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

Bingsheng He
Department of Mathematics
Nanjing University
hebma@nju.edu.cn

The context of this lecture is based on the articles [13, 15]
1  Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

where \( \theta_1(x) : \mathbb{R}^{n_1} \to \mathbb{R}, \ \theta_2(y) : \mathbb{R}^{n_2} \to \mathbb{R} \) are convex functions (but not necessary smooth), \( A \in \mathbb{R}^{m \times n_1}, \ B \in \mathbb{R}^{m \times n_2} \) and \( b \in \mathbb{R}^m, \ X \subset \mathbb{R}^{n_1}, \ Y \subset \mathbb{R}^{n_2} \) are given closed convex sets. Let \( n = n_1 + n_2 \).

The task of solving the problem (1.1) is to find an \((x^*, y^*, \lambda^*) \in \Omega\), such that

\[
\begin{cases}
\theta_1(x) - \theta_1(x^*) + (x - x^*)^T (-A^T \lambda^*) \geq 0, \\
\theta_2(y) - \theta_2(y^*) + (y - y^*)^T (-B^T \lambda^*) \geq 0, \quad \forall \ (x, y, \lambda) \in \Omega, \\
(\lambda - \lambda^*)^T (Ax^* + By^* - b) \geq 0,
\end{cases}
\]

where \( \Omega = X \times Y \times \mathbb{R}^m \).
By denoting

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

and

\[ \theta(u) = \theta_1(x) + \theta_2(y), \]

the first order optimal condition (1.2) can be written in a compact form such as

\[
\text{VI}(\Omega, F, \theta) \quad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \quad (1.3)
\]

Note that the mapping \( F \) is monotone. We use \( \Omega^* \) to denote the solution set of the variational inequality (1.3). We define some matrices which will greatly simplify the
notations in our analysis. More specifically, let

\[ H_0 = \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad Q_0 = \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad (1.4) \]

\[ H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad (1.5) \]

and

\[ M = \begin{pmatrix} I_{n_2} & 0 \\ -\beta B & I_m \end{pmatrix}. \quad (1.6) \]

For the matrices \( H, Q \) and \( M \) be defined in (1.5) and (1.6), we have

\[ Q = HM, \quad (1.7) \]
In addition, because

\[(Q^T + Q) - M^T H M = M^T H + H M - M^T H M\]

\[= H - (M^T - I) H (M - I)\]

\[= \begin{pmatrix}
\beta B^T B & 0 \\
0 & \frac{1}{\beta} I_m
\end{pmatrix}\]

\[- \begin{pmatrix}
0 & -\beta B^T \\
0 & 0
\end{pmatrix}\begin{pmatrix}
\beta B^T B & 0 \\
0 & \frac{1}{\beta} I_m
\end{pmatrix}\begin{pmatrix}
0 & 0 \\
-\beta B & 0
\end{pmatrix}\]

\[= \begin{pmatrix}
0 & 0 \\
0 & \frac{1}{\beta} I_m
\end{pmatrix} \succeq 0.
\]

we have

\[(Q^T + Q) - M^T H M \succeq 0. \tag{1.8}\]

These relations will simplify the convergence rate proofs in this lecture.
2 Alternating Direction Method

Applied ADM to the structure VI \((y^k, \lambda^k) \Rightarrow (y^{k+1}, \lambda^{k+1})\)

1. First, for given \((y^k, \lambda^k)\), \(x^{k+1}\) is the solution of the following problem

\[
x^{k+1} = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2} \|Ax + By^k - b\| - \frac{1}{\beta} \lambda^k \|x\|^2 | x \in \mathcal{X}\}
\]  \hspace{1cm} (2.1a)

2. Use \(\lambda^k\) and the obtained \(x^{k+1}\), \(y^{k+1}\) is the solution of the following problem

\[
y^{k+1} = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2} \|Ax^{k+1} + By - b\| - \frac{1}{\beta} \lambda^k \|y\|^2 | y \in \mathcal{Y}\}
\]  \hspace{1cm} (2.1b)

3. Update the

\[
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b).
\]  \hspace{1cm} (2.1c)

As mentioned in [1], the variable \(x\) is an intermediate variable during the ADM iterations since it essentially requires only \((y^k, \lambda^k)\) to generate the \((k + 1)\)-th iterate. For this reason, in the following analysis, we sometimes use the notations \(v^k = (y^k, \lambda^k)\) and
\( \mathcal{V} = \mathcal{V} \times \mathbb{R}^m \), and we let

\[
\mathcal{V}^* := \{ v^* = (y^*, \lambda^*) \mid w^* = (x^*, y^*, \lambda^*) \in \Omega^* \}.
\]

**Lemma 2.1** Let \( w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega \) be generated by (2.1) from the given \( v^k = (y^k, \lambda^k) \). Then, we have

\[
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \left\{ F(w^{k+1}) + \eta(y^k, y^{k+1}) + H_0(v^{k+1} - v^k) \right\} \geq 0, \ \forall w \in \Omega.
\]  
(2.2)

where

\[
\eta(y^k, y^{k+1}) = \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1})
\]  
(2.3)

and the matrix \( H_0 \) is defined in (1.4).

**Proof.** Note that the solution of (2.1a) satisfies

\[
x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \left\{ -A^T \lambda^k + \beta A^T (Ax^{k+1} + By^k - b) \right\} \geq 0, \ \forall x \in \mathcal{X}.
\]  
(2.4a)
And the solution of (2.1b) satisfies

\[ y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \left\{ -B^T \lambda^k + \beta B^T (Ax^{k+1} + By^{k+1} - b) \right\} \geq 0, \forall y \in \mathcal{Y}. \tag{2.4b} \]

Substituting \( \lambda^{k+1} \) (see (2.1c)) in (2.4) (eliminating \( \lambda^k \) in (2.4)), we get

\[ x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \left\{ -A^T \lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) \right\} \geq 0, \forall x \in \mathcal{X}. \tag{2.5a} \]

and

\[ y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \left\{ -B^T \lambda^{k+1} \right\} \geq 0, \forall y \in \mathcal{Y}. \tag{2.5b} \]
For analysis convenience, we rewrite (2.5) as \( u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y} \).

\[
\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T \lambda^{k+1} \\ -B^T \lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \\
+ \begin{pmatrix} 0 & 0 \\ 0 & \beta B^TB \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.
\]

Combining the last inequality with (2.1c), we have \( w^{k+1} \in \Omega \) and

\[
\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T \lambda^{k+1} \\ -B^T \lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix} \\
+ \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ \beta B^TB & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \geq 0, \quad (2.6)
\]

for any \( w \in \Omega \). The last inequality can be written as a compact form of (2.2). \( \square \)
3 Contractive property of ADM

Contractive property means that the sequence \(\{\|v^k - v^*\|_H^2\}\) in ADM is monotonically deceasing. Based on the analysis in the last section, we have the following lemma.

**Lemma 3.1** Let \(w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega\) be generated by (2.1) from the given \(v^k = (y^k, \lambda^k)\). Then, we have

\[
(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}),
\]

and the matrix \(H\) is defined in (1.5) and \(\eta(y^k, y^{k+1})\) is defined in (2.3).

**Proof.** Setting \(w = w^*\) in (2.2), and using the structures of \(H_0\) and \(H\), we get

\[
(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) + (\theta(u^{k+1}) - \theta(u^*)) + (w^{k+1} - w^*)^T F(w^{k+1}).
\]

Since \(F\) is monotone, it follows that

\[
\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0.
\]
The last inequality is due to $w^{k+1} \in \Omega$ and $w^* \in \Omega^*$ (see (1.3)). Substituting it in (3.2), the lemma is proved. \qed

**Lemma 3.2** Let $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ be generated by (2.1) from the given $v^k = (y^k, \lambda^k)$ and the vector $\eta(y^k, y^{k+1})$ be defined in (2.3). Then, we have

$$
(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) = (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}),
$$

(3.3)

and

$$
(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0.
$$

(3.4)

**Proof.** By using $\eta(y^k, y^{k+1})$ (see (2.3)), $Ax^* + By^* = b$ and (2.1c), we have

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

Because (2.5b) is true for the $k$-th iteration and the previous iteration, we have

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

(3.5)
and
\[ \theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \forall y \in \mathcal{Y}, \quad (3.6) \]

Setting \( y = y^k \) in (3.5) and \( y = y^{k+1} \) in (3.6), respectively, and then adding the two resulting inequalities, we get
\[
(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0.
\]

The assertions of this lemma are proved. \( \square \)

**Lemma 3.3** Let \( w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega \) be generated by (2.1) from the given \( v^k = (y^k, \lambda^k) \). Then, we have
\[
(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq 0, \forall v^* \in \mathcal{V}^*.
\]

**Proof.** The assertion follows (3.1), (3.3) and (3.4) directly. \( \square \)

Even though \( H \) is positive semi-definite (see (1.6) when \( B \) is not full column rank), in this lecture we use \( \|v - \tilde{v}\|_H \) to denote that
\[
\|v - \tilde{v}\|_H^2 = (v - \tilde{v})^T H (v - \tilde{v}) = \beta \|B(y - \tilde{y})\|^2 + \frac{1}{\beta} \|\lambda - \tilde{\lambda}\|^2.
\]
Based on the above mentioned lemmas, the contractive property of the sequence \( \{\|v^k - v^*\|_H\} \) follows directly.

**Theorem 3.1** Let \( w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega \) be generated by (2.1) from the given \( v^k = (y^k, \lambda^k) \). Then, we have

\[
\|v^{k+1} - v^*\|^2_H \leq \|v^k - v^*\|^2_H - \|v^k - v^{k+1}\|^2_H, \ \forall \ v^* \in V^*. 
\]  

(3.8)

**Proof.** By using (3.7), we have

\[
\|v^k - v^*\|^2_H = \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|^2_H \\
= \|v^{k+1} - v^*\|^2_H + 2(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \\
+ \|v^k - v^{k+1}\|^2_H \\
\geq \|v^{k+1} - v^*\|^2_H + \|v^k - v^{k+1}\|^2_H,
\]

and thus (3.8) is proved. \( \square \)

The inequality (3.8) in Theorem 3.1 demonstrates the contractive property of the ADM.
4 Defining an associated sequence $\{\tilde{w}^k\}$

For the convergence rate proof we preferably define an associated sequence $\{\tilde{w}^k\}$ by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix},$$  \hspace{1cm} (4.1)

where $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is generated by the ADM (2.1). Note that only the $\lambda$-part of $\tilde{w}^k$ and $w^{k+1}$ is different. By using (4.1) and (2.1c), we obtain the following useful relationship

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_2} & 0 \\ -\beta B & I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix},$$  \hspace{1cm} (4.2)

which can be rewritten into a compact form:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k),$$  \hspace{1cm} (4.3)

where the matrix $M$ is defined in (1.6). These notations simplify our presentation much.
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 (1.3).

**Lemma 4.1** Let \( \{v^k\} \) be the sequence generated by (2.1) and the associated sequence \( \{\tilde{w}^k\} \) be defined by (4.1). Then, we have

\[
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + Q_0(\tilde{v}^k - v^k)\} \geq 0, \quad \forall w \in \Omega, \quad (4.4)
\]

where \( Q_0 \) is defined in (1.4).

**Proof.** The assertion is followed from Lemma 2.1. To see this, we observe the terms

\[
F(w^{k+1}) + \eta(y^k, y^{k+1}) \quad \text{and} \quad H_0(v^{k+1} - v^k).
\]

in (2.2). Note that (4.1) implies

\[
x^{k+1} = \tilde{x}^k, \quad y^{k+1} = \tilde{y}^k, \quad (4.5a)
\]

and

\[
\lambda^{k+1} = \tilde{\lambda}^k + \beta B(y^k - y^{k+1}). \quad (4.5b)
\]
According to the above relations, we have
\[
\begin{pmatrix}
-A^T \lambda^{k+1} \\
-B^T \lambda^{k+1} \\
A x^{k+1} + B y^{k+1} - b
\end{pmatrix} + \beta \begin{pmatrix}
A^T \\
B^T \\
0
\end{pmatrix} \begin{pmatrix}
y^k - y^{k+1}
\end{pmatrix} = \begin{pmatrix}
-A^T \tilde{\lambda}^k \\
-B^T \tilde{\lambda}^k \\
A \tilde{x}^k + B \tilde{y}^k - b
\end{pmatrix},
\]
and thus
\[
F(w^{k+1}) + \eta(y^k, y^{k+1}) = F(\tilde{w}^k). \tag{4.6}
\]
Again, it follows from (4.5) that
\[
\frac{1}{\beta} (\lambda^{k+1} - \lambda^k) = -B(\tilde{y}^k - y^k) + \frac{1}{\beta} (\tilde{\lambda}^k - \lambda^k),
\]
and consequently
\[
\begin{pmatrix}
0 & 0 \\
\beta B^T B & 0 \\
0 & \frac{1}{\beta} I_m
\end{pmatrix} \begin{pmatrix}
y^{k+1} - y^k \\
\lambda^{k+1} - \lambda^k
\end{pmatrix} = \begin{pmatrix}
0 & 0 \\
\beta B^T B & 0 \\
0 & \frac{1}{\beta} I_m
\end{pmatrix} \begin{pmatrix}
\tilde{y}^k - y^k \\
\tilde{\lambda}^k - \lambda^k
\end{pmatrix}.
\]
By using the matrices $H_0$ and $Q_0$ (see (1.4)), the last equation can be written as means that
\[
H_0 (v^{k+1} - v^k) = Q_0 (\tilde{v}^k - v^k). \tag{4.7}
\]
Substituting (4.6) and (4.7) in (2.2), the assertion is proved.

5 Convergence in an ergodic sense

The reason inspiring us to investigate the convergence rate of ADM via the VI approach, rather than the conventional approaches based on the functional values in the literature, is mainly due to the Theorem 2.3.5 in [5] which provides an insightful characterization for the solution set of a generic VI. In the following theorem, we specific the Theorem 2.3.5 in [5] for the particular VI $\langle \Omega, F, \theta \rangle$ under consideration. The proof is an incremental extension of Theorem 2.3.5 in [5]. But, we include all the details for completeness.

**Theorem 5.1** The solution set of $\text{VI}(\Omega, F, \theta)$ is convex and it can be characterized as

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

(5.1)
Proof. Indeed, if $\bar{w} \in \Omega^*$, we have

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

By using the monotonicity of $F$ on $\Omega$, this implies

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

Thus, $\bar{w}$ belongs to the right-hand set in (5.1). Conversely, suppose $\bar{w}$ belongs to the latter set. Let $w \in \Omega$ be arbitrary. The vector

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

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

$$\theta(\bar{u}) - \theta(\bar{u}) + (\bar{w} - \bar{w})^T F(\bar{w}) \geq 0.$$ (5.2)

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

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

Substituting it in (5.2), we get

$$(\theta(u) - \theta(\bar{u})) + (w - \bar{w})^T F(\alpha \bar{w} + (1 - \alpha) w) \geq 0$$
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 $\text{VI}(\Omega, F, \theta)$ is convex. $\square$

Theorem 5.1 thus implies that $\tilde{w} \in \Omega$ is an approximate solution of $\text{VI}(\Omega, F, \theta)$ with the accuracy $\epsilon > 0$ if it satisfies

$$\theta(u) - \theta(\tilde{u}) + F(w)^T (w - \tilde{w}) \geq -\epsilon, \ \forall w \in \Omega_{\tilde{w}},$$

where

$$\Omega_{\tilde{w}} = \{w \in \Omega \mid \|w - \tilde{w}\| \leq 1\}.$$
In the rest, our purpose is to show that after \( t \) iterations of the ADM (2.1), we can find \( \tilde{w} \in \Omega \) such that

\[
\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \leq \epsilon, \quad \forall w \in \Omega_{\tilde{w}}
\]  

(5.3)

with \( \epsilon = O(1/t) \). The convergence rate \( O(1/t) \) of the ADM (2.1) is thus proved.

Since \( H \) is symmetric and positive semi-definite, for notational convenience we use

\[
\|v - \tilde{v}\|_H \quad \text{to denote}
\]

\[
\|v - \tilde{v}\|_H = \left( (v - \tilde{v})^T H (v - \tilde{v}) \right)^{1/2}.
\]

Together with the notations of \( H \) and \( Q \), it is trivial to verify that (4.4) can be written into

\[
(\theta(u) - \theta(\tilde{u}^k)) + (w - \tilde{w}^k)^T F(w) \geq (v - \tilde{v}^k)^T H M(v^k - \tilde{v}^k), \quad \forall w \in \Omega,
\]

(5.4)

and we omit the proof.

Now, we deal with the right-hand side of (5.4), and we want to find a lower bound of it in terms of \( \|v - v^k\|_H^2 \) and \( \|v - v^{k+1}\|_H^2 \). This is realized in the following lemma.
Lemma 5.1  Let \( \{v^k\} \) be the sequence generated by the ADM (2.1) and the associated sequence \( \{\tilde{w}^k\} \) be defined by (4.1). Then we have

\[
(v - \tilde{v}^k)^T H M (v^k - \tilde{v}^k) \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2), \quad \forall v \in \mathcal{V}.
\]  (5.5)

Proof. First, by using \( M(v^k - \tilde{v}^k) = (v^k - v^{k+1}) \) (see (4.3)), it follows that

\[
(v - \tilde{v}^k)^T H M (v^k - \tilde{v}^k) = (v - \tilde{v}^k)^T H (v^k - v^{k+1}).
\]

Therefore, in order to obtain (5.5) we need only to prove that

\[
(v - \tilde{v}^k)^T H (v^k - v^{k+1}) \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2), \quad \forall v \in \mathcal{V}.
\]  (5.6)

Applying the identity

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

with

\[
a = v, \quad b = \tilde{v}^k, \quad c = v^k \quad \text{and} \quad d = v^{k+1},
\]
we thus get
\[(v - \tilde{v}^k)^T H(v^k - v^{k+1})\]
\[= \frac{1}{2} \left( \|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2 \right) + \frac{1}{2} \left( \|v^k - \tilde{v}^k\|_H^2 - \frac{1}{2} \|v^{k+1} - \tilde{v}^k\|_H^2 \right).\]

To show (5.6), we need only to demonstrate that
\[\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \geq 0.\] (5.7)

Recall (2.1) and (4.1), we then get
\[\|v^{k+1} - \tilde{v}^k\|_H^2 = \frac{1}{\beta} \|\lambda^{k+1} - \tilde{\lambda}^k\|^2 = \frac{1}{\beta} \|\beta B(y^k - \tilde{y}^k)\|^2 = \beta \|B(y^k - \tilde{y}^k)\|^2.\]

On the other hand
\[\|v^k - \tilde{v}^k\|_H^2 = \beta \|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2.\]

Therefore, we have \[\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \frac{1}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2\] and (5.7) is true.

The assertion of this lemma is proved. \(\Box\)

Now, we are ready to present the main result regarding the convergence rate of the ADM (2.1), as the following theorem shows.
**Theorem 5.2** Let \( \{ v^k \} \) be the sequence generated by the ADM (2.1) and the associated sequence \( \{ \tilde{w}^k \} \) be defined by (4.1). 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 (4.1). Then, \( \tilde{w}_t \in \Omega \) and

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

(5.9)

where \( H \) is given by (1.5).

**Proof.** First, because of (4.1) and \( w^k \in \Omega \), it holds that \( \tilde{w}^k \in \Omega \) for all \( k \geq 0 \). Thus, together with convexity of \( \mathcal{X} \) and \( \mathcal{Y} \), (5.8) implies that \( \tilde{w}_t \in \Omega \). Second, the inequalities (5.4) and (5.5) imply that

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

(5.10)
Summing the inequality (5.10) over $k = 0, 1, \ldots, t$, we obtain
\[
(t + 1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \left((t + 1)w - \sum_{k=0}^{t} \tilde{w}^k\right)^T F(w) + \frac{1}{2} \|v - v^0\|_H^2 \geq 0,
\]
for any $w \in \Omega$. Use the notation of $\tilde{w}_t$, it can be written as
\[
\frac{1}{t + 1} \sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t + 1)} \|v - v^0\|_H^2, \ \forall w \in \Omega.
\]
(5.11)

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 (5.11), the assertion of this theorem follows directly. \qed

Recall (5.3). The conclusion (5.9) thus indicates obviously that the ADM (2.1) is able to generate an approximate solution (i.e., $\tilde{w}_t$) of (5.8) with the accuracy $O(1/t)$ after $t$
iterations. That is, the convergence rate $O(1/t)$ of the ADM (2.1) is established.

6 Convergence rate in the non-ergodic sense

This section shows that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ is monotonically non-increasing.

$$\|v^k - v^{k+1}\|^2_H \leq \|v^{k-1} - v^k\|^2_H, \quad \forall k \geq 1.$$  

(6.1)

Based on (3.8) and (6.1), we drive

$$\|v^k - v^{k+1}\|^2_H \leq \frac{1}{(k + 1)} \|v^0 - v^*\|^2_H, \quad \forall v^* \in \mathcal{V}^*.$$  

Since $\|v^k - v^{k+1}\|^2_H$ is viewed as the stopping criterium, we obtain the worst-case $O(1/t)$ convergence rate in a non-ergodic sense.

Again, we prove several lemmas for this purpose. First of all, we observe that $v^{k+1}$ and $\tilde{v}^k$ defined in (4.1) are related by (as pointed in (4.3))

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

(6.2)
where the matrix $M$ is given in (1.6).

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

**Lemma 6.1** Let $\{v^k\}$ be the sequence generated by (2.1), the associated sequence $\{\tilde{w}^k\}$ be defined by (4.1) and $Q_0$ be given in (1.4). Then, we have

$$
(\tilde{w}^k - \tilde{w}^{k+1})^T Q_0 \{ (v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1}) \} \geq 0. 
$$

(6.3)

**Proof.** Set $w = \tilde{w}^{k+1}$ in (4.4), we have

$$
\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T \{ F(\tilde{w}^k) + Q_0 (\tilde{v}^k - v^k) \} \geq 0. 
$$

(6.4)

Note that (4.4) is also true for $k := k + 1$ and thus we have

$$
\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T \{ F(\tilde{w}^{k+1}) + Q_0 (\tilde{v}^{k+1} - v^{k+1}) \} \geq 0, \quad \forall \ w \in \Omega.
$$

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

$$
\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T \{ F(\tilde{w}^{k+1}) + Q_0 (\tilde{v}^{k+1} - v^{k+1}) \} \geq 0. 
$$

(6.5)

Adding (6.4) and (6.5) and using the monotonicity of $F$, we get (6.3) immediately.  \qed
Lemma 6.2 Let \( \{v^k\} \) be the sequence generated by (2.1), the associated sequence \( \{\tilde{w}^k\} \) be defined by (4.1), the matrices \( H \), \( Q \) and \( M \) be given in (1.5) and (1.6). Then, we have

\[
(v^k - \tilde{v}^k)^T M^T H M \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\
\geq \frac{1}{2} \| (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \|_2^2 (Q^T + Q),
\]

where the matrix \( Q \) is defined in (1.5).

Proof. First, according to the structures of \( Q_0 \) and \( Q \), it follows from (6.3) that

\[
(\tilde{v}^k - \tilde{v}^{k+1})^T Q \{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0.
\]

Adding the term

\[
\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\}^T Q \{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\}
\]

to the both sides of (6.7), and using \( v^T Q v = \frac{1}{2} v^T (Q^T + Q) v \), we get

\[
(v^k - v^{k+1})^T Q \{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \\
\geq \frac{1}{2} \| (v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1}) \|_2^2 (Q^T + Q).
\]
Reorder \((v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\) to \((v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\), from the above inequality we get

\[
(v^k - v^{k+1})^T Q \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2} \| (v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1}) \|^2_{(Q^T + Q)}. 
\]

Substituting the term \((v^k - v^{k+1})\) into the left-hand side of the last inequality, and using the relationship in (6.2) and the fact \(Q = HM\) in (1.7), we obtain (6.6).

Finally, we are ready to show the assertion (6.1) in the following theorem.

**Theorem 6.1** Let \(\{v^k\}\) be the sequence generated by (2.1), the associated sequence \(\{\tilde{w}^k\}\) be defined by (4.1). Then we have

\[
\|M(v^{k+1} - \tilde{w}^{k+1})\|_H^2 \leq \|M(v^k - \tilde{w}^k)\|_H^2, 
\]

(6.8)

and thus

\[
\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2. 
\]

(6.9)
Proof. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity
\[
\|a\|_H^2 - \|b\|_H^2 = 2a^T H (a - b) - \|a - b\|_H^2,
\]
we obtain
\[
\|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2
= 2(v^k - \tilde{v}^k)^T M^T H M\left\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\right\}
- \|M\left\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\right\}\|_H^2.
\]
Inserting (6.6) into the first term of the right-hand side of the last equality, we obtain
\[
\|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2
\geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_H^2\left\{(Q^T + Q) - M^T H M\right\}
- \|M\left\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\right\}\|_H^2
= \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_H^2\left\{(Q^T + Q) - M^T H M\right\}.
\]
Because of the positive definiteness of the matrix $(Q^T + Q) - M^T H M$ (indicated in (1.8)). Recall the relationship in (6.2). The assertion (6.9) follows immediately from (6.8).
\[\square\]
With Theorems 3.1 and 6.1, we can prove a worst-case $O(1/t)$ convergence rate in a non-ergodic sense for the ADM scheme (2.1).

**Theorem 6.2** Let $\{v^k\}$ be the sequence generated by (2.1). Then we have

$$\|v^k - v^{k+1}\|_H^2 \leq \frac{1}{(k + 1)} \|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \quad (6.10)$$

**Proof.** First, it follows from (3.8) that

$$\sum_{t=0}^{\infty} \|v^t - v^{t+1}\|_H^2 \leq \|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \quad (6.11)$$

According to Theorem 6.1, the sequence $\{\|v^t - v^{t+1}\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(k + 1)\|v^k - v^{k+1}\|_H^2 \leq \sum_{i=0}^{k} \|v^i - v^{i+1}\|_H^2. \quad (6.12)$$

The assertion (6.10) follows from (6.11) and (6.12) immediately. \qed

Notice that $\mathcal{V}^*$ is convex and closed. Let $d := \inf \{\|v^0 - v^*\|_H \mid v^* \in \Omega^*\}$. Then, for
any given $\epsilon > 0$, Theorem 6.2 shows that the ADM scheme (2.1) needs at most $\left\lfloor \frac{d^2}{\epsilon} \right\rfloor$ iterations to ensure that $\|v^k - v^{k+1}\|_H^2 \leq \epsilon$. Recall that $v^{k+1}$ is a solution of $\text{VI}(\Omega, F, \theta)$ if $\|v^k - v^{k+1}\|_H^2 = 0$ (see Lemma 3.1). A worst-case $O(1/t)$ convergence rate in a non-ergodic sense for the ADM scheme (2.1) is thus established in Theorem 6.2.

References


Comments and Suggestions are welcome!