \|\tilde{y}_k - y_{k-1}\|^2\right) + \sigma \|y_k - y_{k-1}\|^2_H \\
&\quad + \sigma \beta \|B(y_k - y_{k-1})\|^2 + \frac{\sigma - \tau_1 \theta}{\beta \theta} \|\tilde{y}_k - y_{k-1}\|^2 - \frac{1}{\beta \theta} \|\tilde{y}_k - y_k\|^2.
\end{align*}
$$

(33)

Note that the inequality in (8) and the definition of $x_k$ in (10) imply that

$$
0 \leq \tau_2 \|\tilde{x}_k - x_{k-1}\|^2 - \|\tilde{x}_k - x_k\|^2 + \tau_1 \|\tilde{y}_k - y_{k-1}\|^2,
$$

which, combined with (33) and the fact that $\sigma \geq \tau_2$, yields

$$
\begin{align*}
\sigma \|\tilde{z}_k - z_{k-1}\|^2_M - 2\|\tilde{z}_k - z_k\|^2_M &\geq \sigma \|y_k - y_{k-1}\|^2_H + \sigma \beta \|B(y_k - y_{k-1})\|^2 \\
&\quad + \frac{\sigma - \tau_1 \theta}{\beta \theta} \|\tilde{y}_k - y_{k-1}\|^2 - \frac{1}{\beta \theta} \|\tilde{y}_k - y_k\|^2.
\end{align*}
$$

(34)
On the other hand, it follows from Lemma 3.2(a) that

\[
\frac{\sigma - \tau_1 \theta}{\beta \theta} \|\tilde{y}_k - \gamma_{k-1}\|^2 - \frac{1}{\beta \theta} \|\tilde{y}_k - \gamma_k\|^2 \\
= \frac{\sigma - \tau_1 \theta}{\beta \theta} \left( \frac{1}{\theta} (\gamma_k - \gamma_{k-1}) + \beta B(y_k - y_{k-1}) \right)^2 \\
- \frac{1}{\beta \theta} \left( \frac{1 - \theta}{\theta} (\gamma_k - \gamma_{k-1}) + \beta B(y_k - y_{k-1}) \right)^2 \\
= \frac{\sigma - 1}{\beta \theta^3} (\gamma_k - \gamma_{k-1})^2 + \frac{(\sigma - 1 - \tau_1 \theta) \beta}{\theta} \|B(y_k - y_{k-1})\|^2 \\
+ \frac{2[\sigma - 1 + (1 - \tau_1 \theta)]}{\theta^2} (\gamma_k - \gamma_{k-1}, B(y_k - y_{k-1})).
\]

Hence, combining the last equality and (34), we obtain

\[
\sigma \|\tilde{z}_k - z_{k-1}\|^2_M - \|\tilde{z}_k - z_k\|^2_M \geq \sigma \|y_k - y_{k-1}\|^2_H \\
+ \frac{[\sigma - 1 + (2 - \theta - \tau_1 \theta)]}{\beta \theta^3} \|\gamma_k - \gamma_{k-1}\|^2 \\
+ \frac{[\sigma - 1 + (\sigma - \tau_1 \theta)] \beta}{\theta} \|B(y_k - y_{k-1})\|^2 \\
+ \frac{2[\sigma - 1 + (1 - \tau_1 \theta)]}{\theta^2} (\gamma_k - \gamma_{k-1}, B(y_k - y_{k-1})).
\]

(35)

We will now consider two cases: \( k = 1 \) and \( k > 1 \).

**Case 1** \( k = 1 \) Since \( [\sigma - 1 + (1 - \tau_1 \theta)] > 0 \) (see (32)), it follows from (35) with \( k = 1 \) and Lemma 3.2(b) that

\[
\sigma \|\tilde{z}_1 - z_0\|^2_M - \|\tilde{z}_1 - z_1\|^2_M \geq \frac{[\sigma - 1 + (2 - \theta - \tau_1 \theta)]}{\beta \theta^3} \|\gamma_1 - y_0\|^2 \\
+ \frac{[\sigma - 1 + (\sigma - \tau_1 \theta)] \beta}{\theta} \|B(y_1 - y_0)\|^2 \\
+ \left[ \sigma + \frac{[\sigma - 1 + (1 - \tau_1 \theta)]}{\theta^{3/2}} \right] \|y_1 - y_0\|^2_H \\
- \frac{4[\sigma - 1 + (1 - \tau_1 \theta)]}{\theta^{3/2}} \max \left\{ 1, \frac{\theta}{2 - \theta} \right\} d_0,
\]

which, combined with definitions of \( \eta_0 \) and \( \eta_1 \) in (29) and (31), respectively, yields

\[
\sigma \|\tilde{z}_1 - z_0\|^2_M - \|\tilde{z}_1 - z_1\|^2_M + \eta_0 - \eta_1 \\
\geq \frac{[\sigma - 1 + (\sigma - \tau_1 \theta)] \beta}{\theta} \|B(y_1 - y_0)\|^2 \\
+ \left[ \sigma + \frac{[\sigma - 1 + (1 - \tau_1 \theta)]}{\theta^{3/2}} - \frac{[\sigma - 1 + (1 - \tau_1 \theta)]}{\theta} \right] \|y_1 - y_0\|^2_H.
\]
From the last inequality and some algebraic manipulations, we obtain

\[
\sigma \| \tilde{z}_1 - z_0 \|^2_M - \| \tilde{z}_1 - z_1 \|^2_M + \eta_0 - \eta_1 
\geq \left[ \frac{\sigma - 1 + (\sigma - \tau_1)\theta}{\theta} \right] \left( \beta \| B(y_1 - y_0) \|^2 + \frac{1}{\sqrt{\theta}} \| y_1 - y_0 \|^2_H \right) \\
+ \left[ \frac{\sigma - 1 - \sigma}{\sqrt{\theta}} \right] \| y_1 - y_0 \|^2_H \\
= \left[ \frac{\sigma - 1 + (\sigma - \tau_1)\theta}{\theta} \right] \left( \beta \| B(y_1 - y_0) \|^2 + \frac{1}{\sqrt{\theta}} \| y_1 - y_0 \|^2_H \right) \\
+ \left[ \frac{(1 - \sigma)(1 + \sqrt{\theta} - \theta) + \tau_1\theta}{\theta} \right] \| y_1 - y_0 \|^2_H.
\] (36)

Using (7), we have \( \theta \in [0, (1 + \sqrt{5})/2] \), which, in turn, implies that \( 1 + \sqrt{\theta} - \theta \geq 0 \). Hence, inequality (4) with \( k = 1 \) follows from (32), (36) and the fact that \( \sigma < 1 \).

**Case 2** \((k > 1)\) Since \( \sigma - 1 + (1 - \tau_1)\theta > 0 \) (see (32)), it follows from (35) and Lemma 3.2(c) that

\[
\sigma \| \tilde{z}_k - z_{k-1} \|^2_M - \| \tilde{z}_k - z_k \|^2_M 
\geq \frac{\sigma - 1 + (2 - \theta - \tau_1)\theta}{\beta\theta^3} \| y_k - y_{k-1} \|^2 \\
+ \left[ \frac{\sigma - 1 + (\sigma - \tau_1)\theta}{\theta} \right] \| B(y_k - y_{k-1}) \|^2 + \frac{2(1 - \theta)[\sigma - 1 + (1 - \tau_1)\theta]}{\theta^2} \\
\langle y_{k-1} - y_{k-1}, B(y_k - y_{k-1}) \rangle \\
+ \frac{\sigma - 1 + (1 - \tau_1)\theta}{\theta} \left( \| y_k - y_{k-1} \|^2_H - \| y_{k-1} - y_{k-2} \|^2_H \right),
\]

which, combined with the definition of \( \eta_k \) in (31) and the Cauchy–Schwarz inequality, yields

\[
\sigma \| \tilde{z}_k - z_{k-1} \|^2_M - \| \tilde{z}_k - z_k \|^2_M + \eta_{k-1} - \eta_k 
\geq \frac{\sigma - 1 + (2 - \theta - \tau_1)\theta}{\beta\theta^3} \| y_k - y_{k-1} \|^2 \\
+ \left[ \frac{\sigma - 1 + (\sigma - \tau_1)\theta}{\theta} \right] \| B(y_k - y_{k-1}) \|^2 - \frac{2(1 - \theta)[\sigma - 1 + (1 - \tau_1)\theta]}{\theta^2} \\
\| y_{k-1} - y_{k-2} \| \| B(y_k - y_{k-1}) \| \\
= \frac{1}{\theta} \left( G \left[ \frac{\sqrt{\beta} \| B(y_k - y_{k-1}) \|}{\| y_{k-1} - y_{k-2} \|/\theta \sqrt{\beta}} \right], \left[ \frac{\sqrt{\beta} \| B(y_k - y_{k-1}) \|}{\| y_{k-1} - y_{k-2} \|/\theta \sqrt{\beta}} \right] \right),
\]

where \( G \) is as in (28). Therefore, since \( G \) is positive definite (see Lemma 3.3(b)), we conclude that inequality (4) also holds for \( k > 1 \).

To end the proof, note that the last statement of the proposition follows trivially from the first one and Lemma 3.1.
We are now ready to present our main results of this paper, namely we establish pointwise and ergodic iteration-complexity bounds for the partially inexact proximal ADMM.

**Theorem 3.1** Consider the sequences \(\{(x_k, y_k, \gamma_k)\}\) and \(\{(	ilde{x}_k, \tilde{y}_k)\}\) generated by the partially inexact proximal ADMM. Then, for every \(k \geq 1\),

\[
\begin{pmatrix}
\frac{1}{\beta}(x_{k-1} - x_k) \\
(H + \beta B^* B)(y_{k-1} - y_k) \\
\frac{1}{\beta \theta}(\gamma_{k-1} - \gamma_k)
\end{pmatrix} \in \begin{bmatrix}
\partial f(\tilde{x}_k) - A^*\tilde{y}_k \\
\partial g(y_k) - B^*\tilde{y}_k \\
A\tilde{x}_k + By_k - b
\end{bmatrix}
\]

(37)

and there exist \(\sigma \in ]0, 1[\) and \(i \leq k\) such that

\[
\left(\frac{1}{\beta} \|x_i - x_{i-1}\|^2 + \|y_i - y_{i-1}\|^2 + \frac{1}{\beta \theta} \|\gamma_i - \gamma_{i-1}\|^2\right)^{1/2} \leq \frac{\sqrt{d_0}}{\sqrt{k}} \sqrt{\frac{2(1 + \sigma) + 8\mu}{1 - \sigma}},
\]

where \(d_0\) and \(\mu\) are as in (16) and (29), respectively.

**Proof** This result follows by combining Proposition 3.1 and Theorem 2.2. \(\square\)

**Remark 3.2** For a given tolerance \(\bar{\rho} > 0\), Theorem 3.1 ensures that, in at most \(\Theta(1/\bar{\rho}^2)\) iterations, the partially inexact proximal ADMM provides an approximate solution \(\tilde{z} := (\tilde{x}, \tilde{y})\) of the Lagrangian system (14) together with a residual \(r := (r_x, r_y, r_\gamma)\), in the sense that

\[
\frac{1}{\beta}r_x \in \partial f(\tilde{x}) - A^*\tilde{y}, \quad (H + \beta B^* B)r_y \in \partial g(y) - B^*\tilde{y},
\]

\[
\frac{1}{\beta \theta}r_\gamma = A\tilde{x} + By - b, \quad \|(r_x, r_y, r_\gamma)\|_M \leq \bar{\rho},
\]

where \(M\) is as in (15). Note that the above relations are equivalent to \(Mr \in T(\tilde{z})\) and \(\|r\|_M \leq \bar{\rho}\), with \(T\) as in (14).

**Theorem 3.2** Let the sequences \(\{(x_k^a, y_k^a, \gamma_k^a)\}\) and \(\{(	ilde{x}_k^a, \tilde{y}_k^a)\}\) be generated by the partially inexact proximal ADMM. Consider the ergodic sequences \(\{(x_k^a, y_k^a, \gamma_k^a)\}\), \(\{(	ilde{x}_k^a, \tilde{y}_k^a)\}\), \(\{(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a)\}\) and \(\{e_{k,x}^a, e_{k,y}^a, A_{k,\gamma}^a\}\) defined by

\[
(x_k^a, y_k^a, \gamma_k^a) = \frac{1}{k} \sum_{i=1}^k (x_i, y_i, \gamma_i), \quad (\tilde{x}_k^a, \tilde{y}_k^a) = \frac{1}{k} \sum_{i=1}^k (\tilde{x}_i, \tilde{y}_i),
\]

\[
(r_{k,x}^a, r_{k,y}^a, r_{k,\gamma}^a) = \frac{1}{k} \sum_{i=1}^k (r_{i,x}, r_{i,y}, r_{i,\gamma}), \quad (e_{k,x}^a, e_{k,y}^a, A_{k,\gamma}^a)
\]

(38)
\[
\varepsilon_{k,x,k}^{a} = \varepsilon_{k,y,k}^{a} = \frac{1}{k} \sum_{i=1}^{k} \left( \frac{r_{i,x}}{\beta} + A^{*} \tilde{y}_{i} - \tilde{x}_{i}^{a} \right),
\]
\[
\left( (H + \beta B^{*} B) r_{i,y}^{a} + B^{*} \tilde{y}_{i} - y_{i}^{a} \right),
\]
(39)

where
\[
(r_{i,x}, r_{i,y}, r_{i,y}) = (x_{i-1} - x_{i}, y_{i-1} - y_{i}, y_{i-1} - y_{i}).
\]
(40)

Then, for every \( k \geq 1 \), we have \( \varepsilon_{k,x,k}^{a}, \varepsilon_{k,y,k}^{a} \geq 0 \),
\[
\left( \begin{array}{c}
\frac{1}{\beta} r_{k,x}^{a} \\
(H + \beta B^{*} B) r_{k,y}^{a} \\
\frac{1}{\beta \delta} r_{k,y}^{a}
\end{array} \right) \in \left[ \begin{array}{c}
\partial \varepsilon_{v_{x,k}}^{a} f (\tilde{x}_{k}^{a}) - A^{*} \tilde{y}_{k}^{a} \\
\partial \varepsilon_{v_{y,k}}^{a} g (y_{k}^{a}) - B^{*} \tilde{y}_{k}^{a} \\
A \tilde{x}_{k}^{a} + B y_{k}^{a} - b,
\end{array} \right]
\]
(41)

and there exists \( \sigma \in ]0, 1[ \) such that
\[
\left( \frac{1}{\beta} \| r_{k,x}^{a} \|^{2} + \| r_{k,y}^{a} \|^{2} (H + \beta B^{*} B) + \frac{1}{\beta \delta} \| r_{k,y}^{a} \|^{2} \right)^{1/2} \leq \frac{2 \sqrt{(1 + 2 \mu) d_{0}}}{k}
\]
(42)

and
\[
\varepsilon_{k,x}^{a} + \varepsilon_{k,y}^{a} \leq \frac{3(3 - 2 \sigma)(1 + 2 \mu) d_{0}}{2(1 - \sigma) k},
\]
(43)

where \( d_{0} \) and \( \mu \) are as in (16) and (29), respectively.

**Proof** By combining Proposition 3.1 and Theorem 2.3, we conclude that inequality
(42) holds, and
\[
\varepsilon_{k}^{a} \leq \frac{3(3 - 2 \sigma)(1 + 2 \mu) d_{0}}{2(1 - \sigma) k},
\]
(44)

where
\[
\varepsilon_{k}^{a} = \frac{1}{k} \left( \sum_{i=1}^{k} \left( \frac{r_{i,x}}{\beta} + \tilde{x}_{i} - \tilde{x}_{i}^{a} \right) + \sum_{i=1}^{k} \left( (H + \beta B^{*} B) r_{i,y}^{a} + B^{*} \tilde{y}_{i} - y_{i}^{a} \right) + \sum_{i=1}^{k} \left( r_{i,y}^{a} / (\theta \beta), \tilde{y}_{i} - \tilde{y}_{i}^{a} \right) \right)
\]
(45)

On the other hand, (19), (38) and (40) yield
\[
A \tilde{x}_{k} + B y_{k} = \frac{1}{\theta \beta} r_{k,y}^{a} + b, \quad A \tilde{x}_{k}^{a} + B y_{k}^{a} = \frac{1}{\theta \beta} r_{k,y}^{a} + b.
\]
Additionally, it follows from definitions of \( r_{i,\gamma} \) and \( r^a_{k,\gamma} \) that
\[
\frac{1}{k} \sum_{i=1}^{k} (\tilde{y}_i - r_{i,\gamma} - r^a_{k,\gamma}) = \frac{1}{k} \sum_{i=1}^{k} (\tilde{y}_i - \tilde{y}^a_k, r_{i,\gamma} - r^a_{k,\gamma}) = \frac{1}{k} \sum_{i=1}^{k} (\tilde{y}_i - \tilde{y}^a_k, r_{i,\gamma}).
\]

Hence, combining the identity in (45) with the last two equations, we have
\[
\varepsilon^a_k = \frac{1}{k} \sum_{i=1}^{k} \left( (r_{i,x}/\beta, \tilde{x}_i - \tilde{x}^a_k) + \left( (H + \beta B^* B) r_{i,y}, y_i - y^a_k \right) \right) + \frac{1}{k} \sum_{i=1}^{k} (\tilde{y}_i, (r_{i,\gamma} - r^a_{k,\gamma}) / (\theta \beta))
\]
\[
= \frac{1}{k} \sum_{i=1}^{k} \left( (r_{i,x}/\beta, \tilde{x}_i - \tilde{x}^a_k) + \left( (H + \beta B^* B) r_{i,y}, y_i - y^a_k \right) + \left( \tilde{y}_i, A \tilde{x}_i - A \tilde{x}^a_k + B y_i - B y^a_k \right) \right)
\]
\[
= \frac{1}{k} \sum_{i=1}^{k} (r_{i,x}/\beta + A \tilde{x}_i, \tilde{x}_i - \tilde{x}^a_k) + \frac{1}{k} \sum_{i=1}^{k} \left( (H + \beta B^* B) r_{i,y} + B^* \tilde{y}_i, y_i - y^a_k \right) = \varepsilon^a_{k,x} + \varepsilon^a_{k,y},
\]
where the last equality is due to the definitions of \( \varepsilon^a_{k,x} \) and \( \varepsilon^a_{k,y} \) in (39). Therefore, the inequality in (43) follows trivially from the last equality and (44).

To finish the proof of the theorem, note that direct use of Theorem 2.1(b) (for \( f \) and \( g \), (37)–(40) give \( \varepsilon^a_{k,x}, \varepsilon^a_{k,y} \geq 0 \) and the inclusion in (41). \( \square \)

**Remark 3.3** For a given tolerance \( \tilde{\rho} > 0 \), Theorem 3.2 ensures that, in at most \( O(1/\tilde{\rho}) \) iterations, the partially inexact proximal ADMM provides, in the ergodic sense, an approximate solution \( \tilde{z} := (\tilde{x}^a, y^a, \tilde{y}^a) \) of the Lagrangian system (14), together with residues \( r := (r^a_x, r^a_y, r^a_{\gamma}) \) and \( (\varepsilon^a_x, \varepsilon^a_{\gamma}) \), such that
\[
\frac{1}{\beta} r^a_x \in \partial_{\varepsilon^a_x} f(\tilde{x}^a) - A^* \tilde{y}^a, \quad (H + \beta B^* B)y^a \in \partial_{\varepsilon^a_y} g(y^a) - B^* \tilde{y}^a,
\]
\[
\frac{1}{\beta \theta} r^a_{\gamma} = A \tilde{x}^a + B y^a - b, \quad \| (r^a_x, r^a_y, r^a_{\gamma}) \| M \leq \tilde{\rho},
\]
where \( M \) is as in (15). The above ergodic-complexity bound is better than the pointwise one by a factor of \( O(1/\tilde{\rho}) \); however, the above inclusion is, in general, weaker than that of the pointwise case, due to the \( \varepsilon \)-subdifferentials of \( f \) and \( g \) instead of subdifferentials.

**4 Numerical Experiments**

In this section, we report some numerical experiments to illustrate the performance of the partially inexact proximal ADMM (PIP-ADMM) on two classes of problems, namely LASSO and \( l_1 \)-regularized logistic regression. Our main goal is to show that, in some applications, the method performs better with a stepsize parameter \( \theta > 1 \), instead of the choice \( \theta = 1 \) as considered in the related literature. Similarly to [13,14], we also used a hybrid inner stopping criterion for the PIP-ADMM, i.e., the inner-loop terminates when \( v_k \) satisfies either the inequality in (8) or \( \| v_k \| \leq 10^{-8} \). This strategy
is motivated by the fact that, close to approximate solutions, the former condition seems to be more restrictive than the latter. We set $\tau_1 = 0.99(1 + \theta - \theta^2)/((\theta(2 - \theta))$, $\tau_2 = 1 - 10^{-8}$ and $H = 0$. For a comparison purpose, we also implemented [13, Algorithm 2], denoted here by relerr-ADMM, see Remark 3.1(iv) for more details on the relationship between the PIP-ADMM and the relerr-ADMM. As suggested in [13], the error tolerance parameter $\tau_1$ in (13) was taken equal to $0.99$. For all tests, both algorithms used the initial point $(x_0, y_0, \gamma_0) = (0, 0, 0)$, the penalty parameter $\beta = 1$, and stopped when the following condition was satisfied

$$\|(x_k - x_{k-1}, y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_M \leq 10^{-2},$$

where $M$ is as in (15). The computational results were obtained using MATLAB R2015a on a 2.4 GHz Intel(R) Core i7 computer with 8 GB of RAM.

### 4.1 LASSO Problem

We consider to approximately solve the LASSO problem [34,35]

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Cx - d\|^2 + \delta \|x\|_1,$$

where $C \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$, and $\delta$ is a regularization parameter. We set $\delta = 0.1\|C^*d\|_\infty$. By introducing a new variable, we can rewrite the above problem as

$$\min \left\{ \frac{1}{2} \|Cx - d\|^2 + \delta \|y\|_1 : y - x = 0, \ x \in \mathbb{R}^n, y \in \mathbb{R}^n \right\}. \quad (46)$$

Obviously, (46) is an instance of (5) with $f(x) = (1/2)\|Cx - d\|^2$, $g(y) = \delta \|y\|_1$, $A = -I$, $B = I$ and $b = 0$. Note that, in this case, the pair $(\tilde{x}_k, v_k)$ in (8) can be obtained by computing an approximate solution $\tilde{x}_k$ with a residual $v_k$ of the following linear system

$$(C^*C + \beta I)x = (C^*d + \beta y_{k-1} - \gamma_{k-1}).$$

For approximately solving the above linear system, we used the conjugate gradient method [40], with starting point $C^*d + \beta y_{k-1} - \gamma_{k-1}$. Note also that subproblem (10) has a closed-form solution

$$y_k = \text{shrinkage}_{\delta/\beta}(\tilde{x}_k + \gamma_{k-1}/\beta),$$

where the shrinkage operator is defined as

$$\text{shrinkage}_\kappa : \mathbb{R}^n \to \mathbb{R}^n, \quad (\text{shrinkage}_\kappa(a))_i = \text{sign}(a_i) \max(0, |a_i| - \kappa) \quad i = 1, 2, \ldots, n,$$

with $\text{sign}(\cdot)$ denoting the sign function.
We first tested the methods for solving three randomly generated LASSO problem instances. For a given dimension \( m \times n \), we generated a random matrix \( C \) and scaled its columns to have unit \( l_2 \)-norm. The vector \( d \in \mathbb{R}^m \) was chosen as \( d = Cx + \sqrt{0.001} y \), where the \((100/n)\)-sparse vector \( x \in \mathbb{R}^n \) and the noisy vector \( y \in \mathbb{R}^m \) were also generated randomly.

We also tested the methods on five standard data sets from the Elvira biomedical data set repository [41]. The first data set is the colon tumor gene expression [42] with \( m = 62 \) and \( n = 2000 \), the second is the central nervous system (CNS) data [43] with \( m = 60 \) and \( n = 7129 \), the third is the prostate cancer data [44] with \( m = 102 \) and \( n = 12600 \), the fourth is the Leukemia cancer-ALLMLL data [45] with \( m = 38 \) and \( n = 7129 \), and the fifth is the lung cancer-Michigan data [46] with \( m = 96 \) and \( n = 7129 \). As in the randomly generated problems, we scaled the columns of \( C \) in order to have unit \( l_2 \)-norm.

The performances of the relerr-ADMM and PIP-ADMM are listed in Tables 1 and 2, in which “Out” and “Inner” denote the number of iterations and the total number of inner iterations of the methods, respectively, whereas “Time” is the CPU time in seconds. From these tables, we see that the relerr-ADMM and the PIP-ADMM with \( \theta = 1 \) had similar performances. However, the PIP-ADMM with \( \theta = 1.3 \) and \( \theta = 1.6 \) clearly outperformed the relerr-ADMM.

4.2 \( l_1 \)-Regularized Logistic Regression

Consider the \( l_1 \)-regularized logistic regression problem [36]

\[
\min_{(u,t) \in \mathbb{R}^n \times \mathbb{R}} \left\{ \sum_{i=1}^{m} \log \left( 1 + \exp \left( -d_i \langle (c_i, u) + t \rangle \right) \right) + \delta m \| u \|_1 \right\},
\]

where \((c_i, d_i) \in \mathbb{R}^n \times (-1, +1)\), for every \( i = 1, \ldots, m \), and \( \delta \) is a regularization parameter. We set \( \delta = 0.5\lambda_{\text{max}} \), where \( \lambda_{\text{max}} \) is defined as in [36, Subsect. 2.1]. Note that the above problem can be rewritten as

\[
\min_{(x, u, t) \in \mathbb{R}^{n+1} \times \mathbb{R}^n \times \mathbb{R}} \left\{ \sum_{i=1}^{m} \log \left( 1 + \exp \left( -d_i \langle (1, c_i), x \rangle \right) \right) + \delta m \| u \|_1 : (u, t) - x = 0 \right\},
\]

which is an instance of (5), with \( f(x) = \sum_{i=1}^{m} \log \left( 1 + \exp \left( -d_i \langle (1, c_i), x \rangle \right) \right) \), \( g(y) = g(u, t) = m\delta \| u \|_1 \), \( A = -I \), \( B = I \), and \( b = 0 \). In this case, the pair \((\tilde{x}_k, v_k)\) in (8) was obtained as follows: the iterate \( \tilde{x}_k \) was computed by the Newton method [40] with starting point equal to \((0, \ldots, 0)\), as an approximate solution of the following unconstrained optimization problem

\[
\min_{x \in \mathbb{R}^{n+1}} \left\{ h(x) = \sum_{i=1}^{m} \log \left( 1 + \exp \left( -d_i \langle (1, c_i), x \rangle \right) \right) + \langle x, y_{k-1} \rangle + \frac{\beta}{2} \| y_{k-1} - x \|_2^2 \right\},
\]
Table 1 Performance of the relerr-ADMM and PIP-ADMM to solve 3 randomly generated LASSO problems

| Dim. of $A$ | relerr-ADMM | PIP-ADMM ($\theta = 1$) | PIP-ADMM ($\theta = 1.3$) | PIP-ADMM ($\theta = 1.6$) |
|-------------|-------------|-------------------------|-------------------------|-------------------------|
| $m \times n$ | Out Inner Time | Out Inner Time | Out Inner Time | Out Inner Time |
| 900 $\times$ 3000 | 26 195 11.1 | 26 195 10.2 | 22 169 8.8 | 19 172 7.9 |
| 1200 $\times$ 4000 | 26 193 22.7 | 26 193 20.9 | 21 155 20.9 | 19 169 17.9 |
| 1500 $\times$ 5000 | 25 185 40.9 | 25 185 36.7 | 21 158 34.0 | 18 159 29.3 |
| Data set | relerr-ADMM | PIP-ADMM ($\theta = 1$) | PIP-ADMM ($\theta = 1.3$) | PIP-ADMM ($\theta = 1.6$) |
|----------|-------------|--------------------------|--------------------------|--------------------------|
|          | Out | Inner | Time | Out | Inner | Time | Out | Inner | Time | Out | Inner | Time |
| Colon    | 87  | 1535  | 11.9 | 87  | 1517  | 11.9 | 78  | 1378  | 10.8 | 72  | 1390  | 10.2 |
| CNS      | 204 | 5979  | 466.6| 204 | 5967  | 467.1| 179 | 5293  | 425.7| 164 | 5267  | 383.5|
| Prostate | 368 | 16,176| 3523.5| 366 | 16,030| 3502.6| 298 | 13,212| 2791.2| 252 | 12,319| 2642.4|
| Leukemia | 415 | 7435  | 813.3| 415 | 7435  | 811.6| 347 | 6290  | 674.2| 297 | 5710  | 591.4|
| Lung     | 485 | 10,975| 1008.6| 485 | 10,949| 1023.4| 379 | 8612  | 805.6| 314 | 7736  | 679.1|
Table 3  Performance of the relerr-ADMM and PIP-ADMM on 7 data sets

| Data set  | relerr-ADMM | PIP-ADMM (θ = 1) | PIP-ADMM (θ = 1.3) | PIP-ADMM (θ = 1.6) |
|-----------|-------------|------------------|---------------------|---------------------|
|           | Out | Inner | Time | Out | Inner | Time | Out | Inner | Time | Out | Inner | Time |
| CNS       | 153 | 753   | 6545.3 | 153 | 753   | 6797.9 | 128 | 630   | 6298.5 | 113 | 564   | 5357.8 |
| Colon     | 149 | 596   | 172.2 | 149 | 596   | 180.5 | 125 | 500   | 150.5 | 110 | 464   | 139.0 |
| Leukemia  | 139 | 693   | 6264.4 | 139 | 693   | 6248.8 | 120 | 592   | 5203.9 | 112 | 563   | 4951.9 |
| Lung      | 225 | 1333  | 11,676.9 | 225 | 1333  | 11,354.4 | 219 | 1304  | 10,910.7 | 215 | 1321  | 11,152.5 |
| Ionosphere| 54  | 208   | 0.2 | 54  | 208   | 0.2 | 42  | 162   | 0.2 | 35  | 142   | 0.1 |
| Secom     | 21  | 122   | 15.0 | 21  | 121   | 15.0 | 17  | 97    | 13.5 | 15  | 89    | 12.4 |
| Spambase  | 47  | 212   | 29.7 | 47  | 212   | 29.8 | 37  | 168   | 25.7 | 30  | 147   | 22.4 |
where \( v_k \) was taken as \( v_k = \nabla h(\tilde{x}_k) \). Note that (10) has a closed-form solution \( y_k = (u_k, t_k) \) given by

\[
 u_k = \text{shrinkage}_{m\delta/\beta} \left( \tilde{x}_k^u + y_{k-1}^u/\beta \right), \quad t_k = \tilde{x}_k^t + y_{k-1}^t/\beta,
\]

where \( \tilde{x}_k^u, y_k^u \in \mathbb{R}^n \) and \( \tilde{x}_k^t, y_k^t \in \mathbb{R} \) are the components of the vectors \( \tilde{x}_k \) and \( y_k \), i.e., \( (\tilde{x}_k^u, \tilde{x}_k^t) = \tilde{x}_k \) and \( (y_k^u, y_k^t) = y_k \), and the operator \text{shrinkage} \) is as in (47).

We tested the methods for solving seven \( l_1 \)-regularized logistic regression problem instances. We selected four instances of Sect. 4.1, and three from the ICU Machine Learning Repository [47], namely the ionosphere data [48] with \( m = 351 \) and \( n = 34 \), the secom data with \( m = 1567 \) and \( n = 590 \), and the spambase data with \( m = 4601 \) and \( n = 57 \). We also scaled the columns (resp. rows) of \( C = [c_1, \ldots, c_n]^* \) to have unit \( l_2 \)-norm when \( n \geq m \) (resp. \( m > n \)).

Table 3 reports the performances of the relerr-ADMM and PIP-ADMM for solving the aforementioned seven instances of the problem (48). In Table 3, “Out” and “Inner” are the number of iterations and the total of inner iterations of the methods, respectively, whereas “Time” is the CPU time in seconds. Similarly to the numerical results of Sect. 4.1, we observe that the relerr-ADMM and the PIP-ADMM with \( \theta = 1 \) had similar performances, whereas the PIP-ADMM with \( \theta = 1.3 \) and \( \theta = 1.6 \) outperformed the relerr-ADMM. Therefore, the efficiency of the PIP-ADMM for solving real-life applications is illustrated.

We end this section by making some remarks. First, the PIP-ADMM was tested with other values of \( \theta \), different from the ones presented in Tables 1, 2 and 3, and we observed the following: (i) if \( \theta \in [0.1, 1.6] \), then the performance of the PIP-ADMM improved as \( \theta \) was increased; (ii) if \( \theta \in ]1.6, (\sqrt{5} + 1)/2[ \), then the PIP-ADMM performed similarly to its exact version, since the relative error condition (8) became stringent. Second, the classical proximal gradient method and its accelerated versions, such as FISTA, can also be applied to solve LASSO and \( l_1 \)-regularized logistic regression problems. Numerical comparisons showing that the relerr-ADMM is competitive with FISTA for solving the aforementioned problems were reported in [13]. Therefore, since the PIP-ADMM performed better than the relerr-ADMM for these applications, we can conclude that the PIP-ADMM is also competitive with FISTA.

5 Conclusions

In this paper, we proposed a partially inexact proximal ADMM and established pointwise and ergodic iteration-complexity bounds for it. The proposed method allows its first subproblem to be solved inexactly, using a relative approximate criterion, whereas a stepsize parameter is added in the updating rule of the Lagrangian multiplier, in order to improve its computational performance. We presented some computational results illustrating the numerical advantages of the method.

Acknowledgements The work of these authors was supported in part by CAPES, CNPq Grants 302666/2017-6 and 406975/2016-7. We thank the reviewers for their careful reading and comments.
References

1. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. 2(1), 17–40 (1976)
2. Glowinski, R., Marroco, A.: Sur l’approximation, par éléments finis d’ordre un, et la résolution, par penalisation-dualité, d’une classe de problèmes de Dirichlet non linéaires. R.A.I.R.O. Anal. Numér. 9(2), 41–76 (1975)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2011)
4. Attouch, H., Soueycatt, M.: 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)
5. Xu, M.H.: Proximal alternating directions method for structured variational inequalities. J. Optim. Theory Appl. 134(1), 107–117 (2007)
6. He, B., Yuan, X.: On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J. Numer. Anal. 50(2), 700–709 (2012)
7. He, B., Yuan, X.: On non-ergodic convergence rate of Douglas–Rachford alternating direction method of multipliers. Numer. Math. 130(3), 567–577 (2015)
8. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. 66(3), 889–916 (2016)
9. Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Extending the ergodic convergence rate of the proximal ADMM. arXiv preprint arXiv:1611.02903 (2016)
10. He, B., Liao, L.Z., Han, D., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program. 92(1), 103–118 (2002)
11. Cui, Y., Li, X., Sun, D., Toh, K.C.: On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions. J. Optim. Theory Appl. 169(3), 1013–1041 (2016)
12. Gu, Y., Jiang, B., Deren, H.: A semi-proximal-based strictly contractive Peaceman–Rachford splitting method. arXiv preprint arXiv:1506.02221 (2015)
13. Eckstein, J., Yao, W.: Approximate ADMM algorithms derived from Lagrangian splitting. Comput. Optim. Appl. 68(2), 363–397 (2017)
14. Eckstein, J., Yao, W.: Relative-error approximate versions of Douglas–Rachford splitting and special cases of the ADMM. Math. Program. 170(2), 417–444 (2018)
15. Xie, J., Liao, A., Yang, X.: An inexact alternating direction method of multipliers with relative error criteria. Optim. Lett. 11(3), 583–596 (2017)
16. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55(3, Ser. A), 293–318 (1992)
17. Eckstein, J., Silva, P.J.S.: A practical relative error criterion for augmented Lagrangians. Math. Program. 141(1), 319–348 (2013)
18. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set Valued Anal. 7(4), 323–345 (1999)
19. Solodov, M.V., Svaiter, B.F.: A hybrid projection-proximal point algorithm. J. Convex Anal. 6(1), 59–70 (1999)
20. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim. 23(1), 475–507 (2013)
21. Gonçalves, M.L.N.: On the pointwise iteration-complexity of a dynamic regularized ADMM with over-relaxation stepsize. Appl. Math. Comput. 336, 315–325 (2018)
22. Bitterlich, S., Boţ, R.I., Csetnek, E.R., Wanka, G.: The proximal alternating minimization algorithm for two-block separable convex optimization problems with linear constraints. arXiv preprint arXiv:1806.00260 (2018)
23. Adona, V.A., Gonçalves, M.L.N., Melo, J.G.: Iteration-complexity analysis of a generalized alternating direction method of multipliers. J. Glob. Optim. (2018). https://doi.org/10.1007/s10898-018-0697-z
24. Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-euclidean HPE framework. SIAM J. Optim. 27(1), 379–407 (2017)
25. Gonçalves, M.L.N., Alves, M.M., Melo, J.G.: Pointwise and ergodic convergence rates of a variable metric proximal alternating direction method of multipliers. J. Optim. Theory Appl. 177(2), 448–478 (2018)

26. Boţ, R.I., Csetnek, E.R.: ADMM for monotone operators: convergence analysis and rates. Adv. Comput. Math. (2018). https://doi.org/10.1007/s10444-018-9619-3

27. Hager, W.W., Yashhti, M., Zhang, H.: An $O(1/k)$ convergence rate for the variable stepsize Bregman operator splitting algorithm. SIAM J. Numer. Anal. 54(3), 1535–1556 (2016)

28. Fang, E.X., Bingsheng, H., Liu, H., Xiaoming, Y.: Generalized alternating direction method of multipliers: new theoretical insights and applications. Math. Program. Comput. 7(2), 149–187 (2015)

29. Shefi, R., Teboulle, M.: 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)

30. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pac. J. Math. 33, 209–216 (1970)

31. Burachik, R.S., Sagastizábal, C.A., Svatier, B.F.: $\varepsilon$-Enlargements of maximal monotone operators: theory and applications. In: Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods (Lausanne, 1997). Applied Optimization, vol. 22, pp. 25–43. Kluwer Acad. Publ., Dordrecht (1999)

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

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

34. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58(1), 267–288 (1996)

35. Rockafellar, R.T.: On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. SIAM J. Optim. 20(6), 2755–2787 (2010)

36. Tibshirani, R.J.: The lasso problem and uniqueness. Electron. J. Stat. 7, 1456–1490 (2013)

37. Koh, K., Kim, S.J., Boyd, S.: An interior-point method for large-scale $l_1$-regularized logistic regression. J. Mach. Learn. Res. 8, 1519–1555 (2007)

38. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. 1(3), 127–239 (2014)

39. Beck, A.: First-Order Methods in Optimization. SIAM, Philadelphia (2017)

40. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)

41. Cano, A., Masegosa, A., Moral, S.: ELVIRA biomedical data set repository. http://leo.ugr.es/elvira/DBCRepository/ (2005). Accessed 7 Jan 2018

42. Alon, U., Barkai, N., Noterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA 96(12), 6745–6750 (1999)

43. Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M., McLaughlin, M.E., Kim, J.Y.H., Goumnerova, L.C., Black, P.M., Lau, C., et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286(5439), 531–537 (1999)

44. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286(5439), 531–537 (1999)

45. Beer, D.G., Kardia, S.L.R., Huang, C., Giordano, T.J., Levin, A.M., Misek, D.E., Lin, L., Chen, G., Gharib, T.G., Thomas, D.G., et al.: Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat. Med. 8(8), 816 (2002)

46. Dheeru, D., Taniskidou, E.K.: UCI machine learning repository. http://archive.ics.uci.edu/ml/datasets/madelon (2018). Accessed 7 Jan 2018

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.