SIAM J. NUMER. ANAL.
Vol. 50, No. 2, pp. 700–709

ON THE $O(1/n)$ CONVERGENCE RATE OF THE DOUGLAS–RACHFORD ALTERNATING DIRECTION METHOD*

BINGSHENG HE† AND XIAOMING YUAN‡

Abstract. Alternating direction methods (ADMs) have been well studied in the literature, and they have found many efficient applications in various fields. In this note, we focus on the Douglas–Rachford ADM scheme proposed by Glowinski and Marrocco, and we aim at providing a simple approach to estimating its convergence rate in terms of the iteration number. The linearized version of this ADM scheme, which is known as the split inexact Uzawa method in the image processing literature, is also discussed.

Key words. alternating direction method, convergence rate, split inexact Uzawa method, variational inequalities, convex programming

DOI. 10.1137/110836936

1. Introduction.

Operator splitting methods such as alternating direction methods (ADMs) have been well studied in the area of scientific computing; see, e.g., [9, 10, 11, 12, 15, 19, 21, 22, 27]. In this note, we focus on the Douglas–Rachford ADM scheme proposed by Glowinski and Marrocco in [14], and we restrict our discussion to the context of convex programming problems with linear constraints. That is, we study

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

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, and $\theta_1 : \Re^{n_1} \to \Re$ 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 the reader to [9, 15] for some convergence results without this assumption.

For solving (1.1), the iterative scheme of the Douglas–Rachford ADM scheme in [14] reads

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

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

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

*Received by the editors June 10, 2011; accepted for publication (in revised form) January 3, 2012; published electronically April 10, 2012.
http://www.siam.org/journals/sinum/50-2/83693.html

†Department of Mathematics, Nanjing University, Nanjing 210093, China (hebma@nju.edu.cn). This author's work was supported by NSFC grants 10971095 and 91130007, the Cultivation Fund of KSTIP-MOEC 708044, and the MOEC fund 20110091110004.

‡Department of Mathematics, Hong Kong Baptist University, Hong Kong, China (xmyuan@hkbu.edu.hk). This author's work was supported by the Hong Kong General Research Fund 203311 and by HKBU grant FRG2/10-11/069.

where $\lambda^k \in \Re^m$ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter. As elaborated in [3], the scheme (1.2) blends the ideas of the augmented Lagrangian method in [18, 28] with the Gauss–Seidel decomposition, and thus it becomes possible to exploit the properties of θ_1 and θ_2 individually. The close relationship between (1.2) and the Douglas–Rachford iterative schemes in [4, 21] has also been analyzed in [3]. Later, for simplicity, the scheme (1.2) is simply referred to as the ADM.

The ADM (1.2) has received much attention in the fields of partial differential equations and optimization; see, e.g., [3, 9, 10, 11, 15] for earlier literature, and [5, 8, 16, 20, 31] for some studies in variational inequalities and convex programming. Very recently, this ADM scheme has found some efficient applications in a broad spectrum of areas such as imaging processing, statistical learning, and engineering; see, e.g., [1, 2, 6, 17, 26, 29, 30, 33]. Elegant convergence analysis including some estimation on its convergence speed (and other ADM-type methods in more general settings) can be found in [13, 15, 21], from which the empirical efficiency of ADM is theoretically supported. As mentioned in [1], the ADM (1.2) is "at least comparable to very specialized algorithms (even in the serial setting), and in most cases, the simple ADM algorithm will be efficient enough to be useful."

Recently, under some additional assumptions (e.g., the full rank assumption on A), and under the restriction that both the subproblems (1.2a) and (1.2b) must be solved exactly, in [23] the scheme (1.2) is shown to be convergent with the $O(1/n)$ rate, where n denotes the iteration number.[1] These additional assumptions for deriving the $O(1/n)$ convergence rate of ADM, however, exclude many interesting applications where the efficiency of ADM has been shown in the literature. In particular, the requirement to obtain exact solutions of (1.2a) and (1.2b) is somehow too restrictive, even when the resolvent operators of $\partial\theta_1$ and $\partial\theta_2$ have closed-form representations. Here, $\partial(\cdot)$ denotes the subdifferential operator of a convex function, and the resolvent operator of, say, $\partial\theta_1$ is given by

$$\left(I + \frac{1}{r}\partial\theta_1\right)^{-1}(a) = \arg\min\left\{\theta_1(x) + \frac{r}{2}\|x-a\|^2 \,\middle|\, x \in \Re^{n_1}\right\}, \tag{1.3}$$

where $a \in \Re^{n_1}$ and $r > 0$. A simple instance is that occurring when $\theta_1(x) = \|x\|_1$ (thus the resolvent operator of $\partial\theta_1$ has a closed-form representation) but A is not the identity matrix. Then, it is not possible to obtain the exact solution of (1.2a). For these reasons, the convergence rate result in [23] is not valid for some of the aforementioned ADM's novel applications, including those in the area of total variation image restoration problems (e.g., [1, 2, 6, 26]).

The original ADM scheme (1.2) is the basis of many efficient algorithms developed recently. For a general case where the subproblems (1.2a) and (1.2b) do not have closed-form solutions or it is not easy to solve them to a high precision, inner iterative procedures are required to pursue approximate solutions of these subproblems. Thus, customized strategies with respect to particular properties of θ_1 and θ_2 are critical to ensure the efficiency of ADM for these cases. A success in this regard is the split inexact Uzawa method proposed in [34, 35]. Under the assumption that the resolvent operator of $\partial\theta_1$ has a closed-form representation, the authors of [34, 35] suggested linearizing the quadratic term in (1.2a) and solving the following approximate

[1]The discussion in [23] is for finding roots of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable two-block structure.

problem:

$$x^{k+1} = \arg\min \left\{ \theta_1(x) + \beta(x - x^k)^T \left(A^T \left(Ax^k + By^k - b - \frac{1}{\beta}\lambda^k \right) \right) + \frac{r}{2}\|x - x^k\|^2 \middle| x \in \mathcal{X} \right\}, \tag{1.4}$$

where the requirement on r is $r > \beta\|A^T A\|$. Thus, for the linearized subproblem (1.4) it is easy enough to have a closed-form solution. This linearization strategy was shown in [34, 35] to be very efficient for some image restoration problems. Note that the linearization on (1.2b) was not discussed in [34, 35], as B is usually an identity matrix and θ_2 is usually a simple function such as the least-squares function for applications therein. For convenience, we also focus on the linearization merely of (1.2a). To the best of our knowledge, the convergence rate of the split inexact Uzawa method [34, 35] is not yet established in the literature.

Our purpose is to provide a simple unified proof for the $O(1/n)$ convergence rate (in the worst case) for both the original ADM scheme (1.2) (in the general context) and its linearized variant (i.e., the split inexact Uzawa method proposed in [34, 35]). The conjecture in [23] about whether the subproblems (or one of them) in (1.2) can be solved approximately is also answered affirmatively. Our investigation on the convergence rate of (1.2) is motivated by the encouraging achievement in estimating convergence rate or iteration complexity for various first-order algorithms in the literature (see, e.g., [24, 25, 32]). However, our analysis is based on a variational inequality (VI) approach, and it differs significantly from existing approaches in the literature. A key tool for our analysis is a solution-set characterization of variational inequalities introduced in [7], and this result enables us to find a very simple proof for the $O(1/n)$ convergence rate of (1.2) and its linearized variant.

Note that both (1.2a) and (1.4) can be treated uniformly by

$$x^{k+1} = \arg\min \left\{ \theta_1(x) + \frac{\beta}{2} \left\| (Ax + By^k - b) - \frac{1}{\beta}\lambda^k \right\|^2 + \frac{1}{2}\|x - x^k\|_G^2 \middle| x \in \mathcal{X} \right\}, \tag{1.5}$$

where $G \in \Re^{n_1 \times n_1}$ is a symmetric and positive semidefinite matrix (we denote it $\|x\|_G := \sqrt{x^T G x}$). In fact, (1.2a) is recovered when $G = 0$, and (1.4) is recovered when $G = (rI_{n_1} - \beta A^T A)$ (subject to some constant variation in the objective function). Therefore, our analysis is for the following uniform scheme of both (1.2) and the split inexact Uzawa method in [34, 35]:

$$x^{k+1} = \arg\min \left\{ \theta_1(x) + \frac{\beta}{2} \left\| (Ax + By^k - b) - \frac{1}{\beta}\lambda^k \right\|^2 + \frac{1}{2}\|x - x^k\|_G^2 \;\middle|\; x \in \mathcal{X} \right\}, \tag{1.6a}$$

$$y^{k+1} = \arg\min \left\{ \theta_2(y) + \frac{\beta}{2} \left\| (Ax^{k+1} + By - b) - \frac{1}{\beta}\lambda^k \right\|^2 \;\middle|\; y \in \mathcal{Y} \right\}, \tag{1.6b}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{1.6c}$$

2. Preliminaries. In this section, we first reformulate (1.1) into a VI reformulation and then characterize its solution set by extending Theorem 2.3.5 in [7]. This characterization makes it possible to analyze ADM's convergence rate via the VI approach.

It is easy to see that the VI reformulation of (1.1) is: Find $w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \Re^m$ such that

$$\text{VI}(\Omega, F, \theta): \qquad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0 \quad \forall\, w \in \Omega, \tag{2.1a}$$

where

(2.1b)

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

Then, the mapping $F(w)$ is affine with a skew-symmetric matrix, and it is thus monotone. Furthermore, the solution set of $\text{VI}(\Omega, F, \theta)$, denoted by Ω^*, is nonempty under the nonempty assumption on the solution set of (1.1).

Next, we specify Theorem 2.3.5 in [7] for $\text{VI}(\Omega, F, \theta)$, and this characterization is the basis of our analysis for the convergence rate of ADM via the VI approach. The proof of the next result is an incremental extension of Theorem 2.3.5 in [7]. However, we include all the details for completeness.

THEOREM 2.1. *The solution set of $VI(\Omega, F, \theta)$ is convex and can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \left\{ \tilde{w} \in \Omega : \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0 \right\}. \tag{2.2}$$

Proof. Indeed, if $\tilde{w} \in \Omega^*$, we have

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

Since the monotonicity of F implies

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \geq 0 \;\; \forall w \in \Omega,$$

we have

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

Thus, $\tilde{w}$ belongs to the right-hand set in (2.2). Conversely, suppose that $\tilde{w}$ belongs to the latter set. Let $w \in \Omega$ be arbitrary. The vector

$$\bar{w} = \alpha \tilde{w} + (1 - \alpha) w$$

belongs to Ω for all $\alpha \in (0, 1)$. Thus we have

$$\theta(\bar{u}) - \theta(\tilde{u}) + (\bar{w} - \tilde{w})^T F(\bar{w}) \geq 0. \tag{2.3}$$

Because $\theta(\cdot)$ is convex, we have

$$\theta(\bar{u}) \leq \alpha \theta(\tilde{u}) + (1 - \alpha) \theta(u).$$

Substituting this into (2.3), we get

$$\theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\alpha\tilde{w} + (1-\alpha)w) \geq 0$$

for all $\alpha \in (0,1)$. Letting $\alpha \to 1$ yields

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

Thus $\tilde{w} \in \Omega^*$. Now, we turn to proving the convexity of Ω^*. For each fixed but arbitrary $w \in \Omega$, the set

$$\{\tilde{w} \in \Omega : \theta(\tilde{u}) + \tilde{w}^T F(w) \leq \theta(u) + w^T F(w)\}$$

and its equivalent expression

$$\{\tilde{w} \in \Omega : \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0\}$$

are convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of $\mathrm{VI}(\Omega, F, \theta)$ is convex. $\square$

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

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

In the rest, we show that after n iterations of the ADM (1.6), we can find $\tilde{w} \in \Omega$ such that (2.4) is satisfied with $\epsilon = O(1/n)$, thus proving a convergence rate of $O(1/n)$ in the worst case for the ADM algorithm (1.6).

3. Some properties. In this section, we prove several properties which are useful for establishing the main result.

First, to make the notation of the proof more succinct, we introduce some matrices,

$$\begin{gathered} H = \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad M = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -\beta B & I_m \end{pmatrix}, \\ Q = \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix}. \end{gathered} \tag{3.1}$$

Obviously, we have $Q = HM$.

Then, with the sequence $\{w^k\}$ generated by the ADM scheme (1.6), we define a new sequence by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}. \tag{3.2}$$

As we shall show later (see (4.1)), our analysis of the convergence rate is based on the sequence $\{\tilde{w}^k\}$. Note that (3.2) implies the relationship

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{3.3}$$

which is useful later.

Now, we start to prove some properties of the sequence $\{\tilde{w}^k\}$. The first lemma quantifies the discrepancy between the point $\tilde{w}^k$ and a solution point of VI(Ω, F, θ).

LEMMA 3.1. *Let* $\{\tilde{w}^k\}$ *be defined by* (3.2), *and the matrix* Q *be given in* (3.1). *Then we have*

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) - Q(w^k - \tilde{w}^k)\} \geq 0 \quad \forall\, w \in \Omega. \tag{3.4}$$

Proof. First, by deriving the optimality conditions of the minimization problems in (1.6), we have

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{A^T[\beta(Ax^{k+1} + By^k - b) - \lambda^k] + G(x^{k+1} - x^k)\} \geq 0 \;\forall\, x \in \mathcal{X} \tag{3.5}$$

and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{B^T[\beta(Ax^{k+1} + By^{k+1} - b) - \lambda^k]\} \geq 0 \;\forall\, y \in \mathcal{Y}. \tag{3.6}$$

Then, by using the notation $\tilde{w}^k$ in (3.2), the inequalities (3.5) and (3.6) can be respectively written as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \tilde{\lambda}^k + G(\tilde{x}^k - x^k)\} \geq 0 \quad \forall\, x \in \mathcal{X} \tag{3.7}$$

and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \geq 0 \quad \forall\, y \in \mathcal{Y}. \tag{3.8}$$

In addition, it follows from (1.6) and (3.2) that

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{3.9}$$

Combining (3.7), (3.8), and (3.9), we get $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$. For any $w = (x, y, \lambda) \in \Omega$, it holds that

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) \\ &+ \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T \tilde{\lambda}^k \\ -B^T \tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} - \begin{pmatrix} G(x^k - \tilde{x}^k) \\ \beta B^T B(y^k - \tilde{y}^k) \\ -B(y^k - \tilde{y}^k) + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \right\} \geq 0 \end{aligned}$$

for any $w = (x, y, \lambda) \in \Omega$. Recall the definition of Q in (3.1). The assertion (3.4) is thus derived. $\square$

Hence, the discrepancy between the point $\tilde{w}^k$ and a solution point of (2.1) is measured by the term $(w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k)$. In other words, if $Q(w^k - \tilde{w}^k) = 0$, then $\tilde{w}^k$ is a solution of (2.1).

According to its definition in (3.1), H is symmetric and positive semidefinite. Thus, we use the notation

$$\|w - \tilde{w}\|_H := \big((w - \tilde{w})^T H(w - \tilde{w})\big)^{1/2}.$$

In addition, since $Q = HM$ and F is monotone, (3.4) can be rewritten as

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq (w - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) \quad \forall\, w \in \Omega. \tag{3.10}$$

Now, we deal with the right-hand side of (3.10), and we want to find a uniform lower bound in terms of $\|w - w^k\|_H^2$ and $\|w - w^{k+1}\|_H^2$ for all $w \in \Omega$. This is realized in the following lemma.

LEMMA 3.2. *Let* $\{\tilde{w}^k\}$ *be defined by* (3.2), *and the matrices* M *and* H *be given in* (3.1). *Then we have*

$$(w - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) + \frac{1}{2}\big(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2\big) \geq 0 \quad \forall w \in \Omega. \tag{3.11}$$

Proof. First, by using $M(w^k - \tilde{w}^k) = (w^k - w^{k+1})$ (see (3.3)), it follows that

$$(w - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) = (w - \tilde{w}^k)^T H(w^k - w^{k+1}).$$

Therefore, in order to obtain (3.11) we need only to prove that

$$(w - \tilde{w}^k)^T H(w^k - w^{k+1}) + \frac{1}{2}\big(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2\big) \geq 0 \quad \forall w \in \Omega. \tag{3.12}$$

Applying the identity

$$(a - b)^T H(c - d) = \frac{1}{2}\big(\|a - d\|_H^2 - \|a - c\|_H^2\big) + \frac{1}{2}\big(\|c - b\|_H^2 - \|d - b\|_H^2\big)$$

to the term $(w - \tilde{w}^k)^T H(w^k - w^{k+1})$, we thus obtain

$$(w-\tilde{w}^k)^T H(w^k-w^{k+1}) = \frac{1}{2}(\|w-w^{k+1}\|_H^2-\|w-w^k\|_H^2)+\frac{1}{2}(\|w^k-\tilde{w}^k\|_H^2-\|w^{k+1}-\tilde{w}^k\|_H^2). \tag{3.13}$$

On the other hand, it follows from (3.3) that

$$\begin{aligned} &\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2 \\ &= \|w^k - \tilde{w}^k\|_H^2 - \|(w^k - \tilde{w}^k) - (w^k - w^{k+1})\|_H^2 \\ &= \|w^k - \tilde{w}^k\|_H^2 - \|(w^k - \tilde{w}^k) - M(w^k - \tilde{w}^k)\|_H^2 \\ &= (w^k - \tilde{w}^k)^T(2HM - M^T HM)(w^k - \tilde{w}^k). \end{aligned} \tag{3.14}$$

In fact, using the notation of H, M, and Q and recalling $Q = HM$ (see (3.1)), we have

$$2HM - M^T HM = 2Q - M^T Q = \begin{pmatrix} G & 0 & 0 \\ 0 & 0 & B^T \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

Therefore, it holds that

$$(w^k - \tilde{w}^k)^T(2HM - M^T HM)(w^k - \tilde{w}^k) = (x^k - \tilde{x}^k)^T G(x^k - \tilde{x}^k) + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \geq 0.$$

With this fact, the identity (3.14) becomes

$$\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2 \geq 0.$$

Substituting this into (3.13), we get (3.12), and the lemma is proved. ∎

4. The main result. Now, we are ready to present the $O(1/n)$ convergence rate for the ADM (1.6).

THEOREM 4.1. *Let $\{w^k\}$ be the sequence generated by the ADM* (1.6)*, and H be given in* (3.1)*. For any integer number $n > 0$, let $\tilde{w}_n$ be defined by*

$$\tilde{w}_n = \frac{1}{n+1}\sum_{k=0}^{n} \tilde{w}^k, \tag{4.1}$$

where $\tilde{w}^k$ is defined in (3.2)*. Then $\tilde{w}_n \in \Omega$ and*

$$\theta(\tilde{u}_n) - \theta(u) + (\tilde{w}_n - w)^T F(w) \leq \frac{1}{2(n+1)}\|w - w^0\|_H^2 \quad \forall w \in \Omega. \tag{4.2}$$

Proof. First, because of (3.2) and $w^k \in \Omega$, it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Thus, together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (4.1) implies that $\tilde{w}_n \in \Omega$. Second, the inequalities (3.10) and (3.11) imply that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|w - w^k\|_H^2 \geq \frac{1}{2}\|w - w^{k+1}\|_H^2 \quad \forall w \in \Omega. \tag{4.3}$$

Summing the inequality (4.3) over $k = 0, 1, \ldots, n$, we obtain

$$(n+1)\theta(u) - \sum_{k=0}^{n}\theta(\tilde{u}^k) + \left((n+1)w - \sum_{k=0}^{n}\tilde{w}^k\right)^T F(w) + \frac{1}{2}\|w - w^0\|_H^2 \geq 0 \quad \forall w \in \Omega.$$

Recall that $\tilde{w}_n$ is given in (4.1). We thus have

$$\frac{1}{n+1}\sum_{k=0}^{n}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_n - w)^T F(w) \leq \frac{1}{2(n+1)}\|w - w^0\|_H^2 \quad \forall w \in \Omega. \tag{4.4}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_n = \frac{1}{n+1}\sum_{k=0}^{n}\tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_n) \leq \frac{1}{n+1}\sum_{k=0}^{n}\theta(\tilde{u}^k).$$

Substituting this into (4.4), the assertion (4.2) follows directly. $\square$

For a given compact set $\mathcal{D} \subset \Omega$, let $d = \sup\{\|w - w^0\|_H \mid w \in \mathcal{D}\}$, where $w^0 = (x^0, y^0, \lambda^0)$ is the initial iterate. Then, after n iterations of the ADM (1.6), the point $\tilde{w}_n \in \Omega$ defined in (4.1) satisfies

$$\sup_{w \in \mathcal{D}}\left\{\theta(\tilde{u}_n) - \theta(u) + (\tilde{w}_n - w)^T F(w)\right\} \leq \frac{d^2}{2(n+1)},$$

which means that $\tilde{w}_n$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with the accuracy $O(1/n)$ (recall (2.4)). That is, the convergence rate $O(1/n)$ of the ADM (1.6) in the worst case is established in an ergodic sense, where n denotes the iteration number.

Acknowledgment. We are grateful to the anonymous referees for their valuable comments, which have helped us improve the presentation of this paper substantially.

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] R. H. Chan, J. Yang, and X. Yuan, *Alternating direction method for image inpainting in wavelet domains*, SIAM J. Imaging Sci., 4 (2011), pp. 807–826.

[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, Stanford, CA, 1978.

[4] J. Douglas and H. H. Rachford, *On the numerical solution of the heat conduction problem in* 2 *and* 3 *space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[5] 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), pp. 293–318.

[6] E. Esser, *Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman*, UCLA CAM Report 09-31, Department of Mathematics, UCLA, Los Angeles, CA, 2009.

[7] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I. Springer Ser. Oper. Res., Springer-Verlag, New York, 2003.

[8] M. Fukushima, *Application of the alternating directions method of multipliers to separable convex programming problems*, Comput. Optim. Appl., 2 (1992), pp. 93–111.

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

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

[11] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[12] R. Glowinski, *Finite element methods for incompressible viscous flow*, in Handbook of Numerical Analysis, Vol. IX, P. G. Ciarlet and J. L. Lions, eds., North–Holland, Amsterdam, 2003, pp. 3–1176.

[13] R. Glowinski, T. Kärkkäinen, and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scientific Computing, Variational Problems and Applications, Y. Kuznetsov, P. Neittanmaki, and O. Pironneau, eds., Barcelona, 2003.

[14] R. Glowinski and A. Marrocco, *Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires*, RAIRO, R2 (1975), pp. 41–76.

[15] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator Splitting Methods in Nonlinear Mechanics*, SIAM Stud. Appl. Math. 9, SIAM, Philadelphia, 1989.

[16] 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), pp. 103–118.

[17] B. He, M. Xu, and X. Yuan, *Solving large-scale least squares semidefinite programming by alternating direction methods*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 136–152.

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

[19] R. B. Kellogg, *Nonlinear alternating direction algorithm*, Math. Comp., 23 (1969), pp. 23–38.

[20] S. Kontogiorgis and R. R. Meyer, *A variable-penalty alternating directions method for convex optimization*, Math. Program., 83 (1998), pp. 29–53.

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

[22] G. I. Marchuk, *Splitting and alternating direction methods*, in Handbook of Numerical Analysis, Vol. I, P. G. Ciarlet and J. L. Lions, eds., North–Holland, Amsterdam, 1990, pp. 197–462.

[23] R. D. C. Monteiro and B. F. Svaiter, *Iteration-Complexity of Block-Decomposition Algorithms and the Alternating Direction Method of Multipliers*, manuscript, School of Industrial and Systems Engineering, The Georgia Institute of Technology, Atlanta, GA, 2010.

[24] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim., 15 (2004), pp. 229–251.

[25] Y. E. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[26] M. K. Ng, P. Weiss, and X. Yuan, *Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods*, SIAM J. Sci. Comput., 32 (2010), pp. 2710–2736.
[27] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., 72 (1979), pp. 383–390.
[28] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.
[29] J. Sun and S. Zhang, *A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs*, European J. Oper. Res., 207 (2010), pp. 1210–1220.
[30] M. Tao and X. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57–81.
[31] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.
[32] P. Tseng, *On Accelerated Proximal Gradient Methods for Convex-Concave Optimization*, manuscript, University of Washington, WA, 2008.
[33] S. Zhang, J. Ang, and J. Sun, *An alternating direction method for solving convex nonlinear semidefinite programming problem*, Optim., to appear.
[34] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276.
[35] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2010), pp. 20–46.