Numer. Math. (2015) 130:567–577
DOI 10.1007/s00211-014-0673-6

Numerische Mathematik

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

Bingsheng He · Xiaoming Yuan

Received: 11 April 2013 / Revised: 3 November 2014 / Published online: 30 November 2014

Abstract This note proposes a novel approach to derive a worst-case $O(1/k)$ convergence rate measured by the iteration complexity in a non-ergodic sense for the Douglas–Rachford alternating direction method of multipliers proposed by Glowinski and Marrocco.

Mathematics Subject Classification 90C25 · 90C30

1 Introduction

There has been an impressive development on operator splitting methods in the area of partial differential equations, and among them are some alternating direction methods of multipliers (ADMMs for short). In this note, we focus on the Douglas–Rachford ADMM scheme proposed by Glowinski and Marrocco in [7] (see also [5]) and we restrict our discussion into the context of a convex minimization problem with linear

X. Yuan was supported by the General Research Fund from Hong Kong Research Grants Council: 203613.

B. He was supported by the NSFC Grant 91130007 and 11471156.

B. He
Department of Mathematics, International Centre of Management Science and Engineering,
Nanjing University, Nanjing 210093, China
e-mail: hebma@nju.edu.cn

X. Yuan (✉)
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
e-mail: xmyuan@hkbu.edu.hk

constraints and a separable objective function:

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

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). The solution set of (1.1) is assumed to be nonempty, and we refer to [4,8] for some convergence results without this assumption.

As in [10], in order to treat the original ADMM in [7] and the split inexact Uzawa method in [13] uniformly, we study the following ADMM scheme for (1.1):

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

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

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

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 semidefinite matrix. In fact, the original ADMM scheme in [7] and the split inexact Uzawa method in [13] are recovered by taking $G=0$ and $G=(rI_{n_1}-\beta A^TA)$ with $r>\beta\|A^TA\|$ (where $\|\cdot\|$ denotes a matrix norm such as the spectrum norm) in (1.2a), respectively. We refer to review papers [1,3,6] and references therein for the history of ADMMs, and in particular, some efficient applications of ADMMs exploited recently.

Because of its impressive efficiency and wide applicability, it is interesting to investigate the convergence rate of the ADMM scheme (1.2). A first-step work is [10] where we showed a worst-case $O(1/k)$ convergence rate measured by the iteration complexity, where k denotes the iteration counter, for the ADMM scheme (1.2)[1]. With this first-step result, it becomes possible to investigate more intensive results on the convergence rate of the scheme (1.2). Recall that the convergence rate derived in [10] is in the ergodic sense because the approximate solution with an accuracy of $O(1/k)$ is found based on all k iterates generated by (1.2). One may ask if we can establish the same convergence rate directly in a non-ergodic sense for the sequence generated by the scheme (1.2). The main purpose of this note is to answer this question affirmatively. We expect the new technique to be used to analyze convergence rates for other important numerical algorithms of the same kind.

2 Preliminaries

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

[1] As [11,12] and many others, a worst-case $O(1/k)$ convergence rate measured by the iteration complexity means the accuracy to a solution under certain criterion is of the order $O(1/k)$ after k iterations of an iterative scheme; or equivalently, it requires at most $O(1/\epsilon)$ iterations to achieve an approximate solution with an accuracy of ϵ.

2.1 Notations

We first define some matrices which will 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 symmetric matrix $(Q^T + Q) - M^T HM$ *is positive semidefinite:* $(Q^T + Q) - M^T HM \succeq 0$.

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

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

Thus, the proposition is proved. □

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

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

Note that the mapping $F(w)$ is monotone because 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).

3 Sketch of Proof

To establish a worst-case $O(1/k)$ convergence rate for the sequence $\{w^k\}$ generated by (1.2) in a non-ergodic sense, we need the assertion in the following lemma.

Lemma 3.1 *Let the sequence $\{w^k\}$ be generated by (1.2) and H be given in (2.1). Then, we have*

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \left\{F(w^{k+1}) + \eta(y^k, y^{k+1}) + H(w^{k+1} - w^k)\right\} \geq 0, \quad \forall\, w \in \Omega, \tag{3.1}$$

where

$$\eta(y^k, y^{k+1}) := \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B\,(y^k - y^{k+1}). \tag{3.2}$$

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

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

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

$$\begin{aligned} &\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) \\ &\quad + G(x^{k+1} - x^k)\} \geq 0, \forall\, x \in \mathcal{X}. \end{aligned} \tag{3.3}$$

It follows from (1.2) that

$$(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) = 0. \tag{3.4}$$

Combining (3.3), (4.2) and (3.4) together, we get $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$, such that

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

which can be rewritten as

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

Using the notations of $F(w)$, $\eta(y^k, y^{k+1})$ and H, we get the assertion (3.1) immediately. □

Lemma 3.1 indicates that the quantity $\|w^k - w^{k+1}\|_H^2$ can be used to measure the accuracy of the iterate w^{k+1} to a solution point of VI(Ω, F, θ). 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, because of the variational inequality characterization (2.2), when $\|w^k - w^{k+1}\|_H^2 = 0$, we have

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

which means w^{k+1} is a solution of VI(Ω, F, θ). Therefore, $\|w^k - w^{k+1}\|_H^2$ can be viewed as an error measurement after k iterations of the ADMM scheme (1.2), and it is reasonable to seek an upper bound of $\|w^k - w^{k+1}\|_H^2$ in term of the quantity $O(1/k)$ for the purpose of investigating the convergence rate of ADMM. Based on this fact, our proof follows the following steps:

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

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2 \quad \forall\, w^* \in \Omega^*; \tag{3.5}$$

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

$$\|w^k - w^{k+1}\|_H^2 \leq \|w^{k-1} - w^k\|_H^2 \quad \forall\, k \geq 1; \tag{3.6}$$

3. To derive a worst-case $O(1/k)$ convergence rate in a non-ergodic sense based on (3.5) and (3.6), i.e.,

$$\|w^k - w^{k+1}\|_H^2 \leq \frac{1}{(k+1)} \|w^0 - w^*\|_H^2 \quad \forall w^* \in \Omega^*. \tag{3.7}$$

In the following, our analysis is thus divided into three sections to address these three tasks.

4 Strict Contraction

We prove the conclusion (3.5) in this section, and our proof is inspired by Theorem 1 in [9]. We first present several lemmas.

Lemma 4.1 *Let the sequence* $\{w^k\}$ *be generated by (1.2). Then, we have*

$$(y^k - y^{k+1})^T B^T(\lambda^k - \lambda^{k+1}) \geq 0, \quad \forall\, k \geq 0. \tag{4.1}$$

Proof Deriving the optimality conditions of the minimization problem (1.2b), we have

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

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

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T(-B^T\lambda^{k+1}) \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{4.2}$$

Obviously, analogous to (4.2), we have

$$y^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^k) + (y - y^k)^T(-B^T\lambda^k) \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{4.3}$$

Setting $y = y^k$ and $y = y^{k+1}$ in in (4.2) and (4.3), respectively, and then adding the two resulting inequalities, we get (4.1) immediately. □

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

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \geq 0, \quad \forall\, w^* \in \Omega^*. \tag{4.4}$$

Proof Setting w^* in (3.1), we obtain

$$\begin{aligned}(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \\ \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \\ +(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}), \forall w^* \in \Omega^*.\end{aligned} \tag{4.5}$$

In the following we show that the right-hand side of (4.5) 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

$$\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \geq 0. \tag{4.6}$$

On the other hand, by using the notation of $\eta(y^k, y^{k+1})$ [see (3.2)], $Ax^* + By^* = b$ and (1.2c), we have

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

Combining with (4.1), we get

$$(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \geq 0. \tag{4.7}$$

Substituting (4.6) and (4.7) into the right-hand side of (4.5), the lemma is proved. □

With the proved lemmas, we are now ready to show the assertion (3.5).

Theorem 4.1 *Let the sequence* $\{w^k\}$ *be generated by (1.2) and* H *be given in (2.1). Then (3.5) is satisfied for any* $k \geq 0$.

Proof Using (4.4), 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 (3.5) is proved. □

5 Monotonicity

This section shows the assertion (3.6), i.e, the sequence $\{\|w^{k+1} - w^{k+2}\|_H^2\}$ is monotonically non-increasing. Again, we need to prove several lemmas for this purpose.

First of all, for the convenience of analysis, we introduce an auxiliary variable $\tilde{w}^k$ defined as

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{5.1}$$

where w^k is generated by (1.2). Then, we have the relationship

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{5.2}$$

where the matrix M is given in (2.1).

Lemma 5.1 *Let $\{w^k\}$ be the sequence generated by (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (5.1) and Q be given in (2.1). Then, we have*

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \left\{F(\tilde{w}^k) + Q(\tilde{w}^k - w^k)\right\} \geq 0, \quad \forall\, w \in \Omega. \tag{5.3}$$

Proof By using the notation $\tilde{w}^k$ in (5.1), and the facts

$$\frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) = (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m,$$

the inequality (3.1) can be rewritten as

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

The assertion (5.3) thus follows immediately from the definition of Q. □

Lemma 5.1 enables us to establish an important inequality in the following lemma.

Lemma 5.2 *Let $\{w^k\}$ be the sequence generated by (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (5.1) and Q be given in (2.1). Then, we have*

$$(\tilde{w}^k - \tilde{w}^{k+1})^T Q\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\} \geq 0. \tag{5.4}$$

Proof Setting $w = \tilde{w}^{k+1}$ in (5.3), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T\{F(\tilde{w}^k) + Q(\tilde{w}^k - w^k)\} \geq 0. \tag{5.5}$$

Note that (5.3) is also true for $k := k + 1$ and thus we have

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

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

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T\{F(\tilde{w}^{k+1}) + Q(\tilde{w}^{k+1} - w^{k+1})\} \geq 0. \tag{5.6}$$

Adding (5.5) to (5.6) and using the monotonicity of F, we get (5.4) immediately. □

Lemma 5.3 *Let $\{w^k\}$ be the sequence generated by (1.2), the associated sequence $\{\tilde{w}^k\}$ be defined by (5.1), the matrices H, M and Q be given in (2.1). Then, we have*

$$(w^k-\tilde{w}^k)^T M^T H M \left\{(w^k-\tilde{w}^k)-(w^{k+1}-\tilde{w}^{k+1})\right\} \geq \frac{1}{2}\|(w^k-\tilde{w}^k) -(w^{k+1}-\tilde{w}^{k+1})\|^2_{(Q^T+Q)}. \tag{5.7}$$

Proof First, adding the term

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

to the both sides of (5.4), and using $w^TQw=\frac{1}{2}w^T(Q^T+Q)w$, we get

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

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

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

Substituting the term (w^k-w^{k+1}) into the left-hand side of the last inequality, and using the relationship in (5.2) and the fact $Q=HM$ in Proposition 2.1, we obtain (5.7). □

Finally, we are ready to show the assertion (3.6) in the following theorem.

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by (1.2) and H be given in (2.1). Then, (3.6) 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

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

we obtain

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

Inserting (5.7) 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 $(Q^T+Q)-M^THM$ proved in Proposition 2.1. In other words, we derive

$$\|M(w^{k+1}-\tilde{w}^{k+1})\|_H^2\leq\|M(w^k-\tilde{w}^k)\|_H^2. \tag{5.8}$$

Recall the relationship in (5.2). The assertion (3.6) follows immediately from (5.8). □

6 Non-ergodic convergence rate

With Theorems 4.1 and 5.1, we can prove the assertion (3.7). That is, a worst-case $O(1/k)$ convergence rate in a non-ergodic sense for the ADMM scheme (1.2) is established.

Theorem 6.1 *Let $\{w^k\}$ be the sequence generated by (1.2). Then, the assertion (3.7) is satisfied.*

Proof First, it follows from (3.5) that

$$\sum_{t=0}^{\infty}\|w^t-w^{t+1}\|_H^2\leq\|w^0-w^*\|_H^2,\quad\forall\, w^*\in\Omega^*. \tag{6.1}$$

According to Theorem 5.1, the sequence $\{\|w^t-w^{t+1}\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(k+1)\|w^k-w^{k+1}\|_H^2\leq\sum_{t=0}^{k}\|w^t-w^{t+1}\|_H^2. \tag{6.2}$$

The assertion (3.7) follows from (6.1) and (6.2) immediately. □

Notice that Ω^* is convex and closed (see Theorem 2.1 in [10]). Let $d:=\inf\{\|w^0-w^*\|_H \mid w^*\in\Omega^*\}$. Then, for any given $\epsilon>0$, Theorem 6.1 shows that the ADMM 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 3.1). A worst-case $O(1/k)$ convergence rate in a non-ergodic sense for the ADMM scheme (1.2) is thus established in Theorem 6.1.

Finally, we remark that because of the monotonicity of the sequence $\{\|w^k-w^{k+1}\|_H^2\}$ and the fact (6.1), and using Lemma 1.2 in [2], we can immediately refine the worst-case convergence rate in Theorem 6.1 from $O(1/k)$ to $o(1/k)$.

References

1. 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**, 1–122 (2010)
2. Deng, W., Lai, M. J., Peng, Z. M., Yin, W. T.: Parallel multi-block ADMM with $o(1/k)$ convergence. 2014 (in press)
3. Eckstein, J., Yao, W.: Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results. RUTCOR Research Report RRR 32–2012 (2012)
4. Fortin, M., Glowinski, R.: Augmented lagrangian methods: applications to the numerical solutions of boundary value problems. Stud. Math. Appl. 15. NorthHolland, Amsterdam (1983)
5. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**, 17–40 (1976)
6. Glowinski, R.: On alternating direction methods of multipliers: a historical perspective. In: Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmaki, P., Pironneau, O. (eds.) Modeling, Simulation and Optimization for Science and Technology. Computational Methods in Applied Sciences, vol. 34, pp. 59–82. Springer, Dordrecht (2014)
7. Glowinski, R., Marroco, A.: Approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, vol. 9, R2, pp. 41–76 (1975)
8. Glowinski, R., Le Tallec, P.: Augmented lagrangian and operator-splitting methods in nonlinear mechanics. SIAM Studies in Applied Mathematics, Philadelphia (1989)
9. He, B.S., Yang, H.: Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. Oper. Res. Lett. **23**, 151–161 (1998)
10. He, B.S., Yuan, X.M.: On the O(1/n) convergence rate of Douglas–Rachford alternating direction method. SIAM J. Numer. Anal. **50**, 700–709 (2012)
11. Nemirovsky, A. S., Yudin, D. B.: Problem complexity and method efficiency in Optimization. Wiley-Interscience Series in Discrete Mathematics, Wiley, New York (1983)
12. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR **269**, 543–547 (1983)
13. Zhang, X.Q., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3**(3), 253–276 (2010)