On the $O(1/t)$ convergence rate of alternating direction method

Bingsheng He$^1$ and Xiaoming Yuan$^2$

$^1$Department of Mathematics, Nanjing University, Nanjing, 210093, China
$^2$Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

First version, June 6, 2011
Second version: September 10, 2011

Abstract. The old alternating direction method (ADM) has found many new applications recently, and its empirical efficiency has been well illustrated in various fields. However, the estimate of ADM’s convergence rate remains a theoretical challenge for a few decades. In this note, we provide a uniform proof to show the $O(1/t)$ convergence rate for both the original ADM and its linearized variant (known as the split inexact Uzawa method in image processing literature). The proof is based on a variational inequality approach which is novel in the literature, and it is very simple.

Keywords. Alternating direction method, convergence rate, split inexact Uzawa method, variational inequalities, convex programming.

1 Introduction

The alternating direction method (ADM) was proposed in [11], and it was immediately promoted in the community of partial differential equations, see. e.g. [3, 10, 12, 14] for earlier literature. The last two decades have witnessed impressive development on ADM in the areas of variational inequalities and convex programming, see [5, 9, 16, 18, 26] to mention just a few. Very recently, the ADM has received explosively increasing interests again because of its efficient applications in a broad spectrum of areas such as imaging processing, statistical learning and engineering, see e.g. [1, 2, 4, 6, 17, 22, 24, 25, 28, 29]. In particular, we refer to [6, 7, 23] for the relationship between ADM and the split Bregman iteration scheme which was developed in [15] and is very influential in the area of image processing. As mentioned in [1], the ADM 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”.

However, the exciting literature is vastly devoted to the numerical and application perspectives of ADM, while the research on ADM’s theoretical aspects is still in infancy. Despite the elegant convergence analysis in earlier literature (e.g. [11, 14]), some extensions to inexact versions of ADM (e.g. [16]) and some results for other relevant methods (e.g. [13]), the estimate of ADM’s convergence rate remains a theoretical challenge since its original presence in [11]. On the other hand, numerical results reported in the literature often reveal fast convergence of ADM, and existing generic convergence results without accurate estimate of convergence rate seem inadequate for explaining ADM’s empirical efficiency. In this note, our purpose is to provide a simple proof on the $O(1/t)$ convergence rate.
convergence rate for ADM. This result thus fills the gap of ADM between the witness of empirical efficiency and the lack of matched theoretical results.

We concentrate our discussion on the context of linearly constrained convex programming with separable structure

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

where \(A \in \mathbb{R}^{m \times n_1}, \ B \in \mathbb{R}^{m \times n_2}, \ b \in \mathbb{R}^m, \ \mathcal{X} \subset \mathbb{R}^{n_1} \) and \(\mathcal{Y} \subset \mathbb{R}^{n_2}\) are closed convex sets, \(\theta_1 : \mathbb{R}^{n_1} \to \mathbb{R}\) and \(\theta_2 : \mathbb{R}^{n_2} \to \mathbb{R}\) are convex functions (not necessarily smooth). We assume the solution set of (1.1) to be nonempty. The iterative scheme of ADM for (1.1) is

\[
x^{k+1} = \arg\min\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k - b\|_2^2 \mid x \in \mathcal{X}\}, \tag{1.2a}
\]

\[
y^{k+1} = \arg\min\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|_2^2 \mid y \in \mathcal{Y}\}, \tag{1.2b}
\]

\[
\lambda^{k+1} = \lambda^k - \beta(\lambda^{k+1} - b), \tag{1.2c}
\]

where \(\lambda^k \in \mathbb{R}^m\) is the Lagrange multiplier and \(\beta > 0\) is a penalty parameter. According to (1.2), ADM blends the ideas of decomposition and Gauss-Seidel iterations, and makes it possible to exploit the properties of \(\theta_1\) and \(\theta_2\) individually. In fact, for many applications arising in image processing and statistical learning, the decomposed ADM subproblems are often simple enough to have closed-form solutions or can be easily solved up to high precisions. For these cases, ADM is particularly efficient, and this is the main reason inspiring recent burst of ADM’s wide applications in various areas.

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 pursuit approximate solutions of these subproblems. Thus, customized strategies with respect to particular properties of \(\theta_1\) and \(\theta_2\) are critical to ensure the efficiency of ADM for these cases. A success in this regard is the split inexact Uzawa method proposed in [30, 31]. Under the assumption that the resolvent operator of \(\theta_1\) defined by

\[
(I + \frac{1}{r} \partial \theta_1)^{-1}(a) = \text{Argmin}\{\theta_1(x) + \frac{r}{2}\|x - a\|_2^2 \mid x \in \mathcal{X}\}, \tag{1.3}
\]

has a closed-form representation for any given \(a \in \mathbb{R}^{n_1}\) and \(r > 0\) (a popular instance is \(\theta_1(x) = \|x\|_1\)), authors of [30, 31] suggested to linearize the quadratic term in (1.2a) and solve the following approximate problem

\[
x^{k+1} = \arg\min\{\theta_1(x) + \langle x - x^k, \beta A^T(Ax + By^k - b - \frac{1}{\beta} \lambda^k) \rangle + \frac{r}{2}\|x - x^k\|_2^2 \mid x \in \mathcal{X}\}, \tag{1.4}
\]

where the requirement on \(r\) is \(r > \beta\|A^T A\|\). Obviously, under the mentioned assumption on the resolvent operator of \(\theta_1\), the linearized subproblem (1.3) is easy enough to have a closed-form solution. This linearization strategy was shown in [30, 31] to be very efficient for some image restoration problems. Note that the linearization on (1.2b) was not discussed in [30, 31], as \(B\) is usually an identity matrix and \(\theta_2\) is usually a simple function such as the least-squares function for applications therein. For convenience, we also focus on the linearization merely on (1.2a).

Both (1.2a) and (1.4) can be treated uniformly by

\[
x^{k+1} = \arg\min\{\theta_1(x) + \frac{\beta}{2}\|(Ax + By^k - b) - \frac{1}{\beta} \lambda^k\|_2^2 + \frac{1}{2}\|x - x^k\|_G^2 \mid x \in \mathcal{X}\}, \tag{1.5}
\]
where \( G \in \mathbb{R}^{n_1 \times n_1} \) is a symmetric and positive semi-definite matrix (we denote \( \|x\|_G := \sqrt{x^T G x} \)). In fact, (1.2a) and (1.4) are recovered when \( G = 0 \) and \( G = (rI_{n_1} - \beta A^T A) \), respectively. Therefore, our analysis is for ADM in the following uniform context

\[
x^{k+1} = \arg \min \{ \theta_1(x) + \frac{\beta}{2} \|Ax + By^k - b\|_2^2 + \frac{1}{2} \|x - x^k\|_G^2 \mid x \in \mathcal{X} \}, \quad (1.6a)
\]

\[
y^{k+1} = \arg \min \{ \theta_2(y) + \frac{\beta}{2} \|(Ax^{k+1} + By - b) - 1\|_2^2 \mid y \in \mathcal{Y} \}, \quad (1.6b)
\]

\[
\lambda^{k+1} = \lambda^k - \beta (Ax^{k+1} + By^{k+1} - b). \quad (1.6c)
\]

Our investigation on the convergence rate of ADM is motivated by the encouraging achievement in estimating convergence rate or iteration complexity for various first-order algorithms in the literature (see, e.g. [19, 20, 21, 27] for a few). But, 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 [8], and this result enables us to find a very simple proof for the \( O(1/t) \) convergence rate of ADM.

\section{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 [8]. 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 \( w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^n \) such that

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

where

\[
u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix} \quad \text{and} \quad \theta(u) = \theta_1(x) + \theta_2(y). \quad (2.1b)
\]

Obviously, 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 \( \Omega^* \), is nonempty under the nonempty assumption on the solution set of (1.1).

Next, we specify Theorem 2.3.5 in [8] 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 next lemma is an incremental extension of Theorem 2.3.5 in [8]. But, we include all the details for completeness.

\begin{theorem}
The solution set of \( \text{VI}(\Omega, F, \theta) \) is convex and it can be characterized as

\[
\Omega^* = \bigcap_{w \in \Omega} \{ \tilde{w} \in \Omega : \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(u) \geq 0 \}. \quad (2.2)
\]

\end{theorem}

\begin{proof}
Indeed, if \( \tilde{w} \in \Omega^* \), we have

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

\end{proof}
By using the monotonicity of $F$ on $\Omega$, this implies
\[
\theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0, \ \forall w \in \Omega.
\]
Thus, $\tilde{w}$ belongs to the right-hand set in (2.2). Conversely, suppose $\tilde{w}$ belongs to the latter set. Let $w \in \Omega$ be arbitrary. The vector
\[
\tilde{w} = \alpha \tilde{w} + (1 - \alpha)w
\]
belongs to $\Omega$ for all $\alpha \in (0, 1)$. Thus we have
\[
\theta(\tilde{u}) - \theta(\tilde{u}) + (\tilde{w} - \tilde{w})^T F(\tilde{w}) \geq 0. \tag{2.3}
\]
Because $\theta(\cdot)$ is convex, we have
\[
\theta(\bar{u}) \leq \alpha \theta(\tilde{u}) + (1 - \alpha)\theta(u).
\]
Substituting it in (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 prove the convexity of $\Omega^*$. 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
\[
\{ \tilde{w} \in \Omega : \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq 0 \}
\]
is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of VI($\Omega$, $F$, $\theta$) is convex. \qed

Theorem 2.1 thus implies that $\tilde{w} \in \Omega$ is an approximate solution of VI($\Omega$, $F$, $\theta$) with the accuracy $\epsilon > 0$ if it satisfies
\[
\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \leq \epsilon, \ \forall w \in \Omega. \tag{2.4}
\]
In the rest, we show that after $t$ iterations of the ADM (1.6), we can find $\tilde{w} \in \Omega$ such that (2.4) is satisfied with $\epsilon = O(1/t)$. The convergence rate $O(1/t)$ of the ADM (1.6) is thus established.

3 Some properties

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

First of all, to make the notation of proof more succinct, we introduce some matrices
\[
M = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -\beta B & I_m \end{pmatrix}, \quad Q = \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad H = \begin{pmatrix} G & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.1}
\]
Obviously, we have $Q = HM$.

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

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

(3.2)

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

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

(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($\Omega, F, \theta$).

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

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

(3.4)

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

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

(3.5)

and

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

(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

$$
\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T\tilde{\lambda}^k + (r_{I_{u^1}} - \beta A^T)(\tilde{x}^k - x^k)\} \geq 0, \quad \forall x \in X,
$$

(3.7)

and

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

(3.8)

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

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

(3.9)

Combining (3.7), (3.8) and (3.9) together, 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

$$
\theta(u) - \theta(\tilde{w}^k) + \begin{pmatrix}
x - \tilde{x}^k \\
y - \tilde{y}^k \\
\lambda - \tilde{\lambda}^k
\end{pmatrix}^T \begin{pmatrix}
-A^T\tilde{\lambda}^k \\
-B\tilde{\lambda}^k \\
A\tilde{x}^k + By^k - b
\end{pmatrix} - \begin{pmatrix}
G(x^k - \tilde{x}^k) \\
\beta B^TB(\tilde{y}^k - y^k) \\
-B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)
\end{pmatrix} \geq 0.
$$

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

Hence, the discrepancy between the point $\tilde{w}^k$ and a solution point of (2.1) is measured by the term $(w^k - \tilde{w}^k)^T Q(w^k - \tilde{w}^k)$. 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 semi-definite. Thus, we use the notation
\[ \|w - \tilde{w}\|_H := ((w - \tilde{w})^T H (w - \tilde{w}))^{1/2}. \]
In addition, since $Q = HM$, (3.4) can be rewritten as
\[ \theta(u) - \theta(\tilde{w})^k + (w - \tilde{w})^T F(w) \geq (w - \tilde{w})^T HM(w^k - \tilde{w}^k), \quad \forall w \in \Omega. \tag{3.10} \]

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), the matrices $M$ and $H$ be given in (3.1). Then we have
\[ (w - \tilde{w})^T HM(w^k - \tilde{w})^k + \frac{1}{2}(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2) \geq 0, \quad \forall w \in \Omega. \tag{3.11} \]

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

Therefore, in order to obtain (3.11) we need only to prove that
\[ (w - \tilde{w})^T H(w^k - w^{k+1}) + \frac{1}{2}(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2) \geq 0, \quad \forall w \in \Omega. \tag{3.12} \]

Applying the identity
\[ -a^T H b = \frac{1}{2}(\|a - b\|_H^2 - \|a\|_H^2 - \|b\|_H^2), \]
we thus get
\[ (w^{k+1} - \tilde{w}^k)^T H(w^k - w^{k+1}) = \frac{1}{2}\|w^k - \tilde{w}^k\|_H^2 - \frac{1}{2}\|w^{k+1} - \tilde{w}^k\|_H^2 - \frac{1}{2}\|w^k - w^{k+1}\|_H^2. \tag{3.13} \]

Similarly, due to the identity
\[ a^T H b = \frac{1}{2}\{\|a\|_H^2 + \|b\|_H^2 - \|a - b\|_H^2\}, \]
we have that
\[ (w - w^{k+1})^T H(w^k - w^{k+1}) = \frac{1}{2}(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2) + \frac{1}{2}\|w^k - w^{k+1}\|_H^2. \tag{3.14} \]

Adding (3.13) and (3.14), we obtain
\[ (w - \tilde{w})^T H(w^k - w^{k+1}) + \frac{1}{2}(\|w - w^k\|_H^2 - \|w - w^{k+1}\|_H^2) = \frac{1}{2}(\|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2). \tag{3.15} \]

On the other hand, recall (1.6) and (3.2). We then get
\[ \|w^{k+1} - \tilde{w}^k\|_H^2 = \frac{1}{\beta}\|\lambda^{k+1} - \tilde{\lambda}\|_2^2 = \frac{1}{\beta}\|\beta B(y^k - \tilde{y}^k)\|_2^2 = \beta\|B(y^k - \tilde{y}^k)\|_2^2, \]
and thus
\[ \|w^k - \tilde{w}^k\|_H^2 - \|w^{k+1} - \tilde{w}^k\|_H^2 = \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}\|_2^2. \]

Substituting it in (3.15), we obtain (3.11) and the lemma is proved. \[ \square \]
4 The main result

Now, we are ready to present the $O(1/t)$ 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 (3.7). For any integer number $t > 0$, let $\tilde{w}_t$ be defined by

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k,$$

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

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

**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 convexity of $X$ and $Y$, (4.1) implies that $\tilde{w}_t \in \Omega$. Second, the inequalities (3.10) and (3.11) imply that

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

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

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

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

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

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{u}^k,$$

we have that

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

Substituting it in (4.4), the assertion (4.2) follows directly. \qed

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

$$\sup_{w \in D} \{\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\} \leq \frac{d^2}{2(t+1)},$$

which means $\tilde{w}_t$ is an approximate solution of VI($\Omega, F, \theta$) with the accuracy $O(1/t)$ (recall (2.4)). That is, the convergence rate $O(1/t)$ of the ADM (1.6) is established in an ergodic sense.
References


