On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers

Bingsheng He[†1] and Xiaoming Yuan[♮2]

[*]Department of Mathematics, Nanjing University, Nanjing, 210093, China
[♮]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

January 16, 2012

Abstract. Recently, a worst-case $O(1/t)$ convergence rate was established for the Douglas-Rachford alternating direction method of multipliers in an ergodic sense. This note proposes a novel approach to derive the same convergence rate while in a non-ergodic sense.

Keywords. Alternating direction method of multipliers, convergence rate, non-ergodic, separable convex programming.

1 Introduction

There has been an impressive development on operator splitting methods in the area of scientific computing, and among them are some alternating direction methods of multipliers (ADMs for short). In this note, we focus on the Douglas-Rachford ADM scheme proposed by Glowinski and Marrocco in [5] (see also [4]) and we restrict our discussion into the context of convex minimization problems with linear constraints and separable objective functions:

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

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$ and $\mathcal{Y} \subset \Re^{n_2}$ are closed convex sets, $\theta_1 : \Re^{n_1} \to \Re$ and $\theta_2 : \Re^{n_2} \to \Re$ are convex functions (not necessarily smooth). We assume the solution set of (1.1) to be nonempty, and we refer to [3, 6] for some convergence results without this assumption.

As in [8], with the attempt to treat the original ADM in [5] and the split inexact Uzawa method in [9, 10] uniformly, we study the following ADM scheme for (1.1):

$$x^{k+1} = \arg\min\big\{\theta_1(x) + \frac{\beta}{2}\|(Ax + By^k - b) - \frac{1}{\beta}\lambda^k\|^2 + \frac{1}{2}\|x - x^k\|_G^2 \;\big|\; x \in \mathcal{X}\big\}, \tag{1.2a}$$

$$y^{k+1} = \arg\min\big\{\theta_2(y) + \frac{\beta}{2}\|(Ax^{k+1} + By - b) - \frac{1}{\beta}\lambda^k\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \tag{1.2b}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \tag{1.2c}$$

where $\lambda^k \in \Re^m$ is the Lagrange multiplier, $\beta > 0$ is a penalty parameter, and $G \in \Re^{n_1\times n_1}$ is a symmetric and positive definite matrix. In fact, the original ADM scheme in [5] and the split inexact Uzawa method in [9, 10] are recovered by taking $G = 0$ and $G = (rI_{n_1} - \beta A^TA)$ (with $r > \|A^TA\|$) in (1.2a), respectively. We refer to a review paper [1] and references therein for the history of ADMs, and in particular, some efficient applications of ADMs exploited recently.

[1]Email: hebma@nju.edu.cn. This author was supported by the NSFC (grant 10971095 and 91130007), the Cultivation Fund of KSTIP, MOEC (708044) and the fund MOEC (20110091110004)
[2]Corresponding author. Email: xmyuan@hkbu.edu.hk. This author was supported by a grant from Hong Kong General Research Fund.

In [8], we have shown a worst-case $O(1/t)$ convergence rate (where t denotes the number of iterations) of the ADM scheme (1.2). To the best of our knowledge, it was the first time to establish a convergence rate for this ADM scheme. With the first-step result in [8], it becomes possible to investigate more intensive results on the convergence rate of the scheme (1.2). One possible improvement is to establish the same $O(1/t)$ worst-case convergence rate as in [8], but in a non-ergodic sense. Recall that the convergence rate derived in [8] is in an ergodic sense (see Theorem 2.2 for precise explanation). Here, our purpose is to provide a simple approach to derive such a convergence rate for (1.2) in a non-ergodic sense.

2 Preliminaries

In this section, we provide some preliminaries which are useful in later analysis.

2.1 Notations

We first define some matrices which will greatly simplify the notations in our analysis. More specifically, let

$$H = \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad M = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -\beta B & I_m \end{pmatrix}, \quad Q = \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.1}$$

Without further assumption on B, the matrix H defined above can only be guaranteed as symmetric and positive semidefinite. But we still use the notation $\|w\|_H^2$ to represent the non-negative number $w^T H w$ in our analysis. Based on these matrices, some relationship can be easily derived and we summarize them in the following proposition.

Proposition 2.1 *Let the matrices H, M and Q be defined in (2.1). Then we have*
1) $Q = HM$;
2) The matrix $(Q^T + Q) - M^T H M \succeq 0$, i.e., it is positive semidefinite.

Proof. The first conclusion is trivial, and we omit it. For the second one, we notice that

$$\begin{aligned}
(Q^T + Q) - M^T H M &= (Q^T + Q) - M^T Q \\
&= \begin{pmatrix} 2G & 0 & 0 \\ 0 & 2\beta B^T B & -B^T \\ 0 & -B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & -\beta B^T \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix} \\
&= \begin{pmatrix} 2G & 0 & 0 \\ 0 & 2\beta B^T B & -B^T \\ 0 & -B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} G & 0 & 0 \\ 0 & 2\beta B^T B & -B^T \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix} \\
&= \begin{pmatrix} G & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \succeq 0.
\end{aligned}$$

Thus, the proposition is proved. $\square$

2.2 Variational Inequality Characterization of (1.1)

It is easy to see that (1.1) is characterized by a variational inequality (VI) problem: Find $w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \Re^m$ such that

$$\text{VI}(\Omega, F, \theta): \qquad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.2a}$$

where

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

Note that the mapping $F(w)$ is monotone as it is affine with a skew-symmetric matrix. We denote by Ω^* the solution set of VI(Ω, F, θ). Then, Ω^* is nonempty under the nonempty assumption on the solution set of (1.1).

With this VI characterization, we can imitate Theorem 2.3.5 in [2] and prove the following theorem which provides a useful characterization on Ω^*.

Theorem 2.1 *(Theorem 2.1 in [8]) The solution set of VI(Ω, F, θ) is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \big\{\tilde{w} \in \Omega : \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0\big\}. \tag{2.3}$$

2.3 Ergodic Convergence Rate of (1.2)

Theorem 2.1 implies that $\tilde{w} \in \Omega$ is an approximate solution of VI(Ω, F, θ) with the accuracy $\epsilon > 0$ if it satisfies

$$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \leq \epsilon, \;\; \forall w \in \Omega. \tag{2.4}$$

Let $\{w^k\}$ be generated by the ADM scheme (1.2) and an associated sequence $\{\tilde{w}^k\}$ be defined by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}. \tag{2.5}$$

Then, the worst-case $O(1/t)$ convergence rate of (1.2) established in [8] can be summarized as the following theorem.

Theorem 2.2 *Let $\{w^k\}$ be the sequence generated by the scheme (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (2.5) and H be given in (2.1). For any integer number $t > 0$, let*

$$\tilde{w}_t := \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \tag{2.6}$$

Then, we have $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)} \|w - w^0\|_H^2, \quad \forall w \in \Omega. \tag{2.7}$$

Note this convergence rate is in an ergodic sense and $\tilde{w}_t$ is a convex combination of the previous vectors $\{\tilde{w}^0, \tilde{w}^1, \ldots, \tilde{w}^t\}$ with equal weights. One may ask if we can establish the same convergence rate directly for the sequence $\{w^k\}$ generated by the scheme (1.2), and this is the main purpose of this note.

2.4 Sketch of Proof

To establish a worst-case $O(1/t)$ convergence rate for the sequence $\{w^k\}$ generated by (1.2) in a non-ergodic sense, our proof follows the following steps:

1). to show that $\{w^k\}$ is contractive with respect to Ω^*, *i.e.*,

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2 \quad \forall\, w^* \in \Omega^*; \tag{2.8}$$

2). to show that $\{\|w^k - w^{k+1}\|_H^2\}$ is monotonically non-increasing, *i.e.*,

$$\|w^k - w^{k+1}\|_H^2 \leq \|w^{k-1} - w^k\|_H^2 \quad \forall\, k \geq 1; \tag{2.9}$$

3). to derive a worst-case $O(1/t)$ convergence rate in a non-ergodic sense based on (2.8) and (2.9), *i.e.*,

$$\|w^k - w^{k+1}\|_H^2 \leq \frac{1}{(k+1)}\|w^0 - w^*\|_H^2 \quad \forall w^* \in \Omega^*. \tag{2.10}$$

The reason why we measure the accuracy of iterates by $\|w^k - w^{k+1}\|_H^2$ will be explained in Lemma 3.2. In the following, our analysis is thus divided into three sections to address these three tasks.

3 Contraction

We prove the conclusion (2.8) in this section, and our proof is inspired by Theorem 1 in [7]. We first present several lemmas.

Lemma 3.1 *Let the sequence $\{w^k\}$ be generated by (1.2). Then, we have*

$$(y^k - y^{k+1})^T B^T(\lambda^k - \lambda^{k+1}) \geq 0, \quad \forall\, k \geq 0. \tag{3.1}$$

Proof. By deriving the optimality conditions of the minimization problem (1.2b), we have

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{B^T[\beta(Ax^{k+1} + By^{k+1} - b) - \lambda^k]\} \geq 0, \ \forall\, y \in \mathcal{Y}.$$

Substituting (1.2c) into the last inequality, we obtain

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T(-B^T\lambda^{k+1}) \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.2}$$

Obviously, analogous to (3.2), we have

$$y^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^k) + (y - y^k)^T(-B^T\lambda^k) \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.3}$$

Setting $y = y^k$ and $y = y^{k+1}$ in in (3.2) and (3.3), respectively, and then adding the two resulting inequalities, we get (3.1) immediately. $\square$

Lemma 3.2 *Let the sequence $\{w^k\}$ be generated by (1.2) and H be given in (2.1). Then, we have*

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T\{F(w^{k+1}) + \eta(y^k, y^{k+1}) + H(w^{k+1} - w^k)\} \geq 0, \quad \forall\, w \in \Omega, \tag{3.4}$$

where

$$\eta(y^k, y^{k+1}) := \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}). \tag{3.5}$$

Proof. First, by deriving the optimality condition of the minimization problem (1.2a), we have

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{A^T[\beta(Ax^{k+1} + By^k - b) - \lambda^k] + G(x^{k+1} - x^k)\} \ge 0,\ \forall\, x \in \mathcal{X}.$$

By using (1.2c), it can be written as

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) + G(x^{k+1} - x^k)\} \ge 0,\ \forall\, x \in \mathcal{X}. \quad (3.6)$$

It follows from (1.2) that

$$(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) = 0. \quad (3.7)$$

Combining (3.6), (3.2) and (3.7) together, we get $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$, such that

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) \\ -B^T\lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix} + \begin{pmatrix} G(x^{k+1} - x^k) \\ 0 \\ +\frac{1}{\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \right\} \ge 0, \quad \forall\, w \in \Omega,$$

which can be rewritten as

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix} + \beta \begin{pmatrix} A^T B(y^k - y^{k+1}) \\ B^T B(y^k - y^{k+1}) \\ 0 \end{pmatrix} + \begin{pmatrix} G(x^{k+1} - x^k) \\ \beta B^T B(y^{k+1} - y^k) \\ +\frac{1}{\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \right\} \ge 0, \ \ \forall\, w \in \Omega.$$

Using the notations of $F(w)$, $\eta(y^k, y^{k+1})$ and H, we get the assertion (3.4) immediately. □

Lemma 3.2 indicates that the quantity $\|w^k - w^{k+1}\|_H^2$ can be used to measure how accurate w^{k+1} is for being a solution of $\mathrm{VI}(\Omega, F, \theta)$. More specifically, since H is positive semidefinite, we conclude that $H(w^{k+1} - w^k) = 0$ and $\eta(y^k, y^{k+1}) = 0$ if $\|w^k - w^{k+1}\|_H^2 = 0$. In other words, if $\|w^k - w^{k+1}\|_H^2 = 0$, we have

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \ge 0, \quad \forall\, w \in \Omega,$$

which means w^{k+1} is a solution of $\mathrm{VI}(\Omega, F, \theta)$ according to (2.2). Therefore, $\|w^k - w^{k+1}\|_H^2$ can be viewed as an error measurement after k iterations of (1.2). This explains why we analyze the convergence rate of (1.2) in accordance with (2.10).

Lemma 3.3 *Let the sequence $\{w^k\}$ be generated by (1.2) and H be given in (2.1). Then, we have*

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \ge 0, \quad \forall\, w^* \in \Omega^*. \quad (3.8)$$

Proof. By setting w^* in (3.4), we obtain

$$\begin{aligned} &(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \\ &\ge \ \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) + (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}),\ \forall w^* \in \Omega^*. \end{aligned} \quad (3.9)$$

In the following we show that the right-hand side of (3.9) is non-negative. First, since $w^{k+1} \in \Omega$ and $w^* \in \Omega^*$, it follows from (2.2a) that

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0.$$

Because of the monotonicity of F, we have

$$(w^{k+1} - w^*)^T F(w^{k+1}) \geq (w^{k+1} - w^*)^T F(w^*).$$

Thus, we obtain

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \geq 0. \tag{3.10}$$

On the other hand, by using the notation of $\eta(y^k, y^{k+1})$ (see (3.5)), $Ax^* + By^* = b$ and (1.2c), we have

$$\begin{aligned} &(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \\ &= (y^k - y^{k+1})^T B^T \beta\{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\} \\ &= (y^k - y^{k+1})^T B^T \beta(Ax^{k+1} + By^{k+1} - b) \\ &= (y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}). \end{aligned}$$

Combining with (3.1), we get

$$(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \geq 0. \tag{3.11}$$

Substituting (3.10) and (3.11) into the right-hand side of (3.9), the lemma is proved. □

With the proved lemmas, we are now ready to show the assertion (2.8).

Theorem 3.1 *Let the sequence $\{w^k\}$ be generated by (1.2) and H be given in (2.1). Then (2.8) is satisfied for any $k \geq 0$.*

Proof. By using (3.8), we have

$$\begin{aligned} \|w^k - w^*\|_H^2 &= \|(w^{k+1} - w^*) + (w^k - w^{k+1}\|_H^2 \\ &= \|w^{k+1} - w^*\|_H^2 + 2(w^{k+1} - w^*)^T H(w^k - w^{k+1}) + \|w^k - w^{k+1}\|_H^2 \\ &\geq \|w^{k+1} - w^*\|_H^2 + \|w^k - w^{k+1}\|_H^2, \end{aligned}$$

and thus the assertion (2.8) is proved. □

4 Monotonicity

This section shows the assertion (2.9), i.e, the sequence $\{\|w^{k+1} - w^{k+2}\|_H^2\}$ is monotonically non-increasing. Again, we prove several lemmas for this purpose. First of all, we observe that w^{k+1} and $\tilde{w}^k$ defined in (2.5) are related by

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{4.1}$$

where the matrix M is given in (2.1).

Lemma 4.1 *Let $\{w^k\}$ be the sequence generated by (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (2.5) and Q be given in (2.1). Then, we have*

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + Q(\tilde{w}^k - w^k)\} \geq 0, \quad \forall\, w \in \Omega. \tag{4.2}$$

Proof. Using the notation $\tilde{w}^k$ in (2.5), and the facts

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b),$$

the inequality (3.8) can be rewritten as

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \begin{pmatrix} G(\tilde{x}^k - x^k) \\ \beta B^T B(\tilde{y}^k - y^k) \\ -B(y^k - \tilde{y}^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0, \quad \forall\, w \in \Omega.$$

The assertion (4.2) thus follows immediately from the definition of Q. □

Lemma 4.1 enables us to establish an important inequality in the following lemma.

Lemma 4.2 *Let $\{w^k\}$ be the sequence generated by (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (2.5) and Q be given in (2.1). Then, we have*

$$(\tilde{w}^k - \tilde{w}^{k+1})^T Q\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\} \geq 0. \tag{4.3}$$

Proof. Set $w = \tilde{w}^{k+1}$ in (4.2), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T\{F(\tilde{w}^k) + Q(\tilde{w}^k - w^k)\} \geq 0. \tag{4.4}$$

Note that (4.2) is also true for $k := k+1$ and thus we have

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T\{F(\tilde{w}^{k+1}) + Q(\tilde{w}^{k+1} - w^{k+1})\} \geq 0, \quad \forall\, w \in \Omega.$$

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T\{F(\tilde{w}^{k+1}) + Q(\tilde{w}^{k+1} - w^{k+1})\} \geq 0. \tag{4.5}$$

Adding (4.4) and (4.5) and using the monotonicity of F, we get (4.3) immediately. □

Lemma 4.3 *Let $\{w^k\}$ be the sequence generated by (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (2.5), the matrices H, M and Q be given in (2.1). Then, we have*

$$(w^k - \tilde{w}^k)^T M^T H M\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\} \geq \frac{1}{2}\|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|^2_{(Q^T+Q)}. \tag{4.6}$$

Proof. First, adding the term

$$\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\}^T Q\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\}$$

to the both sides of (4.3), and using $w^T Q w = \frac{1}{2} w^T(Q^T + Q)w$, we get

$$(w^k - w^{k+1})^T Q\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\} \geq \frac{1}{2}\|(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\|^2_{(Q^T+Q)}.$$

Reorder $(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})$ in the above inequality to $(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})$, we get

$$(w^k - w^{k+1})^T Q\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\} \geq \frac{1}{2}\|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|^2_{(Q^T+Q)}.$$

Substituting the term $(w^k - w^{k+1})$ into the left-hand side of the last inequality, and using the relationship in (4.1) and the fact $Q = HM$ in Proposition 2.1, we obtain (4.6). □

Finally, we are ready to show the assertion (2.9) in the following theorem.

Theorem 4.1 *Let $\{w^k\}$ be the sequence generated by (1.2) and H be given in (2.1). Then, (2.9) is satisfied for any $k \geq 0$.*

Proof. Setting $a = M(w^k - \tilde{w}^k)$ and $b = M(w^{k+1} - \tilde{w}^{k+1})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

$$\begin{aligned}&\|M(w^k - \tilde{w}^k)\|_H^2 - \|M(w^{k+1} - \tilde{w}^{k+1})\|_H^2 \\ &\quad= 2(w^k - \tilde{w}^k)^T M^T HM\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\} - \|M\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\}\|_H^2.\end{aligned}$$

Inserting (4.6) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned}&\|M(w^k - \tilde{w}^k)\|_H^2 - \|M(w^{k+1} - \tilde{w}^{k+1})\|_H^2 \\ &\quad\geq \|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|_{(Q^T+Q)}^2 - \|M\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\}\|_H^2 \\ &\quad= \|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|_{\{(Q^T+Q)-M^THM\}}^2 \\ &\quad\geq 0,\end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - M^T HM$ proved in Proposition 2.1. In other words, we derive

$$\|M(w^{k+1} - \tilde{w}^{k+1})\|_H^2 \leq \|M(w^k - \tilde{w}^k)\|_H^2. \tag{4.7}$$

Recall the relationship in (4.1). The assertion (2.9) follows immediately from (4.7). □

5 Non-ergodic Convergence Rate

With Theorems 3.1 and 4.1, we can prove the assertion (2.10). That is, a worst-case $O(1/t)$ convergence rate in a non-ergodic sense for the ADM scheme (1.2) is established.

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by (1.2). Then, the assertion (2.10) is satisfied.*

Proof. First, it follows from (2.8) that

$$\sum_{t=0}^{\infty} \|w^t - w^{t+1}\|_H^2 \leq \|w^0 - w^*\|_H^2, \quad \forall\, w^* \in \Omega^*. \tag{5.1}$$

According to Theorem 4.1, the sequence $\{\|w^t - w^{t+1}\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(k+1)\|w^k - w^{k+1}\|_H^2 \leq \sum_{i=0}^{k} \|w^i - w^{i+1}\|_H^2. \tag{5.2}$$

The assertion (2.10) follows from (5.1) and (5.2) immediately. □

Notice that Ω^* is convex and closed (see Theorem 2.1). Let $d := \inf\{\|w^0 - w^*\|_H \mid w^* \in \Omega^*\}$. Then, for any given $\epsilon > 0$, Theorem 5.1 shows that the ADM scheme (1.2) needs at most $\lfloor d^2/\epsilon \rfloor$ iterations to ensure that $\|w^k - w^{k+1}\|_H^2 \leq \epsilon$. Recall that w^{k+1} is a solution of VI(Ω, F, θ) if $\|w^k - w^{k+1}\|_H^2 = 0$ (see Lemma 3.2). A worst-case $O(1/t)$ convergence rate in a non-ergodic sense for the ADM scheme (1.2) is thus established in Theorem 5.1.

References

[1] 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), pp. 1-122.

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

[3] M. Fortin and R. Glowinski, Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems, Stud. Math. Appl. 15, NorthCHolland, Amsterdam, 1983.

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

[5] 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., R2 (1975), pp. 41-76.

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

[7] B. S. He and H. Yang, *Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities*, Oper. Res. Let., 23 (1998), pp. 151-161.

[8] B. S. He and X. M. Yuan, *On the $O(1/t)$ convergence rate of the alternating direction method*, SIAM J. Num. Analy., to appear.

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

[10] X. Q. Zhang, M. Burger and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Computing, 46(1)(2010), pp. 20-46.