Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 14, No. 1, pp. 101–124, 2018

INEXACT GENERALIZED PROXIMAL ALTERNATING DIRECTION METHODS OF MULTIPLIERS AND THEIR CONVERGENCE RATES*

LIMING SUN, ZHIKAI JIANG† AND XINXIN LI

Abstract: The alternating direction method of multiplies (ADMM) has been well studied in the literature; and it has inspired some variants such as inexact versions, generalized versions and proximal versions which are efficient for different circumstances. We propose a general algorithmic framework of ADMM by combining these variants together, in the setting of convex minimization model with linear constraints and a separable objective function. Some ADMM type methods in the literature are subsumed by this general algorithmic framework. More specifically, we allow ADMM's subproblems to be regularized by proximal terms and solved approximately; and then the output is further relaxed by a generalized scheme as suggested by Eckstein and Bertsekas. By choosing different inexactness criteria for the proximal subproblems, two concrete algorithms of the inexact generalized proximal ADMM kind can be derived. We prove the global convergence for these new ADMM type algorithms; and establish their worst-case $O(1/t)$ convergence rates in both ergodic and nonergodic senses. This is a more general and comprehensive work than existing convergence rate results in ADMM literature.

Key words: *convex programming, alternating direction method of multipliers, inexact, convergence rate*

Mathematics Subject Classification: *90C25, 90C30*

---

1 Introduction

We consider the convex minimization problem with linear constraints and a separable objective function

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{1.1}$$

where $A \in \mathcal{R}^{m\times n}$, $B \in \mathcal{R}^{m\times p}$, $b \in \mathcal{R}^m$, $\mathcal{X} \subseteq \mathcal{R}^n$ and $\mathcal{Y} \subseteq \mathcal{R}^p$ are closed and convex sets, $\theta_1 : \mathcal{R}^n \to \mathcal{R}$ and $\theta_2 : \mathcal{R}^p \to \mathcal{R}$ are convex functions. Note both θ_1 and θ_2 could be nonsmooth. Throughout, the solution set of (1.1) denoted by $\mathcal{S}^*$ is assumed to be nonempty.

The Douglas-Rachford alternating direction method of multipliers (ADMM for short) proposed in [16] (see also [3, 13]) is a benchmark solver for the model (1.1), and it becomes very popular recently because of its efficient applications in many areas such as continuum mechanics [18], image processing [4, 5, 29], statistical learning [23, 34], video processing [24],

*This work was supported by the National Natural Science Foundation of China (Grant No. 71401060) and the Natural Science Foundation of Jiangsu Province (Grant No. BK2012662).

†Corresponding author.

and so on; We refer to [2,7,15] for some excellent reviews on the history and applications of ADMM. The iterative scheme of ADMM for solving (1.1) can be written as

$$\begin{cases} x^{k+1} \in \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\|Ax + By^k - b\|_H^2 \;\middle|\; x \in \mathcal{X}\right\}, \\ y^{k+1} \in \arg\min\left\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\|Ax^{k+1} + By - b\|_H^2 \;\middle|\; y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^k - H(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.2}$$

where $\lambda \in \mathcal{R}^m$ is the Lagrange multiplier and $H \in \mathcal{R}^{m\times m}$ is a symmetric positive definite matrix playing the role of a penalty parameter. The most popular choice for H is $H = \beta \cdot I_{m\times m}$ where $\beta > 0$ is a scalar and $I_{m\times m}$ is the identity matrix in $\mathcal{R}^{m\times m}$. The ADMM scheme (1.2) originates from the idea of splitting the subproblem at each iteration of the augmented Lagrangian method in [22,30] in Gauss-Seidel order so as to generate smaller and easier subproblems which could exploit the properties of θ_1 and θ_2 individually.

In [14], it was shown that the ADMM scheme (1.2) can be obtained by applying the Douglas-Rachford splitting method in [25] to the dual of (1.1). Then, in [9], Eckstein and Bertsekas demonstrated that the Douglas-Rachford splitting method is a special form of the proximal point algorithm (PPA) in [26]. They thus followed the relaxed PPA in [17] and proposed the generalized alternating direction method of multipliers (GADMM):

$$\begin{cases} x^{k+1} \in \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \dfrac{1}{2}\|Ax + By^k - b\|_H^2 \;\middle|\; x \in \mathcal{X}\right\}, \\ y^{k+1} \in \arg\min\left\{\theta_2(y) - y^T B^T \lambda^k + \dfrac{1}{2}\|\alpha Ax^{k+1} - (1-\alpha)(By^k - b) + By - b\|_H^2 \;\middle|\; y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^k - H\big[\alpha Ax^{k+1} - (1-\alpha)(By^k - b) + By^{k+1} - b\big], \end{cases} \tag{1.3}$$

where the parameter $\alpha \in (0,2)$ is a relaxation factor. Clearly, the original ADMM (1.2) is the special case of the GADMM (1.3) with $\alpha = 1$. In some articles such as [1,8,10], it has been verified that an over-relaxation factor (i.e., $\alpha \in (1,2)$) can accelerate ADMM's convergence empirically.

Both (1.2) and (1.3) are generic algorithmic frameworks for the abstract model (1.1); how to implement them to solve a concrete application of (1.1) depends on the specific structure/property of the involving objective functions and constraints. For instance, an application of (1.1) is the total variation image denoising model

$$\min\left\{\frac{1}{2}\|x - x^0\|_2^2 + \tau\|\nabla x\|_1 \mid x \in \mathcal{R}^n\right\}, \tag{1.4}$$

where x is the vector representation of a two-dimensional digital image in lexicographical order, x^0 is the observed image corrupted by Gaussian noise, $\nabla \in \mathcal{R}^{n\times n}$ is the matrix representation of the nonsmooth isotropic total variation operator proposed in [32], $\|x\|_1 := \sum_{i=1}^n |x_i|$ and $\tau > 0$ is a trade-off parameter balancing the data fidelity term $\frac{1}{2}\|x - x^0\|_2^2$ and the regularization term $\|\nabla x\|_1$. By introducing an auxiliary variable y, we can reformulate (1.4) as

$$\begin{array}{ll} \min & \dfrac{1}{2}\|x - x^0\|_2^2 + \tau\|y\|_1 \\ s.t. & \nabla x - y = 0. \end{array} \tag{1.5}$$

which is a special case of (1.1) with $\theta_1(x) = \frac{1}{2}\|x - x^0\|_2^2$, $\theta_2(y) = \tau\|y\|_1$, $A = \nabla$, $B = -I_{n\times n}$, $b = 0$, $m = n = p$ and $\mathcal{X} = \mathcal{Y} = \mathcal{R}^n$. Then, because of the simplicity of the functions

The special case where $B = -I_{p\times p}$, $m = p$, $b = 0$ and $H = \beta \cdot I_{m\times m}$ was considered in [9].

in the objective, applying the scheme (1.2) or (1.3) is extremely easy — both the x- and y-subproblems at each iteration are simple enough to have closed-form solutions.

For more complicated functions (θ_1 and θ_2) or coefficient matrices (A and B), the x- and y-subproblems in (1.2) or (1.3) might not have closed-form solutions. In such cases, how to solve these subproblems is crucial for implementing ADMM or GADMM efficiently. One case having widespread applications is that the function (say θ_1) itself is still simple (e.g., $\theta_1(x) = \|x\|_1$) while the corresponding matrix A is not an identity — the x-subproblem in (1.2) or (1.3) thus can only be solved approximately via inner iterations. But obviously, we still want to use the advantageous simplicity of θ_1; an effective and simple strategy towards this purpose is to linearize the quadratic term of the x-subproblem in (1.2) or (1.3) and obtain its approximate problem:

$$\begin{aligned} x^{k+1} = \arg\min \Big\{ & \theta_1(x) + (x - x^k)^T \beta A^T \Big(Ax^k + By^k - b - \frac{1}{\beta}\lambda^k \Big) \\ & + \frac{r}{2}\|x - x^k\|^2 \;\Big|\; x \in \mathcal{X} \Big\}, \end{aligned} \tag{1.6}$$

where the parameter r is required to satisfy $r > \beta\|A^T A\|$ in order to control the accuracy of the linearization. Note for simplicity we take $H = \beta \cdot I_{m\times m}$ in (1.6). Hence, when $\theta_1(x) = \|x\|_1$ and $\mathcal{X} = \mathcal{R}^n$, the closed-form solution of (1.6) is given by the soft-shrinkage operator (e.g. [6]). This ADMM linearization strategy is the essential idea of the well-known split inexact Uzawa method proposed in [36, 37]. Obviously, the approximate x-subproblem (1.6) can be written as

$$x^{k+1} = \arg\min \Big\{ \theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\|Ax + By^k - b\|_H^2 + \frac{1}{2}\|x - x^k\|_G^2 \;\Big|\; x \in \mathcal{X} \Big\}, \tag{1.7}$$

where $G = r \cdot I_{n\times n} - A^T H A$. Therefore, the original ADMM (1.2) and the split inexact Uzawa method in [36, 37] are both special cases of the following scheme

$$\begin{cases} x^{k+1} \in \arg\min \Big\{ \theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\|Ax + By^k - b\|_H^2 + \frac{1}{2}\|x - x^k\|_G^2 \;\Big|\; x \in \mathcal{X} \Big\}, \\ y^{k+1} \in \arg\min \Big\{ \theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\|Ax^{k+1} + By - b\|_H^2 \;\Big|\; y \in \mathcal{Y} \Big\}, \\ \lambda^{k+1} = \lambda^k - H(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.8}$$

where the x-subproblem is regularized by a proximal term and $G \in \mathcal{R}^{n\times n}$ could be an arbitrary symmetric positive semidefinite matrix. Because of its ability to take advantage of the properties of θ_1 effectively, the scheme (1.8) is very efficient for solving a broad spectrum of applications (see, e.g., [23, 31, 33, 35, 37]) and it has received wide attention from various areas. Note in (1.8) we allow the matrix G to be only positive semi-definite and accordingly $\|x\|_G^2 := x^T G x$. This also explains "$\in$" rather than "$=$" in (1.8), since the uniqueness of the x- or y-subproblem in (1.8) is not guaranteed. In fact, the scheme (1.8) is a special case of the proximal ADMM (PADMM) in [19, 37] which suggests to regularize both ADMM's subproblems by proximal terms:

$$\begin{cases} x^{k+1} \in \arg\min \Big\{ \theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\|Ax + By^k - b\|_H^2 + \frac{1}{2}\|x - x^k\|_G^2 \;\Big|\; x \in \mathcal{X} \Big\}, \\ y^{k+1} \in \arg\min \Big\{ \theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\|Ax^{k+1} + By - b\|_H^2 + \frac{1}{2}\|y - y^k\|_S^2 \;\Big|\; y \in \mathcal{Y} \Big\}, \\ \lambda^{k+1} = \lambda^k - H(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.9}$$

where $S \in \mathcal{R}^{p\times p}$ is also an arbitrary symmetric positive semidefinite matrix.

Because of the obvious advantages of the GADMM (1.3) and the PADMM (1.9), we are inspired to consider combining these two ideas for ADMM and thus propose the following generalized proximal ADMM (GPADMM for short):

$$\left\{\begin{array}{l} x^{k+1} \in \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \dfrac{1}{2}\|Ax + By^k - b\|_H^2 + \frac{1}{2}\|x - x^k\|_G^2 \mid x \in \mathcal{X}\right\}, \\ y^{k+1} \in \arg\min\Big\{\theta_2(y) - y^T B^T \lambda^k \\ \qquad\qquad + \dfrac{1}{2}\|\alpha A x^{k+1} - (1-\alpha)(By^k - b) + By - b\|_H^2 + \frac{1}{2}\|y - y^k\|_S^2 \mid y \in \mathcal{Y}\Big\}, \\ \lambda^{k+1} = \lambda^k - H[\alpha A x^{k+1} - (1-\alpha)(By^k - b) + By^{k+1} - b], \end{array}\right. \tag{1.10}$$

with $\alpha \in (0,2)$. Again, the scheme (1.10) is a generic algorithmic framework applicable to the abstract model (1.1). Clearly, ADMM (1.2), GADMM (1.3), PADMM (1.9) and the split inexact Uzawa method (1.6) are all special cases of this scheme. For special cases of (1.1) with simple functions, the GPADMM scheme (1.10) could be used directly as its subproblems might be solved easily. On the other hand, for general cases of (1.1) with generic functions, the x- and y-subproblems in (1.10) must be solved iteratively; we thus can only implement the following inexact version of the GPADMM scheme:

$$\left\{\begin{array}{l} x^{k+1} \approx \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \dfrac{1}{2}\|Ax + By^k - b\|_H^2 + \frac{1}{2}\|x - x^k\|_G^2 \mid x \in \mathcal{X}\right\}, \\ y^{k+1} \approx \arg\min\Big\{\theta_2(y) - y^T B^T \lambda^k \\ \qquad\qquad + \dfrac{1}{2}\|\alpha A x^{k+1} - (1-\alpha)(By^k - b) + By - b\|_H^2 + \dfrac{1}{2}\|y - y^k\|_S^2 \mid y \in \mathcal{Y}\Big\}, \\ \lambda^{k+1} = \lambda^k - H\big[\alpha A x^{k+1} - (1-\alpha)(By^k - b) + By^{k+1} - b\big]. \end{array}\right. \tag{1.11}$$

When some standard inexact criteria in the literature are employed for its subproblems, the scheme (1.11) can be solidified as some implementable algorithms. We refer to [28] and [19] for some inexact versions of the original ADMM (1.2) and the PADMM (1.9), respectively. As we shall show later, the inexact ADMM in [28] can be subsumed by the general inexact GPADMM scheme (1.11).

In this paper, we investigate the convergence for some concrete algorithms derived from the inexact GPADMM scheme (1.11) under different inexactness criteria in a uniform way. In addition to proving the global convergence, we establish the worst-case $O(1/t)$ convergence rate in both ergodic and nonergodic senses for these ADMM type algorithms. Recall that the ergodic and nonergodic worst-case $O(1/t)$ convergence rates of the original ADMM (1.2) and the split inexact Uzawa method (1.6) have been established simultaneously in [20] and [21], respectively; and the same convergence rate of the GADMM (1.3) has been proved in [12]. This work represents a more general and comprehensive analysis than the existing work in [12, 20, 21].

The rest of this paper is organized as follows. In Section 2, we recall some definitions and properties which are useful for further analysis. In Section 3, some preliminary assertions are proved by simple algebra. Then, we propose two concrete algorithms based on the inexact GPADMM scheme (1.11) under different inexactness criteria in Sections 4-5. For each algorithm, we prove its global convergence and establish its worst-case convergence rate in both ergodic and nonergodic senses. Finally, we make some conclusions in Section 6.

2 Preliminaries

In this section, we recall some basic definitions and properties which will be frequently used in our later analysis. Some useful notations are also summarized.

2.1 Variational Reformulation of (1.1)

As the work [20, 21], our analysis requires a variational reformulation of (1.1). Thus we first state it. More specifically, by attaching a Lagrange multiplier $\lambda \in \mathcal{R}^m$ to the linear constraints, solving (1.1) is equivalent to the variational inequality problem: Finding $w^* \in \mathcal{W} := \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^m$ such that

$$\theta(u) - \theta(u^*) + (w - w^*)^T Q(w^*) \geq 0, \qquad \forall w \in \mathcal{W}, \tag{2.1}$$

where

$$u := \begin{pmatrix} x \\ y \end{pmatrix}, \quad w := \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \qquad \theta(u) := \theta_1(x) + \theta_2(y) \quad \text{and} \quad Q(w) := \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}. \tag{2.2}$$

We denote by VI$(\mathcal{W}, Q, \theta)$ the variational inequality problem (2.1)-(2.2). Clearly, the mapping $Q(w)$ defined in (2.2) is affine with a skew-symmetric matrix; it is thus monotone. Furthermore, under our nonempty assumption onto $\mathcal{S}^*$, the solution set of VI$(\mathcal{W}, Q, \theta)$ (denoted by $\mathcal{W}^*$) is also nonempty.

The following theorem originates from [11], and it provides us a very useful characterization on $\mathcal{W}^*$ for establishing worst-case $O(1/t)$ convergence rates for ADMM type algorithms. Since the proof can be found in [11, 20], it is omitted.

Theorem 2.1. *The solution set of VI$(\mathcal{W}, Q, \theta)$ is convex and it can be characterized as*

$$\mathcal{W}^* := \bigcap_{w \in \mathcal{W}} \{\bar{w} \in \mathcal{W} \mid \theta(u) - \theta(\bar{u}) + (w - \bar{w})^T Q(w) \geq 0\}.$$

Based on Theorem 2.1, $\bar{w} \in \mathcal{W}$ can be regarded as an ε-approximation solution of VI$(\mathcal{W}, Q, \theta)$ if it satisfies

$$\sup_{w \in \mathcal{D}} \left\{\theta(\bar{u}) - \theta(u) + (\bar{w} - w)^T Q(w)\right\} \leq \varepsilon,$$

where $\mathcal{D} \subseteq \mathcal{W}$ is some compact set. As Definition 1 in [27], we can take

$$\mathcal{D} = \mathcal{B}_{\mathcal{W}}(\bar{w}) := \left\{w \in \mathcal{W} \mid \|w - \bar{w}\| \leq 1\right\}.$$

In our later analysis, we shall establish ergodic worst-case $O(1/t)$ convergence rates for some algorithms based on the inexact GPADMM (1.11) in the sense that after t iterations of such an algorithm, we can find $\bar{w} \in \mathcal{W}$ such that

$$\theta(\bar{u}) - \theta(u) + (\bar{w} - w)^T Q(w) \leq \varepsilon, \quad \forall w \in \mathcal{B}_{\mathcal{W}}(\bar{w}),$$

with $\varepsilon = O(1/t)$.

2.2 Some Notations

Let Ω be a nonempty closed and convex subset of $\mathcal{R}^l$, $N \in \mathcal{R}^{l \times l}$ be a positive definite matrix, and the N-norm of a vector $v \in \mathcal{R}^l$ be denoted by $\|v\|_N = \sqrt{v^T N v}$.

Let $\alpha \in (0, 2)$, $G \in \mathcal{R}^{n \times n}$, $S \in \mathcal{R}^{p \times p}$ and $H \in \mathcal{R}^{m \times m}$ be positive definite matrices. Throughout we define the matrices P_α and M_α as

$$P_\alpha := \begin{pmatrix} G & 0 & 0 \\ 0 & S & 0 \\ 0 & 0 & (2-\alpha)H^{-1} \end{pmatrix} \quad \text{and} \quad M_\alpha := \begin{pmatrix} G & 0 & 0 \\ 0 & \frac{1}{\alpha}B^T H B + S & \frac{1-\alpha}{\alpha}B^T \\ 0 & \frac{1-\alpha}{\alpha}B & \frac{1}{\alpha}H^{-1} \end{pmatrix}. \tag{2.3}$$

Both P_α and M_α are positive definite under our assumption.

3 Some Preliminary Assertions

In this section, we prove some simple assertions which will be used later when we analyze the convergence for the inexact GPADMM (1.11) under different circumstances. To prove these assertions, only preliminary algebra is needed. Although we attach the superscript k to the letters, the proof of the following lemma is irrelevant to the specific scheme (1.11).

Lemma 3.1. *Let $\alpha \in (0,2)$, $\bar{x}^k \in \mathcal{R}^n$, $y^k, \bar{y}^k \in \mathcal{R}^p$ and $\lambda^k \in \mathcal{R}^m$. If we define $\bar{\lambda}^k$ and $\hat{\lambda}^{k+1}$ as*

$$\bar{\lambda}^k := \lambda^k - H(A\bar{x}^k + By^k - b) \quad \textit{and} \quad \hat{\lambda}^{k+1} := \lambda^k - H\left\{[\alpha A\bar{x}^k - (1-\alpha)(By^k - b)] + B\bar{y}^k - b\right\},$$

respectively; then for any $y \in \mathcal{R}^p$ and $\lambda \in \mathcal{R}^m$, we have

$$\begin{aligned}
&2(\lambda - \bar{\lambda}^k)^T(A\bar{x}^k + B\bar{y}^k - b) + \frac{2}{\alpha}(y - \bar{y}^k)^T B^T H B(y^k - \bar{y}^k) \\
&= \frac{1}{\alpha}(\|y - \bar{y}^k\|^2_{B^T HB} - \|y - y^k\|^2_{B^T HB}) + \frac{1}{\alpha}(\|\lambda - \hat{\lambda}^{k+1}\|^2_{H^{-1}} - \|\lambda - \lambda^k\|^2_{H^{-1}}) \\
&\quad + (2-\alpha)\|\lambda^k - \bar{\lambda}^k\|^2_{H^{-1}} + \frac{2(1-\alpha)}{\alpha}(\lambda - \lambda^k)^T(By^k - B\bar{y}^k).
\end{aligned}$$

Proof. Using the definition of $\hat{\lambda}^{k+1}$, we have

$$\begin{aligned}
\hat{\lambda}^{k+1} &= \lambda^k - H\left\{[\alpha A\bar{x}^k - (1-\alpha)(By^k - b)] + B\bar{y}^k - b\right\} \\
&= \lambda^k - H\left[\alpha(A\bar{x}^k + B\bar{y}^k - b) + (\alpha - 1)(By^k - B\bar{y}^k)\right],
\end{aligned}$$

from which we get

$$A\bar{x}^k + B\bar{y}^k - b = \frac{1}{\alpha}H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) + \frac{1-\alpha}{\alpha}(By^k - B\bar{y}^k).$$

Then, then we obtain

$$\begin{aligned}
&2(\lambda - \bar{\lambda}^k)^T(A\bar{x}^k + B\bar{y}^k - b) \\
&= \frac{2}{\alpha}(\lambda - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) + \frac{2(1-\alpha)}{\alpha}(\lambda - \bar{\lambda}^k)^T(By^k - B\bar{y}^k) \\
&= \frac{2}{\alpha}(\lambda - \hat{\lambda}^{k+1})^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) + \frac{2}{\alpha}(\hat{\lambda}^{k+1} - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) \\
&\quad + \frac{2(1-\alpha)}{\alpha}(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k) + \frac{2(1-\alpha)}{\alpha}(\lambda - \lambda^k)^T(By^k - B\bar{y}^k). \qquad (3.1)
\end{aligned}$$

For the first item in (3.1), we have the following identity

$$\frac{2}{\alpha}(\lambda - \hat{\lambda}^{k+1})^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) = \frac{1}{\alpha}\|\lambda - \hat{\lambda}^{k+1}\|^2_{H^{-1}} - \frac{1}{\alpha}\|\lambda - \lambda^k\|^2_{H^{-1}} + \frac{1}{\alpha}\|\lambda^k - \hat{\lambda}^{k+1}\|^2_{H^{-1}}. \qquad (3.2)$$

Then, it follows from the definitions of $\hat{\lambda}^{k+1}$ and $\bar{\lambda}^k$ that

$$\hat{\lambda}^{k+1} = \lambda^k - H\left[\alpha(A\bar{x}^k + By^k - b) + (B\bar{y}^k - By^k)\right] = \lambda^k - \alpha(\lambda^k - \bar{\lambda}^k) + H(By^k - B\bar{y}^k).$$

Therefore, we obtain

$$\hat{\lambda}^{k+1} - \bar{\lambda}^k = (1-\alpha)(\lambda^k - \bar{\lambda}^k) + H(By^k - B\bar{y}^k) \qquad (3.3)$$

and

$$\lambda^k - \hat{\lambda}^{k+1} = \alpha(\lambda^k - \bar{\lambda}^k) + H(B\bar{y}^k - By^k). \tag{3.4}$$

Using the above equations and by simple manipulations, we get

$$\begin{aligned}
&\frac{2}{\alpha}(\hat{\lambda}^{k+1} - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) + \frac{2(1-\alpha)}{\alpha}(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k) \\
&= \frac{2}{\alpha}[(1-\alpha)(\lambda^k - \bar{\lambda}^k) + H(By^k - B\bar{y}^k)]^T H^{-1}[\alpha(\lambda^k - \bar{\lambda}^k) + H(B\bar{y}^k - By^k)] \\
&\quad + \frac{2(1-\alpha)}{\alpha}(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k) \\
&= 2(1-\alpha)\|\lambda^k - \bar{\lambda}^k\|^2_{H^{-1}} - \frac{2}{\alpha}\|y^k - \bar{y}^k\|^2_{B^THB} + 2(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k). \quad (3.5)
\end{aligned}$$

It follows from (3.4) that

$$\|\lambda^k - \hat{\lambda}^{k+1}\|^2_{H^{-1}} = \alpha^2\|\lambda^k - \bar{\lambda}^k\|^2_{H^{-1}} + \|y^k - \bar{y}^k\|^2_{B^THB} - 2\alpha(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k),$$

and thus

$$2(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k) = \alpha\|\lambda^k - \bar{\lambda}^k\|^2_{H^{-1}} + \frac{1}{\alpha}\|y^k - \bar{y}^k\|^2_{B^THB} - \frac{1}{\alpha}\|\lambda^k - \hat{\lambda}^{k+1}\|^2_{H^{-1}}.$$

From the above equality and (3.5), we obtain

$$\begin{aligned}
&\frac{2}{\alpha}(\hat{\lambda}^{k+1} - \bar{\lambda}^k)^T H^{-1}(\lambda^k - \hat{\lambda}^{k+1}) + \frac{2(1-\alpha)}{\alpha}(\lambda^k - \bar{\lambda}^k)^T(By^k - B\bar{y}^k) \\
&\quad = (2-\alpha)\|\lambda^k - \bar{\lambda}^k\|^2_{H^{-1}} - \frac{1}{\alpha}\|y^k - \bar{y}^k\|^2_{B^THB} - \frac{1}{\alpha}\|\lambda^k - \hat{\lambda}^{k+1}\|^2_{H^{-1}}.
\end{aligned}$$

Substituting this and (3.2) into (3.1), we have

$$\begin{aligned}
&2(\lambda - \bar{\lambda}^k)^T(A\bar{x}^k + B\bar{y}^k - b) \\
&= \frac{1}{\alpha}(\|\lambda - \hat{\lambda}^{k+1}\|^2_{H^{-1}} - \|\lambda - \lambda^k\|^2_{H^{-1}}) + (2-\alpha)\|\lambda^k - \bar{\lambda}^k\|^2_{H^{-1}} - \frac{1}{\alpha}\|y^k - \bar{y}^k\|^2_{B^THB} \\
&\quad + \frac{2(1-\alpha)}{\alpha}(\lambda - \lambda^k)^T(By^k - B\bar{y}^k).
\end{aligned}$$

The following is an identity

$$\frac{2}{\alpha}(y - \bar{y}^k)^T B^T HB(y^k - \bar{y}^k) = \frac{1}{\alpha}\|y - \bar{y}^k\|^2_{B^THB} - \frac{1}{\alpha}\|y - y^k\|^2_{B^THB} + \frac{1}{\alpha}\|y^k - \bar{y}^k\|^2_{B^THB}.$$

Adding the above two equalities, the assertion is proved. □ □

Next, we show an assertion which will be used for proving the convergence of the inexact GPADMM (1.11) to be proposed.

Theorem 3.2. *Let c_0 and $\alpha \in (0,2)$ be positive constants; $\{\zeta_k\}$ and $\{\eta_k\}$ be nonnegative sequences with $\sum_{k=0}^{\infty}\zeta_k < +\infty$ and $\sum_{k=0}^{\infty}\eta_k < +\infty$; and P_α and M_α be defined in (2.3). For any $w^* \in \mathcal{W}^*$, if there are two sequences $\{w^k\}$ and $\{\bar{w}^k\}$ satisfying*

$$\|w^{k+1} - w^*\|^2_{M_\alpha} \le (1+\zeta_k)\|w^k - w^*\|^2_{M_\alpha} + \eta_k - c_0\|w^k - \bar{w}^k\|^2_{P_\alpha}, \quad \forall k \ge 0, \tag{3.6}$$

then $\{w^k\}$ is bounded and

$$\lim_{k\to\infty}\|w^k - \bar{w}^k\|_{P_\alpha} = 0. \tag{3.7}$$

Furthermore, if the mapping Q is continuous and

$$\liminf_{k\to\infty} \left\{\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k)\right\} \geq 0, \qquad \forall w \in \mathcal{W}, \tag{3.8}$$

then the sequence $\{w^k\}$ converges to a point in $\mathcal{W}^*$.

Proof. From $\sum_{k=0}^{\infty} \zeta_k < +\infty$ and $\zeta_k \geq 0$, it follows that $\prod_{k=0}^{\infty}(1+\zeta_k) < +\infty$. We denote

$$C_s := \sum_{k=0}^{\infty} \zeta_k, \qquad C_p := \prod_{k=0}^{\infty}(1+\zeta_k) \qquad \text{and} \qquad C_\eta := \sum_{k=0}^{\infty} \eta_k.$$

Let $w^* \in \mathcal{W}^*$. From (3.6) we get

$$\begin{aligned}
\|w^{k+1} - w^*\|_{M_\alpha}^2 &\leq (1+\zeta_k)\|w^k - w^*\|_{M_\alpha}^2 + \eta_k \\
&\leq (1+\zeta_k)\big[(1+\zeta_{k-1})\|w^{k-1} - w^*\|_{M_\alpha}^2 + \eta_{k-1}\big] + \eta_k \\
&\leq (1+\zeta_k)(1+\zeta_{k-1})\big(\|w^{k-1} - w^*\|_{M_\alpha}^2 + \eta_{k-1} + \eta_k\big).
\end{aligned}$$

Thus for any $l \leq k$, we have

$$\begin{aligned}
\|w^{k+1} - w^*\|_{M_\alpha}^2 &\leq \prod_{i=l}^{k}(1+\zeta_i)\Big(\|w^l - w^*\|_{M_\alpha}^2 + \sum_{i=l}^{k}\eta_i\Big) \\
&\leq C_p\Big(\|w^l - w^*\|_{M_\alpha}^2 + \sum_{i=l}^{\infty}\eta_i\Big) \\
&\leq C_p\|w^l - w^*\|_{M_\alpha}^2 + C_p C_\eta.
\end{aligned} \tag{3.9}$$

Therefore, there exists a constant $C > 0$ such that

$$\|w^k - w^*\|_{M_\alpha}^2 \leq C, \qquad \forall k \geq 0. \tag{3.10}$$

Then, the sequence $\{w^k\}$ is bounded. Combining (3.6) and (3.10), we have

$$\begin{aligned}
c_0 \sum_{k=0}^{\infty} \|w^k - \bar{w}^k\|_{P_\alpha}^2 &\leq \|w^0 - w^*\|_{M_\alpha}^2 + \sum_{k=0}^{\infty} \zeta_k \|w^k - w^*\|_{M_\alpha}^2 + \sum_{k=0}^{\infty} \eta_k \\
&\leq C + C\sum_{k=0}^{\infty}\zeta_k + C_\eta \\
&\leq (1+C_s)C + C_\eta.
\end{aligned}$$

It follows that

$$\lim_{k\to\infty} \|w^k - \bar{w}^k\|_{P_\alpha} = 0.$$

Thus the first assertion (3.7) is proved.

Since $\{w^k\}$ is bounded and $\lim_{k\to\infty}\|w^k - \bar{w}^k\|_{P_\alpha} = 0$, we have that $\{\bar{w}^k\}$ is also bounded and then it has at least one cluster point. Let w^∞ be a cluster point of $\{\bar{w}^k\}$ and the subsequences $\{\bar{w}^{k_j}\}$ and $\{w^{k_j}\}$ both converge to w^∞. It follows from (3.8) that

$$\liminf_{j\to\infty} \left\{\theta(u) - \theta(\bar{u}^{k_j}) + (w - \bar{w}^{k_j})^T Q(\bar{w}^{k_j})\right\} \geq 0, \qquad \forall w \in \mathcal{W},$$

and consequently

$$\theta(u) - \theta(u^\infty) + (w - w^\infty)^T Q(w^\infty) \geq 0, \qquad \forall w \in \mathcal{W}.$$

This means that w^∞ is a solution of VI$(\mathcal{W}, Q, \theta)$. Note that inequality (3.9) is true for all solution points of VI$(\mathcal{W}, Q, \theta)$, hence we have

$$\|w^{k+1} - w^\infty\|_{M_\alpha}^2 \leq C_p(\|w^l - w^\infty\|_{M_\alpha}^2 + \sum_{i=l}^{\infty} \eta_i), \quad \forall k \geq 0, \forall l \leq k. \tag{3.11}$$

Since $w^{k_j} \to w^\infty$ $(j \to \infty)$ and $\sum_{i=0}^{\infty} \eta_i < +\infty$, for any given $\varepsilon > 0$, there exists a $j_0 > 0$ such that

$$\|w^{k_{j_0}} - w^\infty\|_{M_\alpha}^2 \leq \frac{\varepsilon^2}{2C_p} \quad \text{and} \quad \sum_{i=k_{j_0}}^{\infty} \eta_i \leq \frac{\varepsilon^2}{2C_p}. \tag{3.12}$$

Therefore, for any $k \geq k_{j_0}$, it follows from (3.11) and (3.12) that

$$\|w^{k+1} - w^\infty\|_{M_\alpha} \leq \sqrt{C_p(\|w^{k_{j_0}} - w^\infty\|_{M_\alpha}^2 + \sum_{i=k_{j_0}}^{\infty} \eta_i)} \leq \varepsilon.$$

This implies that the sequence $\{w^k\}$ converges to a point w^∞ in $\mathcal{W}^*$. □ □

4 An Implementable Inexact GPADMM with Absolute Error Control

In the following sections, we embed the inexactness criteria proposed in [28] into the inexact version of GPADMM (1.11) and propose some implementable algorithms based on the inexact GPADMM (1.11). The algorithmic framework of the new algorithms with the criteria in [28] can be described as follows: Find $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \mathcal{W}$, $\xi_x^k \in \mathcal{R}^n$, and $\xi_y^k \in \mathcal{R}^p$ such that for any $w = (x, y) \in \mathcal{X} \times \mathcal{Y}$ we have

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{-A^T[\lambda^k - H(Ax^{k+1} + By^k - b)] + G(x^{k+1} - x^k) + \xi_x^k\} \geq 0, \tag{4.1}$$

$$\begin{aligned} \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T[\lambda^k - H(\alpha A x^{k+1} \\ - (1-\alpha)(By^k - b) + By^{k+1} - b)] + S(y^{k+1} - y^k) + \xi_y^k\} \geq 0, \end{aligned} \tag{4.2}$$

$$\lambda^{k+1} := \lambda^k - H[\alpha A x^{k+1} - (1-\alpha)(By^k - b) + By^{k+1} - b]. \tag{4.3}$$

In (4.1)-(4.3), the terms ξ_x^k and ξ_y^k are error terms for solving the respective x- and y-subproblems; and their specific choices are referred to [28]. In the implementation, we can control either the absolute or the relative errors of these error terms, and it leads to two concrete algorithms under the framework (4.1)-(4.3). This section focuses on the case where the absolute error of those error terms are controlled.

4.1 Algorithm

Algorithm 1: An implementable inexact GPADMM with absolute error control

Step 0. *Let $\varepsilon > 0$; $\alpha \in (0, 2)$; $w^0 := (x^0, y^0, \lambda^0) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^m$, G, S and H be positive definite matrices and $\{\nu_k\}$ be a nonnegative sequence satisfying $\sum_{k=0}^{\infty} \nu_k < +\infty$. Set $k := 0$.*

Step 1. *Find* $x^{k+1} \in \mathcal{X}$ *and* $\xi_x^k \in \mathcal{R}^n$ *such that*

$$\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\Big\{-A^T[\lambda^k-H(Ax^{k+1}+By^k-b)]+G(x^{k+1}-x^k)+\xi_x^k\Big\}\geq 0, \quad \forall x\in\mathcal{X}, \tag{4.4}$$

where ξ_x^k *satisfies the inexactness criterion* $\|\xi_x^k\|_{G^{-1}} \leq \nu_k$.
Step 2. *Find* $y^{k+1} \in \mathcal{Y}$ *and* $\xi_y^k \in \mathcal{R}^p$ *such that*

$$\begin{aligned}\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\Big\{&-B^T\big[\lambda^k-H\big(\alpha Ax^{k+1}-(1-\alpha)(By^k-b)+By^{k+1}-b\big)\big]\\ &+S(y^{k+1}-y^k)+\xi_y^k\Big\}\geq 0, \qquad \forall y\in\mathcal{Y},\end{aligned} \tag{4.5}$$

where ξ_y^k *satisfies the inexactness criterion* $\|\xi_y^k\|_{S^{-1}} \leq \nu_k$.
Step 3. *Update* λ^{k+1} *via*

$$\lambda^{k+1} := \lambda^k - H\big[\alpha Ax^{k+1}-(1-\alpha)(By^k-b)+By^{k+1}-b\big]. \tag{4.6}$$

Step 4. *Set* $w^{k+1} := (x^{k+1}, y^{k+1}, \lambda^{k+1})$. *If* $\|w^{k+1}-w^k\| \leq \varepsilon$, *stop; otherwise set* $k = k+1$ *and goto Step 1.*

Remark 4.1. If $\alpha = 1$, Algorithm 1 is just the inexact alternating direction method with Criterion 1 in [28].

4.2 Convergence

In this subsection, we prove the convergence of Algorithm 1. First we define

$$\bar{w}^k := \begin{pmatrix} \bar{x}^k \\ \bar{y}^k \\ \bar{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - H(Ax^{k+1}+By^k-b) \end{pmatrix} \tag{4.7}$$

to simplify our notation in the following analysis.

We now prove a useful lemma which will be used in Sections 4 and 5.

Lemma 4.2. *For given* w^k, *let* w^{k+1} *be generated by* (4.1)-(4.3), *and* $\bar{w}^k$ *be defined by* (4.7). *Then for any* $w = (x, y, \lambda) \in \mathcal{W}$, *we have*

$$\begin{aligned}&\theta(u)-\theta(\bar{u}^k)+(w-\bar{w}^k)^TQ(\bar{w}^k)+(x-\bar{x}^k)^T\xi_x^k+(y-\bar{y}^k)^T\xi_y^k\\ &\geq \frac{1}{2}(\|w^{k+1}-w\|_{M_\alpha}^2-\|w^k-w\|_{M_\alpha}^2)+\frac{1}{2}\|w^k-\bar{w}^k\|_{P_\alpha}^2,\end{aligned} \tag{4.8}$$

where P_α *and* M_α *are defined by* (2.3).

Proof. With the notation $\bar{w}^k$ given in (4.7), the VI (4.4) can be written as

$$\begin{aligned}&\theta_1(x)-\theta_1(\bar{x}^k)+(x-\bar{x}^k)^T(-A^T\bar{\lambda}^k)+(x-\bar{x}^k)^T\xi_x^k\\ &\geq (x^{k+1}-x)^TG(x^{k+1}-x^k)\\ &=\frac{1}{2}(\|x^{k+1}-x\|_G^2-\|x^k-x\|_G^2)+\frac{1}{2}\|x^k-\bar{x}^k\|_G^2, \quad \forall x\in\mathcal{X}.\end{aligned} \tag{4.9}$$

Analogously, from (4.5) and (4.6) we get

$$\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T[-B^T\lambda^{k+1}+S(y^{k+1}-y^k)]+(y-\bar{y}^k)^T\xi_y^k \geq 0, \qquad \forall y \in \mathcal{Y}. \quad (4.10)$$

Using the notation $\bar{w}^k$ given in (4.7), λ^{k+1} in (4.6) can be written as

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - H[\alpha(Ax^{k+1}+By^k-b)+B(y^{k+1}-y^k)] \\ &= \lambda^k - \alpha(\lambda^k-\bar{\lambda}^k) - HB(\bar{y}^k-y^k) \\ &= \bar{\lambda}^k - (1-\alpha)(\bar{\lambda}^k-\lambda^k) - HB(\bar{y}^k-y^k). \end{aligned} \quad (4.11)$$

Substituting this into (4.10) and using the notation $\bar{w}^k$, for any $y \in \mathcal{Y}$ we obtain

$$\begin{aligned} &\theta_2(y)-\theta_2(\bar{y}^k)+(y-\bar{y}^k)^T(-B^T\bar{\lambda}^k)+(y-\bar{y}^k)^T\xi_y^k \\ &\geq (y-y^{k+1})^TS(y^k-y^{k+1})+(y-\bar{y}^k)^TB^T[HB(y^k-\bar{y}^k)+(1-\alpha)(\lambda^k-\bar{\lambda}^k)] \\ &= \frac{1}{2}(\|y^{k+1}-y\|_S^2-\|y^k-y\|_S^2)+\frac{1}{2}\|y^k-\bar{y}^k\|_S^2+(y-\bar{y}^k)^TB^THB(y^k-\bar{y}^k) \\ &\quad +(1-\alpha)(y-\bar{y}^k)^TB^T(\lambda^k-\bar{\lambda}^k). \end{aligned} \quad (4.12)$$

Setting $\hat{\lambda}^{k+1}=\lambda^{k+1}$ in Lemma 3.1 and using $\bar{y}^k=y^{k+1}$, we have

$$\begin{aligned} &(\lambda-\bar{\lambda}^k)^T(A\bar{x}^k+B\bar{y}^k-b) \\ &= \frac{1}{\alpha}(y-\bar{y}^k)^TB^THB(\bar{y}^k-y^k)+\frac{1}{2\alpha}(\|y^{k+1}-y\|_{B^THB}^2-\|y^k-y\|_{B^THB}^2) \\ &\quad +\frac{1}{2\alpha}(\|\lambda^{k+1}-\lambda\|_{H^{-1}}^2-\|\lambda^k-\lambda\|_{H^{-1}}^2) \\ &\quad +\frac{2-\alpha}{2}\|\lambda^k-\bar{\lambda}^k\|_{H^{-1}}^2+\frac{1-\alpha}{\alpha}(y^k-\bar{y}^k)^TB^T(\lambda-\lambda^k). \end{aligned} \quad (4.13)$$

Combining (4.9), (4.12) and (4.13) together, we get

$$\begin{aligned} &\theta(u)-\theta(\bar{u}^k)+(w-\bar{w}^k)^TQ(\bar{w}^k)+(x-\bar{x}^k)^T\xi_x^k+(y-\bar{y}^k)^T\xi_y^k \\ &\geq \frac{1}{2}\Big[\big(\|x^{k+1}-x\|_G^2-\|x^k-x\|_G^2\big)+\big(\|y^{k+1}-y\|_{\frac{1}{\alpha}B^THB+S}^2-\|y^k-y\|_{\frac{1}{\alpha}B^THB+S}^2\big) \\ &\quad +\big(\|\lambda^{k+1}-\lambda\|_{\frac{1}{\alpha}H^{-1}}^2-\|\lambda^k-\lambda\|_{\frac{1}{\alpha}H^{-1}}^2\big)\Big]+\frac{1-\alpha}{\alpha}\Big[(y-\bar{y}^k)^TB^THB(\bar{y}^k-y^k) \\ &\quad +\alpha(y-\bar{y}^k)^TB^T(\lambda^k-\bar{\lambda}^k)+(y^k-\bar{y}^k)^TB^T(\lambda-\lambda^k)\Big] \\ &\quad +\frac{1}{2}\Big[\|x^k-\bar{x}^k\|_G^2+\|y^k-\bar{y}^k\|_S^2+(2-\alpha)\|\lambda^k-\bar{\lambda}^k\|_{H^{-1}}^2\Big]. \end{aligned} \quad (4.14)$$

From (4.11) and by simple manipulations, we obtain

$$\begin{aligned} &(y-\bar{y}^k)^TB^THB(\bar{y}^k-y^k)+\alpha(y-\bar{y}^k)^TB^T(\lambda^k-\bar{\lambda}^k)+(y^k-\bar{y}^k)B^T(\lambda-\lambda^k) \\ &= (y-\bar{y}^k)^TB^THB(\bar{y}^k-y^k)+(y-\bar{y}^k)^TB^T[\lambda-\alpha\bar{\lambda}^k-(1-\alpha)\lambda^k]-(y-y^k)^TB^T(\lambda-\lambda^k) \\ &= (y-\bar{y}^k)^TB^T[\lambda-\bar{\lambda}^k+(1-\alpha)(\bar{\lambda}^k-\lambda^k)+HB(\bar{y}^k-y^k)]-(y-y^k)^TB^T(\lambda-\lambda^k) \\ &= (y-\bar{y}^k)^TB^T(\lambda-\lambda^{k+1})-(y-y^k)^TB^T(\lambda-\lambda^k). \end{aligned}$$

Substituting the above inequality into (4.14) and using the notation P_α and M_α, we get (4.8) immediately. The proof is completed. □

Let us prove two more lemmas before proving the convergence of Algorithm 1.

Lemma 4.3. *Let the sequence $\{w^k\}$ be generated by Algorithm 1, the accompanying sequence $\{\bar{w}^k\}$ be defined by (4.7). Then for any $w = (x, y, \lambda) \in \mathcal{W}$ and $k \geq 0$, we have*

$$\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k) \geq \frac{1-\nu_k}{2}\|w^{k+1} - w\|^2_{M_\alpha} - \frac{1}{2}\|w^k - w\|^2_{M_\alpha} + \frac{1}{2}\|w^k - \bar{w}^k\|^2_{P_\alpha} - \nu_k, \tag{4.15}$$

where P_α and M_α are defined by (2.3).

Proof. It follows from Lemma 4.2 that

$$\begin{aligned} &\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k) + (x - \bar{x}^k)^T \xi_x^k + (y - \bar{y}^k)^T \xi_y^k \\ &\geq \frac{1}{2}(\|w^{k+1} - w\|^2_{M_\alpha} - \|w^k - w\|^2_{M_\alpha}) + \frac{1}{2}\|w^k - \bar{w}^k\|^2_{P_\alpha}. \end{aligned}$$

Note that $\bar{x}^k = x^{k+1}$. Using Cauchy-Schwarz Inequality and the inexactness criterion $\|\xi_x^k\|_{G^{-1}} \leq \nu_k$, we obtain

$$(\bar{x}^k - x)^T \xi_x^k \geq -\frac{\nu_k}{2}\|\bar{x}^k - x\|^2_G - \frac{1}{2\nu_k}\|\xi_x^k\|^2_{G^{-1}} \geq -\frac{\nu_k}{2}\|x^{k+1} - x\|^2_G - \frac{\nu_k}{2}, \qquad \forall x \in \mathcal{X}. \tag{4.16}$$

Similarly, we have

$$(\bar{y}^k - y)^T \xi_y^k \geq -\frac{\nu_k}{2}\|y^{k+1} - y\|^2_S - \frac{\nu_k}{2}, \qquad \forall y \in \mathcal{Y}. \tag{4.17}$$

Summing the above three inequalities and using the notation M_α, we get

$$\begin{aligned} &\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k) \\ &\geq \frac{1}{2}\|w^{k+1} - w\|^2_{M_\alpha} - \frac{\nu_k}{2}(\|x^{k+1} - x\|^2_G + \|y^{k+1} - y\|^2_S) - \frac{1}{2}\|w^k - w\|^2_{M_\alpha} + \frac{1}{2}\|w^k - \bar{w}^k\|^2_{P_\alpha} - \nu_k \\ &\geq \frac{1-\nu_k}{2}\|w^{k+1} - w\|^2_{M_\alpha} \\ &-\frac{1}{2}\|w^k - w\|^2_{M_\alpha} + \frac{1}{2}\|w^k - \bar{w}^k\|^2_{P_\alpha} - \nu_k. \end{aligned}$$

The proof is completed. □

The following result shows the contraction of the sequence generated by Algorithm 1, based on which the convergence of Algorithm 1 can be established easily.

Lemma 4.4. *Let the sequence $\{w^k\}$ be generated by Algorithm 1. Then for any $w^* \in \mathcal{W}^*$, we have*

$$\|w^{k+1} - w^*\|^2_{M_\alpha} \leq (1 + 2\nu_k)\|w^k - w^*\|^2_{M_\alpha} + (2\nu_k + 4\nu_k^2) - \|w^k - \bar{w}^k\|^2_{P_\alpha},$$

where P_α and M_α are defined by (2.3).

Proof. Setting $w = w^*$ in (4.15), we get

$$\begin{aligned} &2\big[\theta(u^*) - \theta(\bar{u}^k) + (w^* - \bar{w}^k)^T Q(\bar{w}^k)\big] \\ &\geq (1 - \nu_k)\|w^{k+1} - w^*\|^2_{M_\alpha} - \|w^k - w^*\|^2_{M_\alpha} + \|w^k - \bar{w}^k\|^2_{P_\alpha} - 2\nu_k \\ &\geq (1 - \nu_k)\|w^{k+1} - w^*\|^2_{M_\alpha} - \|w^k - w^*\|^2_{M_\alpha} + (1 - \nu_k)\|w^k - \bar{w}^k\|^2_{P_\alpha} - 2\nu_k. \end{aligned}$$

On the other hand, since Q is monotone and $w^* \in \mathcal{W}^*$, we have

$$0 \ge \theta(u^*) - \theta(\bar{u}^k) + (w^* - \bar{w}^k)^T Q(w^*) \ge \theta(u^*) - \theta(\bar{u}^k) + (w^* - \bar{w}^k)^T Q(\bar{w}^k).$$

Recall that $\sum_{k=0}^{\infty} \nu_k < +\infty$. Without loss of generality (say, $0 < \nu_k < 1/2$), we assume that

$$\frac{1}{1-\nu_k} \le 1 + 2\nu_k.$$

It follows from the above three inequalities that

$$\begin{aligned} \|w^{k+1} - w^*\|_{M_\alpha}^2 &\le \frac{1}{1-\nu_k}\|w^k - w^*\|_{M_\alpha}^2 + \frac{2\nu_k}{1-\nu_k} - \|w^k - \bar{w}^k\|_{P_\alpha}^2 \\ &\le (1+2\nu_k)\|w^k - w^*\|_{M_\alpha}^2 + 2\nu_k(1+2\nu_k) - \|w^k - \bar{w}^k\|_{P_\alpha}^2. \end{aligned}$$

The proof is completed. □

Now we are ready to prove the global convergence for Algorithm 1.

Theorem 4.5. *The sequence $\{w^k\}$ generated by Algorithm 1 converges to some w^∞ which is a solution of VI($\mathcal{W}, Q, \theta$).*

Proof. From $\sum_{k=0}^{\infty} \nu_k < +\infty$ and $\nu_k \ge 0$, it follows that

$$\sum_{k=0}^{\infty}(2\nu_k + 4\nu_k^2) < +\infty.$$

Setting $\zeta_k = 2\nu_k$, $\eta_k = 2\nu_k + 4\nu_k^2$, $c_0 = 1$ in (3.6), from Lemma 4.4 and Theorem 3.2 we have

$$\lim_{k\to\infty} \|w^k - \bar{w}^k\|_{P_\alpha} = 0.$$

And thus we get

$$\lim_{k\to\infty} \|w^k - \bar{w}^k\|_{M_\alpha} = 0.$$

Then it follows from (4.6), (4.7) and (3.3) that

$$\begin{aligned} \lim_{k\to\infty} \|w^{k+1} - \bar{w}^k\|_{M_\alpha} &= \lim_{k\to\infty} \frac{1}{\alpha}\|\lambda^{k+1} - \bar{\lambda}^k\|_{H^{-1}} \\ &= \lim_{k\to\infty} \frac{1}{\alpha}\|(1-\alpha)(\lambda^k - \bar{\lambda}^k) + H(By^k - B\bar{y}^k)\|_{H^{-1}} \\ &= 0. \end{aligned}$$

Note that $\lim_{k\to\infty} \nu_k = 0$. From (4.15) and the above three formulae, we get

$$\liminf_{k\to\infty} \left\{\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k)\right\} \ge 0, \qquad \forall w \in \mathcal{W}.$$

The convergence of Algorithm 1 is then obtained immediately from Theorem 3.2. □

4.3 Ergodic Worst-case $O(1/t)$ Convergence Rate

Now, we establish a worst-case $O(1/t)$ convergence rate in ergodic sense for Algorithm 1.

Theorem 4.6. *For any integer $t > 0$, there is a $\bar{w}_t \in \mathcal{W}$ which is a convex combination of the iterates $\bar{w}^0, \bar{w}^1, \cdots, \bar{w}^t$ defined by (4.7). Then for any $w \in \mathcal{W}$, we have*

$$\theta(\bar{u}_t)-\theta(u)+(\bar{w}_t-w)^TQ(w) \le \frac{1}{2(t+1)}\Big[\|w-w^0\|_{M_\alpha}^2+\sum_{k=0}^{t}\nu_k(\|w-w^{k+1}\|_{M_\alpha}^2+2)\Big], \quad (4.18)$$

where $\bar{w}_t := (\sum_{k=0}^{t}\bar{w}^k)/(t+1)$ and M_α is defined by (2.3).

Proof. From (4.15), we have

$$\theta(u)-\theta(\bar{u}^k)+(w-\bar{w}^k)^TQ(\bar{w}^k)+\frac{1}{2}\|w-w^k\|_{M_\alpha}^2 \ge \frac{1-\nu_k}{2}\|w-w^{k+1}\|_{M_\alpha}^2-\nu_k, \quad \forall w \in \mathcal{W}.$$

Since Q is monotone, we have

$$(w-\bar{w}^k)^TQ(w) \ge (w-\bar{w}^k)^TQ(\bar{w}^k), \quad \forall w \in \mathcal{W}.$$

It follows from the above two inequalities that

$$\theta(u)-\theta(\bar{u}^k)+(w-\bar{w}^k)^TQ(w)+\frac{1}{2}\|w-w^k\|_{M_\alpha}^2 \ge \frac{1-\nu_k}{2}\|w-w^{k+1}\|_{M_\alpha}^2-\nu_k, \ \ \forall w \in \mathcal{W}. \quad (4.19)$$

Summing the inequality (4.19) over $k = 0, 1, \cdots, t$, we obtain

$$\begin{aligned}
&(t+1)\theta(u)-\sum_{k=0}^{t}\theta(\bar{u}^k)+\Big[(t+1)w-(\sum_{k=0}^{t}\bar{w}^k)\Big]^TQ(w)+\frac{1}{2}\|w-w^0\|_{M_\alpha}^2\\
&\ge \frac{1}{2}\|w-w^{t+1}\|_{M_\alpha}^2-\frac{1}{2}\sum_{k=0}^{t}\nu_k(\|w-w^{k+1}\|_{M_\alpha}^2+2)\\
&\ge -\frac{1}{2}\sum_{k=0}^{t}\nu_k(\|w-w^{k+1}\|_{M_\alpha}^2+2), \qquad \forall w \in \mathcal{W}.
\end{aligned}$$

Since $\sum_{k=0}^{t}1/(t+1) = 1$, $\bar{w}_t$ is a convex combination of $\bar{w}^0, \bar{w}^1, \cdots, \bar{w}^t$ and thus $\bar{w}_t \in \mathcal{W}$. Using the notation of $\bar{w}_t$, we derive

$$\begin{aligned}
&\frac{1}{t+1}\sum_{k=0}^{t}\theta(\bar{u}^k)-\theta(u)+(\bar{w}_t-w)^TQ(w)\\
&\le \frac{1}{2(t+1)}\Big[\|w-w^0\|_{M_\alpha}^2+\sum_{k=0}^{t}\nu_k(\|w-w^{k+1}\|_{M_\alpha}^2+2)\Big], \quad \forall w \in \mathcal{W}. \quad (4.20)
\end{aligned}$$

Since $\theta(u)$ is convex and

$$\bar{u}_t := \frac{1}{t+1}\sum_{k=1}^{t}\bar{u}^k,$$

we have that

$$\theta(\bar{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\bar{u}^k).$$

Substituting it in (4.20), the assertion (4.18) follows immediately. □

We first notice that since the sequence $\{w^k\}$ generated by Algorithm 1 converges to the solution, it is bounded. According to (4.2), the sequence $\{\bar{w}^k\}$ defined by (4.7) is also bounded. Therefore, there exists a constant $D > 0$ such that

$$\|w^k\|_{M_\alpha} \le D \qquad \text{and} \qquad \|\bar{w}^k\|_{M_\alpha} \le D, \quad \forall k \ge 0.$$

Recall that $\bar{w}_t$ is the average of $\{\bar{w}^0, \bar{w}^1, \cdots, \bar{w}^t\}$. Thus, we have $\|\bar{w}_t\|_{M_\alpha} \le D$. Denote

$$E_1 := \sum_{k=0}^{\infty} \nu_k < +\infty. \tag{4.21}$$

For any $w \in \mathcal{B}_{\mathcal{W}}(\bar{w}_t) := \big\{w \in \mathcal{W} \mid \|w - \bar{w}_t\|_{M_\alpha} \le 1\big\}$, we get

$$\begin{aligned}
&\theta(\bar{u}_t) - \theta(u) + (\bar{w}_t - w)^T Q(w) \\
&\le \frac{1}{2(t+1)} \Big[\|w - w^0\|^2_{M_\alpha} + \sum_{k=0}^{t} \nu_k \big(\|w - w^{k+1}\|^2_{M_\alpha} + 2\big)\Big] \\
&\le \frac{1}{2(t+1)} \Big\{ \big(\|w - \bar{w}_t\|_{M_\alpha} + \|\bar{w}_t - w^0\|_{M_\alpha}\big)^2 + \sum_{k=0}^{t} \nu_k \Big[\big(\|w - \bar{w}_t\|_{M_\alpha} \\
&\qquad + \|\bar{w}_t - w^{k+1}\|_{M_\alpha}\big)^2 + 2\Big]\Big\} \\
&\le \frac{1}{2(t+1)} \Big\{ \big(\|w - \bar{w}_t\|_{M_\alpha} + \|\bar{w}_t\|_{M_\alpha} + \|w^0\|_{M_\alpha}\big)^2 \\
&\quad + \sum_{k=0}^{t} \nu_k \Big[\big(\|w - \bar{w}_t\|_{M_\alpha} + \|\bar{w}_t\|_{M_\alpha} + \|w^{k+1}\|_{M_\alpha}\big)^2 + 2\Big]\Big\} \\
&\le \frac{1}{2(t+1)} \Big\{(1+2D)^2 + E_1\big[\big(1+2D\big)^2 + 2\big]\Big\} \\
&= \frac{1}{2(t+1)} \big[(1+2D)^2(1+E_1) + 2E_1\big].
\end{aligned}$$

Thus, for any given $\varepsilon > 0$, after most $t := \lceil \frac{(1+2D)^2(1+E_1)+2E_1}{2\varepsilon} - 1 \rceil$ iterations, we have

$$\theta(\bar{u}_t) - \theta(u) + (\bar{w}_t - w)^T Q(w) \le \varepsilon, \quad \forall w \in \mathcal{B}_{\mathcal{W}}(\bar{w}_t),$$

which means $\bar{w}_t$ is an approximate solution of VI$(\mathcal{W}, Q, \theta)$ with an accuracy of $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate of Algorithm 1 is established in ergodic sense.

4.4 Nonergodic Worst-case $O(1/t)$ Convergence Rate

In this subsection, we establish a worst-case $O(1/t)$ convergence rate in nonergodic sense for Algorithm 1.

Recall that the sequence $\{w^k\}$ generated by Algorithm 1 is bounded. Thus, for any given $w^* \in \mathcal{W}^*$, there is a constant $\bar{C}_{w^*} > 0$ such that for any $k \ge 0$ we have

$$\|w^k - w^*\|_{M_\alpha} \le \bar{C}_{w^*}. \tag{4.22}$$

Theorem 4.7. *Let the sequence $\{w^k\}$ be generated by Algorithm 1, the accompanying sequence $\{\bar{w}^k\}$ be defined by (4.7). Then, for any $w^* \in \mathcal{W}^*$ we have*

$$\min_{i \in \{0,\cdots,k\}} \|w^i - \bar{w}^i\|^2_{P_\alpha} \le \frac{1}{k+1} \left[\|w^0 - w^*\|^2_{M_\alpha} + (\bar{C}^2_{w^*} + 2)E_1\right], \tag{4.23}$$

where P_α and M_α are defined by (2.3).

Proof. Setting $w = w^*$ in (4.15), for any $i \geq 0$ we get

$$2\left[\theta(u^*) - \theta(\bar{u}^i) + (w^* - \bar{w}^i)^T Q(\bar{w}^i)\right] \geq (1-\nu_i)\|w^{i+1} - w^*\|^2_{M_\alpha} - \|w^i - w^*\|^2_{M_\alpha} + \|w^i - \bar{w}^i\|^2_{P_\alpha} - 2\nu_i.$$

On the other hand, since Q is monotone, $\bar{w}^i \in \mathcal{W}$, and $w^* \in \mathcal{W}^*$, we have

$$0 \geq \theta(u^*) - \theta(\bar{u}^i) + (w^* - \bar{w}^i)^T Q(w^*) \geq \theta(u^*) - \theta(\bar{u}^i) + (w^* - \bar{w}^i)^T Q(\bar{w}^i).$$

It follows from the above two inequalities that

$$\|w^i - \bar{w}^i\|^2_{P_\alpha} \leq \|w^i - w^*\|^2_{M_\alpha} - (1-\nu_i)\|w^{i+1} - w^*\|^2_{M_\alpha} + 2\nu_i.$$

Summing the above inequality over $i = 0, 1, \cdots, k$ and using (4.22) and (4.21), we obtain

$$\sum_{i=0}^{k} \|w^i - \bar{w}^i\|^2_{P_\alpha} \leq \|w^0 - w^*\|^2_{M_\alpha} + \sum_{i=0}^{k} \nu_i \left(\|w^{i+1} - w^*\|^2_{M_\alpha} + 2\right) \leq \|w^0 - w^*\|^2_{M_\alpha} + (\bar{C}^2_{w^*} + 2)E_1.$$

The assertion (4.23) follows from the above inequality immediately. □

If $\|w^k - \bar{w}^k\|_{P_\alpha} = 0$, from (4.4), (4.5) and (4.7) we have

$$x^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(x^k) + (x - x^k)^T(-A^T\lambda^k) \geq -(x - x^k)^T \xi^k_x, \qquad \forall x \in \mathcal{X},$$

$$y^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^k) + (y - y^k)^T(-B^T\lambda^k) \geq -(y - y^k)^T \xi^k_y, \quad \forall y \in \mathcal{Y},$$

$$Ax^k + By^k - b = 0.$$

Combining the above three formulae together and using the inexactness criteria $\|\xi^k_x\|_{G^{-1}} \leq \nu_k$ and $\|\xi^k_y\|_{S^{-1}} \leq \nu_k$, we get

$$\begin{aligned} \theta(u) - \theta(u^k) + (w - w^k)^T Q(w^k) &\geq -(x - x^k)^T\xi^k_x - (y - y^k)^T\xi^k_y \\ &\geq -\|x - x^k\|_G \|\xi^k_x\|_{G^{-1}} - \|y - y^k\|_S \|\xi^k_y\|_{S^{-1}} \\ &\geq -\nu_k \left(\|x - x^k\|_G + \|y - y^k\|_S\right) \\ &\geq -\nu_k \|w - w^k\|_{P_\alpha} \\ &\geq -\nu_k, \qquad \forall w \in \mathcal{B}_{\mathcal{W}}(w^k), \end{aligned}$$

where

$$\mathcal{B}_{\mathcal{W}}(w^k) := \left\{w \in \mathcal{W} \mid \|w - w^k\|_{P_\alpha} \leq 1\right\}.$$

This means that $w^k \in \mathcal{W}$ can be regarded as a ν_k-approximation solution of VI$(\mathcal{W}, Q, \theta)$ according to (2.1). Therefore, $\|w^k - \bar{w}^k\|_{P_\alpha}$ can be viewed as an error measurement in terms of the distance to the solution set of VI$(\mathcal{W}, Q, \theta)$ for the $(k+1)$-th iteration of Algorithm 1. Hence, Theorem 4.7 shows a worst-case $O(1/t)$ convergence rate in nonergodic sense for Algorithm 1.

5 An Implementable Inexact GPADMM with Relative Error Control

In Section 4, the absolute errors of the error terms ξ_x^k and ξ_y^k are controlled by the summable sequence $\{\nu_k\}$. As in [28], we also investigate the case where the relative errors of the error terms ξ_x^k and ξ_y^k are controlled in this section. The control sequence $\{\nu_k\}$ can be also required to be summable: $\sum_{k=0}^{\infty}\nu_k < +\infty$. But in this section, we propose a more relaxable requirement on $\{\nu_k\}$ in [28]: $\sum_{k=0}^{\infty}\nu_k^2 < +\infty$.

5.1 Algorithm

Algorithm 2: An implementable inexact GPADMM with relative error control

Step 0. *Let* $\varepsilon > 0$; $\alpha \in (0,2)$; $w^0 := (x^0, y^0, \lambda^0) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{R}^m$; *G, S and H be positive definite matrices; and* $\{\nu_k\}$ *be a nonnegative sequence satisfying* $\sum_{k=0}^{\infty}\nu_k^2 < +\infty$. *Set* $k := 0$.

Step 1. *Find* $x^{k+1} \in \mathcal{X}$ *and* $\xi_x^k \in \mathcal{R}^n$ *such that*

$$\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\Big\{-A^T[\lambda^k-H(Ax^{k+1}+By^k-b)]+G(x^{k+1}-x^k)+\xi_x^k\Big\}\geq 0,\quad \forall x\in\mathcal{X}, \quad (5.1)$$

where ξ_x^k *satisfies the following inexactness criterion:*

$$\|\xi_x^k\|_{G^{-1}} \leq \nu_k\|x^k - x^{k+1}\|_G. \quad (5.2)$$

Step 2. *Find* $y^{k+1} \in \mathcal{Y}$ *and* $\xi_y^k \in \mathcal{R}^p$ *such that*

$$\begin{aligned}&\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\Big\{-B^T\big[\lambda^k-H\big(\alpha Ax^{k+1}-(1-\alpha)(By^k-b)+By^{k+1}-b\big)\big]\\&\quad+S(y^{k+1}-y^k)+\xi_y^k\Big\}\geq 0,\qquad \forall y\in\mathcal{Y},\end{aligned} \quad (5.3)$$

where ξ_y^k *satisfies the following inexactness criterion:*

$$\|\xi_y^k\|_{S^{-1}} \leq \nu_k\|y^k - y^{k+1}\|_S. \quad (5.4)$$

Step 3. *Update* λ^{k+1} *via*

$$\lambda^{k+1} := \lambda^k - H[\alpha Ax^{k+1} - (1-\alpha)(By^k-b) + By^{k+1} - b]. \quad (5.5)$$

Step 4. *Set* $w^{k+1} := (x^{k+1}, y^{k+1}, \lambda^{k+1})$. *If* $\|w^{k+1}-w^k\| \leq \varepsilon$, *stop; otherwise set* $k = k+1$ *and go to Step 1.*

Remark 5.1. Algorithm 2 with $\alpha = 1$ reduces to the inexact ADMM with Criterion 2 in [28].

5.2 Convergence

In this subsection, we prove the convergence of Algorithm 2. In addition, we define

$$\bar{w}^k := \begin{pmatrix}\bar{x}^k\\ \bar{y}^k\\ \bar{\lambda}^k\end{pmatrix} = \begin{pmatrix}x^{k+1}\\ y^{k+1}\\ \lambda^k - H(Ax^{k+1}+By^k-b)\end{pmatrix} \quad (5.6)$$

to further alleviate the complication of notation.

Let us start the proof with a lemma.

Lemma 5.2. *Let the sequence $\{w^k\}$ be generated by Algorithm 2, the accompanying sequence $\{\bar{w}^k\}$ be defined by (5.6). Then for any $w := (x, y, \lambda) \in \mathcal{W}$ and $k \geq 0$, we have*

$$\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k) \geq \frac{1 - 2\nu_k^2}{2}\|w^{k+1} - w\|_{M_\alpha}^2 - \frac{1}{2}\|w^k - w\|_{M_\alpha}^2 + \frac{1}{4}\|w^k - \bar{w}^k\|_{P_\alpha}^2, \tag{5.7}$$

where P_α and M_α are defined by (2.3).

Proof. From Lemma 4.2, we have

$$\begin{aligned} &\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k) + (x - \bar{x}^k)^T \xi_x^k + (y - \bar{y}^k)^T \xi_y^k \\ &\geq \frac{1}{2}(\|w^{k+1} - w\|_{M_\alpha}^2 - \|w^k - w\|_{M_\alpha}^2) + \frac{1}{2}\|w^k - \bar{w}^k\|_{P_\alpha}^2. \end{aligned} \tag{5.8}$$

It follows from (5.2) that

$$\|\xi_x^k\|_{G^{-1}}^2 \leq \nu_k^2 \|x^k - x^{k+1}\|_G^2.$$

Note that $\bar{x}^k = x^{k+1}$. Using Cauchy-Schwarz Inequality and the above inequality, we obtain

$$(\bar{x}^k - x)^T \xi_x^k \geq -\frac{2\nu_k^2}{2}\|\bar{x}^k - x\|_G^2 - \frac{1}{4\nu_k^2}\|\xi_x^k\|_{G^{-1}}^2 \geq -\nu_k^2\|x^{k+1} - x\|_G^2 - \frac{1}{4}\|x^k - \bar{x}^k\|_G^2, \qquad \forall x \in \mathcal{X}.$$

Similarly, we have

$$(\bar{y}^k - y)^T \xi_y^k \geq -\nu_k^2 \|y^{k+1} - y\|_S^2 - \frac{1}{4}\|y^k - \bar{y}^k\|_S^2, \qquad \forall y \in \mathcal{Y}.$$

Adding (5.8) and the above two inequalities, we get

$$\begin{aligned} &\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T Q(\bar{w}^k) \\ &\geq \frac{1}{2}\|w^{k+1} - w\|_{M_\alpha}^2 - \nu_k^2(\|x^{k+1} - x\|_G^2 + \|y^{k+1} - y\|_S^2) \\ &\quad -\frac{1}{2}\|w^k - w\|_{M_\alpha}^2 + \frac{1}{2}\|w^k - \bar{w}^k\|_{P_\alpha}^2 \\ &\quad -\frac{1}{4}(\|x^k - \bar{x}^k\|_G^2 + \|y^k - \bar{y}^k\|_S^2) \\ &\geq \frac{1 - 2\nu_k^2}{2}\|w^{k+1} - w\|_{M_\alpha}^2 - \frac{1}{2}\|w^k - w\|_{M_\alpha}^2 + \frac{1}{4}\|w^k - \bar{w}^k\|_{P_\alpha}^2, \quad \forall w \in \mathcal{W}. \end{aligned}$$

The proof is completed. □

The following result shows the contraction of the sequence generated by Algorithm 2, based on which the convergence of Algorithm 2 can be established easily.

Lemma 5.3. *Let the sequence $\{w^k\}$ be generated by Algorithm 2. Then for any $w^* \in \mathcal{W}^*$, we have*

$$\|w^{k+1} - w^*\|_{M_\alpha}^2 \leq (1 + 3\nu_k^2)\|w^k - w^*\|_{M_\alpha}^2 - \frac{1}{2}\|w^k - \bar{w}^k\|_{P_\alpha}^2,$$

where P_α and M_α are defined by (2.3).

Proof. Setting $w = w^*$ in (5.7), we get

$$2\big[\theta(u^*) - \theta(\bar{u}^k) + (w^* - \bar{w}^k)^T Q(\bar{w}^k)\big]$$

$$\geq (1-2\nu_k^2)\|w^{k+1}-w^*\|_{M_\alpha}^2 - \|w^k-w^*\|_{M_\alpha}^2 + \frac{1}{2}\|w^k-\bar{w}^k\|_{P_\alpha}^2$$
$$\geq (1-2\nu_k^2)\|w^{k+1}-w^*\|_{M_\alpha}^2 - \|w^k-w^*\|_{M_\alpha}^2 + \frac{1-2\nu_k^2}{2}\|w^k-\bar{w}^k\|_{P_\alpha}^2.$$

On the other hand, since Q is monotone and $w^* \in \mathcal{W}^*$, we have

$$0 \geq \theta(u^*) - \theta(\bar{u}^k) + (w^*-\bar{w}^k)^T Q(w^*) \geq \theta(u^*) - \theta(\bar{u}^k) + (w^*-\bar{w}^k)^T Q(\bar{w}^k).$$

Recall that $\sum_{k=0}^{\infty} \nu_k^2 < +\infty$. Without loss of generality (say, $0 < \nu_k < \sqrt{6}/6$), we assume that

$$\frac{1}{1-2\nu_k^2} \leq 1 + 3\nu_k^2.$$

It follows from the above three inequalities that

$$\begin{aligned}\|w^{k+1}-w^*\|_{M_\alpha}^2 &\leq \frac{1}{1-2\nu_k^2}\|w^k-w^*\|_{M_\alpha}^2 - \frac{1}{2}\|w^k-\bar{w}^k\|_{P_\alpha}^2 \\ &\leq (1+3\nu_k^2)\|w^k-w^*\|_{M_\alpha}^2 - \frac{1}{2}\|w^k-\bar{w}^k\|_{P_\alpha}^2.\end{aligned}$$

The proof is completed. □

Theorem 5.4. *The sequence $\{w^k\}$ generated by Algorithm 2 converges to some w^∞ which is a solution of VI$(\mathcal{W}, Q, \theta)$.*

Proof. Setting $\zeta_k = 3\nu_k^2$, $\eta_k \equiv 0$, $c_0 = 1/2$ in (3.6), from Lemma 5.3 and Theorem 3.2 we have

$$\lim_{k\to\infty} \|w^k - \bar{w}^k\|_{P_\alpha} = 0.$$

Then it follows from (5.5), (5.6) and (3.3) that

$$\begin{aligned}\lim_{k\to\infty} \|w^{k+1}-\bar{w}^k\|_{M_\alpha} &= \lim_{k\to\infty} \frac{1}{\alpha}\|\lambda^{k+1}-\bar{\lambda}^k\|_{H^{-1}} \\ &= \lim_{k\to\infty} \frac{1}{\alpha}\|(1-\alpha)(\lambda^k-\bar{\lambda}^k) + H(By^k - B\bar{y}^k)\|_{H^{-1}} \\ &= 0.\end{aligned}$$

From (5.7) and the above two formulae, we get

$$\liminf_{k\to\infty} \left\{\theta(u) - \theta(\bar{u}^k) + (w-\bar{w}^k)^T Q(\bar{w}^k)\right\} \geq 0, \qquad \forall w \in \mathcal{W}.$$

The convergence of Algorithm 2 is then obtained immediately from Theorem 3.2. □ □

5.3 Ergodic Worst-case $O(1/t)$ Convergence Rate

Now, we establish a worst-case $O(1/t)$ convergence rate in ergodic sense for Algorithm 2.

Theorem 5.5. *For any integer $t > 0$, there is a $\bar{w}_t \in \mathcal{W}$ which is a convex combination of the iterates $\bar{w}^0, \bar{w}^1, \cdots, \bar{w}^t$ defined by (5.6). Then for any $w \in \mathcal{W}$, we have*

$$\theta(\bar{u}_t) - \theta(u) + (\bar{w}_t - w)^T Q(w) \leq \frac{1}{t+1}\Big(\frac{1}{2}\|w-w^0\|_{M_\alpha}^2 + \sum_{k=0}^{t} \nu_k^2 \|w-w^{k+1}\|_{M_\alpha}^2\Big), \qquad (5.9)$$

where $\bar{w}_t := (\sum_{k=0}^{t} \bar{w}^k)/(t+1)$, and M_α is defined by (2.3).

Proof. From (5.7), we have

$$\theta(u)-\theta(\bar{u}^k)+(w-\bar{w}^k)^TQ(\bar{w}^k)+\frac{1}{2}\|w-w^k\|^2_{M_\alpha}\geq\frac{1-2\nu_k^2}{2}\|w-w^{k+1}\|^2_{M_\alpha},\quad \forall w\in\mathcal{W}.$$

Since Q is monotone, we have

$$(w-\bar{w}^k)^TQ(w)\geq(w-\bar{w}^k)^TQ(\bar{w}^k),\quad \forall w\in\mathcal{W}.$$

It follows from the above two inequalities that

$$\theta(u)-\theta(\bar{u}^k)+(w-\bar{w}^k)^TQ(w)+\frac{1}{2}\|w-w^k\|^2_{M_\alpha}\geq\frac{1-2\nu_k^2}{2}\|w-w^{k+1}\|^2_{M_\alpha},\qquad \forall w\in\mathcal{W}.\quad (5.10)$$

Summing the inequality (5.10) over $k=0,1,\cdots,t$, we obtain

$$\begin{aligned}&(t+1)\theta(u)-\sum_{k=0}^{t}\theta(\bar{u}^k)+\Big[(t+1)w-(\sum_{k=0}^{t}\bar{w}^k)\Big]^TQ(w)+\frac{1}{2}\|w-w^0\|^2_{M_\alpha}\\ &\geq\frac{1}{2}\|w-w^{t+1}\|^2_{M_\alpha}-\sum_{k=0}^{t}\nu_k^2\|w-w^{k+1}\|^2_{M_\alpha}\\ &\geq-\sum_{k=0}^{t}\nu_k^2\|w-w^{k+1}\|^2_{M_\alpha},\qquad \forall w\in\mathcal{W}.\end{aligned}$$

Since $\sum_{k=0}^{t}1/(t+1)=1$, $\bar{w}_t$ is a convex combination of $\bar{w}^0,\bar{w}^1,\cdots,\bar{w}^t$ and thus $\bar{w}_t\in\mathcal{W}$. Using the notation of $\bar{w}_t$, we derive

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\bar{u}^k)-\theta(u)+(\bar{w}_t-w)^TQ(w)\leq\frac{1}{t+1}\Big(\frac{1}{2}\|w-w^0\|^2_{M_\alpha}+\sum_{k=0}^{t}\nu_k^2\|w-w^{k+1}\|^2_{M_\alpha}\Big),\quad \forall w\in\mathcal{W}.\quad (5.11)$$

Since $\theta(u)$ is convex and

$$\bar{u}_t:=\frac{1}{t+1}\sum_{k=1}^{t}\bar{u}^k,$$

we have that

$$\theta(\bar{u}_t)\leq\frac{1}{t+1}\sum_{k=0}^{t}\theta(\bar{u}^k).$$

Substituting it in (5.11), the assertion (5.9) follows immediately. □ □

Analogous to Algorithm 1, Theorem 5.5 implies a worst-case $O(1/t)$ convergence rate in ergodic sense of Algorithm 2.

5.4 Nonergodic Worst-case $O(1/t)$ Convergence Rate

In this subsection, we establish a worst-case $O(1/t)$ convergence rate in nonergodic sense for Algorithm 2.

Since the sequence $\{w^k\}$ generated by Algorithm 2 converges to the solution, it is bounded. For given $w^* \in \mathcal{W}^*$, there is a constant $\tilde{C}_{w^*} > 0$ such that for any $k \geq 0$ we have

$$\|w^k - w^*\|_{M_\alpha} \leq \tilde{C}_{w^*}. \tag{5.12}$$

Denote

$$E_2 := \sum_{k=0}^{\infty} \nu_k^2 < +\infty. \tag{5.13}$$

Theorem 5.6. *Let the sequence $\{w^k\}$ be generated by Algorithm 2, the accompanying sequence $\{\bar{w}^k\}$ be defined by (5.6). Then, for any $w^* \in \mathcal{W}^*$ we have*

$$\min_{i\in\{0,\cdots,k\}} \|w^i - \bar{w}^i\|_{P_\alpha}^2 \leq \frac{1}{k+1}\Big(2\|w^0 - w^*\|_{M_\alpha}^2 + 4\tilde{C}_{w^*}^2 E_2\Big), \tag{5.14}$$

where P_α and M_α are defined by (2.3).

Proof. Setting $w = w^*$ in (5.7), for any $i \geq 0$ we get

$$4\left[\theta(u^*) - \theta(\bar{u}^i) + (w^* - \bar{w}^i)^T Q(\bar{w}^i)\right] \geq 2(1-2\nu_i^2)\|w^{i+1}-w^*\|_{M_\alpha}^2 - 2\|w^i-w^*\|_{M_\alpha}^2 + \|w^i-\bar{w}^i\|_{P_\alpha}^2.$$

On the other hand, since Q is monotone, $\bar{w}^i \in \mathcal{W}$, and $w^* \in \mathcal{W}^*$, we have

$$0 \geq \theta(u^*) - \theta(\bar{u}^i) + (w^* - \bar{w}^i)^T Q(w^*) \geq \theta(u^*) - \theta(\bar{u}^i) + (w^* - \bar{w}^i)^T Q(\bar{w}^i).$$

It follows from the above two inequalities that

$$\|w^i - \bar{w}^i\|_{P_\alpha}^2 \leq 2\|w^i - w^*\|_{M_\alpha}^2 - 2(1-2\nu_i^2)\|w^{i+1} - w^*\|_{M_\alpha}^2, \quad \forall i \geq 0.$$

Summing the above inequality over $i = 0, 1, \cdots, k$ and using (5.12) and (5.13), we obtain

$$\sum_{i=0}^{k} \|w^i - \bar{w}^i\|_{P_\alpha}^2 \leq 2\|w^0 - w^*\|_{M_\alpha}^2 + 4\sum_{i=0}^{k} \nu_i^2 \|w^{i+1} - w^*\|_{M_\alpha}^2 \leq 2\|w^0 - w^*\|_{M_\alpha}^2 + 4\tilde{C}_{w^*}^2 E_2.$$

The assertion (5.14) follows from the above inequality immediately. □ □

If $\|w^k - \bar{w}^k\|_{P_\alpha} = 0$, from (5.1)-(5.6) we have

$$\theta(u) - \theta(u^k) + (w - w^k)^T Q(w^k) \geq 0, \qquad \forall w \in \mathcal{W},$$

which means that w^k is a solution of $\mathrm{VI}(\mathcal{W}, Q, \theta)$ according to (2.1). Therefore, $\|w^k - \bar{w}^k\|_{P_\alpha}$ can be viewed as an error measurement in terms of the distance to the solution set of $\mathrm{VI}(\mathcal{W}, Q, \theta)$ for the $(k+1)$-th iteration of Algorithm 2. Hence, Theorem 5.6 shows a worst-case $O(1/t)$ convergence rate in nonergodic sense for Algorithm 2.

6 Conclusions

This paper studies the convergence of a general algorithmic framework which blends the inexact, generalized and proximal versions of the alternating direction method of multipliers (ADMM). The global convergence and worst-case $O(1/t)$ convergence rates in both ergodic and nonergodic senses are established for two concrete algorithms based on the inexact generalized proximal ADMM framework. The convergence rate results of some ADMM type algorithms in the literature are thus further developed in a uniform manner.

References

[1] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982.

[2] 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. Learning* 3 (2010) 1–122.

[3] T.F. Chan and R. Glowinski, Finite element approximation and iterative solution of a class of mildly nonlinear elliptic equations, Technical report, Computer Science Department, Stanford University, 1978.

[4] R.H. Chan, M. Tao and X.M. Yuan, Constrained total variational deblurring models and fast algorithms based on alternating direction method of multipliers, *SIAM J. Imaging Sci.* 6 (2013) 680–697.

[5] R.H. Chan, X.M. Yuan and W.X. Zhang, Point spread function reconstruction in ground-based astronomy by l1-lp model, *J. Opt. Soc. Am. A.* 29 (2012) 2263–2271.

[6] S.S. Chen, D.L. Donoho and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM Rev.* 43 (2001) 129–159.

[7] J. Eckstein, Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some Illustrative computational results, *RUTCOR Research Report* 32 (2012) 1–34.

[8] J. Eckstein, Parallel alternating direction multiplier decomposition of convex programs, *J. Optim. Theory Appli.* 80 (1994) 39–62.

[9] J. Eckstein and D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.* 55 (1992) 293–318.

[10] J. Eckstein and M.C. Ferris, Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control, *INFORMS J. Comput.* 10 (1998) 218–238.

[11] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I and II, Springer Series in Operations Research, Springer Verlag, New York, 2003.

[12] X.Y. Fang, B.S. He, H. Liu and X.M. Yuan, Generalized alternating direction method of multipliers: New theoretical insights and applications, *Math. Program. Comput.* 7 (2015) 149–187.

[13] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, *Comput. Math. Appl.* 2 (1976) 17–40.

[14] D. Gabay, Applications of the method of multipliers to variational inequalities, in *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski (eds.), North-Holland, Amsterdam, 1983, pp. 299–331.

[15] R. Glowinski, On alternating directon methods of multipliers: a historical perspective, *Modeling, Simulation and Optimization for Science and Technology Computational Methods in Applied Sciences* 34 (2014) 59–82.

[16] R. Glowinski and A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires, *R.A.I.R.O.* 2 (1975) 41–76.

[17] E.G. Golshtein and N.V. Tretyakov, Modified Lagrangians in convex programming and their generalizations, *Point-to-Set Maps and Mathematical Programming*, 1979, pp. 86–97.

[18] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[19] B.S. He, L.Z. Liao, D.R. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, *Math. Program.* 92 (2002) 103–118.

[20] B.S. He and X.M. Yuan, On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, *SIAM J. Num. Anal.* 50 (2012) 700–709.

[21] B.S. He and X. M. Yuan, On nonergodic convergence rate of Douglas-Rachford alternating direction method of multipliers, *Numer. Math.* 130 (2015) 567–577.

[22] M.R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.* 4 (1969) 303–320.

[23] X.X. Li, L.L. Mo, X.M. Yuan and J. Z. Zhang, Linearized alternating direction method of multipliers for sparse group and fused LASSO models, *Comput. Statist. Data Anal.* 79 (2014) 203–221.

[24] X.X. Li, M. Ng and X.M. Yuan, Median filtering based variational models for background extraction from surveillance video, *Numer. Linear Algebra Appl.* 22 (2015) 845–865.

[25] P.L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Num. Anal.* 16 (1979) 964–979.

[26] B. Martinet, Regularision d'inéquations variationnelles par approximations successive, *Rev. Francaise Inf. Rech, Oper.* 126 (1970) 154–159.

[27] Y. Nesterov, Gradient methods for minimizing composite objective function, *Core Discussion Paper* 76 (2007) 1–30.

[28] M. Ng, F. Wang and X.M. Yuan, Inexact alternating direction methods for image recovery, *SIAM J. Sci. Comput.* 33 (2011) 1643–1668.

[29] M. Ng, P. A. Weiss and X. M. Yuan, Solving constrained total-variation image reconstruction problems via alternating direction methods, *SIAM J. Sci. Comput.* 32 (2010) 2710–2736.

[30] M.J.D. Powell, A method for nonlinear constraints in minimization problems, in *Optimization*, R. Fletcher (eds.), Academic Press, 1969.

[31] X. Ren and Z. Lin, Linearized alternating direction method with adaptive penalty and warm starts for fast solving transform invariant low-rank textures, *Int. J. Comput. Vis.* 104 (2013) 1–14.

[32] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D* 60 (1992) 227–238.

[33] X.F. Wang and X.M. Yuan, The linearized alternating direction method for Dantzig Selector, *SIAM J. Sci. Comput.* 34 (2012) A2792–A2811.

[34] X.M. Yuan, Alternating direction method of multipliers for covariance selection models, *J. Sci. Comput.* 51 (2012), 261–273.

[35] X.M. Yuan and J.F. Yang, Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization, *Math. Comp.* 82 (2013) 301–329.

[36] X.Q. Zhang, M. Burger, X. Bresson and S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, *SIAM J. Imaging Sci.* 3 (2010) 253–276.

[37] X.Q. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, *J. Sci. Comput.* 46 (2010) 20–46.

Manuscript received 17 August 2015
revised 26 February 2016
accepted for publication 29 February 2016

Liming Sun
School of Science, Nanjing Audit University
Nanjing, 211815, China
E-mail address: lmsun@nau.edu.cn

Zhikai Jiang
School of Business, Huaihai Institute of Technology
Lianyungang, 222005, China
E-mail address: jiangzk@hhit.edu.cn

Xinxin Li
School of Mathematics, Jilin University
Changchun, 130000, China
E-mail address: xinxinli@jlu.edu.cn