

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

Bingsheng He^{*1} and Xiaoming Yuan^{‡2}

^{*}Department of Mathematics, Nanjing University, Nanjing, 210093, China

[‡]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

First version: June 6, 2011

Second version: September 10, 2011

Third version: October 12, 2011

Fourth version: October 31, 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 unified 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, 27] 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, 23, 25, 26, 29, 30]. In particular, we refer to [6, 7, 24] 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”.

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

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

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, $\theta_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ and $\theta_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are convex functions (not necessarily smooth). We assume the solution set of (1.1)

¹Email: hebma@nju.edu.cn. This author was supported by the NSFC grant 10971095.

²Email: xmyuan@hkbu.edu.hk. This author was supported by a grant from Hong Kong General Research Fund.

to be nonempty. The iterative scheme of ADM for (1.1) is

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

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

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

where $\lambda^k \in \mathfrak{R}^m$ 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 θ_1 and θ_2 individually. This is the main reason inspiring recent burst of ADM's wide applications in various areas.

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. We refer to 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]). 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, the $O(1/t)$ convergence rate of ADM was shown in [19] in the more general context of finding roots of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable two-block structure. These additional assumptions for deriving the $O(1/t)$ convergence rate of ADM, however, exclude many interesting applications where the efficiency of ADM has been shown in the literature. In particular, the requirement on obtaining exact solutions of (1.2a) and (1.2b) is somehow too restricted to realize, even when the resolvent operators of θ_1 and θ_2 have closed-form representations. Here, the resolvent operator of a convex function, say θ_1 , is given by

$$\left(I + \frac{1}{r} \partial \theta_1 \right)^{-1}(a) = \arg \min \left\{ \theta_1(x) + \frac{r}{2} \|x - a\|^2 \mid x \in \mathcal{X} \right\}, \quad (1.3)$$

where $a \in \mathfrak{R}^{n_1}$ and $r > 0$. A simple instance is that when $\theta_1(x) = \|x\|_1$ (thus its resolvent operator 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 [19] is not valid for some aforementioned ADM's novel applications, including those in the area of total variation image restoration problems (e.g. [1, 2, 6, 23]).

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 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 [31, 32]. Under the assumption that the resolvent operator of θ_1 has a closed-form representation, authors of [31, 32] suggested to linearize the quadratic term in (1.2a) and solve the following approximate problem

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

where the requirement on r is $r > \beta \|A^T A\|$. Obviously, under the mentioned assumption on the resolvent operator of θ_1 , the linearized subproblem (1.4) is easy enough to have a closed-form solution.

This linearization strategy was shown in [31, 32] to be very efficient for some image restoration problems. Note that the linearization on (1.2b) was not discussed in [31, 32], 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 on (1.2a). To the best of our knowledge, the convergence rate of the split inexact Uzawa method [31, 32] is not yet established in the literature.

Our purpose is to provide a simple unified proof on the $O(1/t)$ convergence rate for both the original ADM (1.2) (in the general context) and its linearized variant (i.e., the split inexact Uzawa method proposed in [31, 32]). This result thus fills the gap of ADM between the witness of empirical efficiency and the lack of matched theoretical results. The conjecture in [19] 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 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. [20, 21, 22, 28] 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 and its linearized variant.

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

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

where $G \in \mathfrak{R}^{n_1 \times n_1}$ 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

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

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

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

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 [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 $w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathcal{Y} \times \mathfrak{R}^m$ 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 Ω^* , 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.

Theorem 2.1 *The solution set of $\text{VI}(\Omega, F, \theta)$ is convex and it can be characterized as*

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

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

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

By using the monotonicity of F on Ω , this implies

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

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

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

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

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

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 $\alpha \rightarrow 1$ 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 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

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

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 2.1 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

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

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

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

Obviously, we have $Q = HM$.

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

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

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

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

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*

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

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

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

and

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

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

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

and

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

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

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

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

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

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 semi-definite. Thus, we use the notation

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

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

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

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*

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

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

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

Applying the identity

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

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

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

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

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

In fact, using the notation of H , M and Q and recall $Q = HM$ (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

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

With this fact, the identity (3.14) shows

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

Substituting it into (3.13), we get (3.12) 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 (3.1). For any integer number $t > 0$, let \tilde{w}_t be defined by*

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^t \tilde{w}^k, \quad (4.1)$$

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

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

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

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

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

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

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

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

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. \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 t iterations of the ADM (1.6), the point $\tilde{w}_t \in \Omega$ defined in (4.1) satisfies

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

which means \tilde{w}_t is an approximate solution of $\text{VI}(\Omega, F, \theta)$ 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

- [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. F. Yang and X. M. Yuan, *Alternating direction method for image inpainting in wavelet domain*, 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 non-linear elliptic equations*, technical report, Stanford University, 1978.
- [4] C. H. Chen, B. S. He and X. M. Yuan, *Matrix completion via alternating direction method*, IMA J. Numer. Analysis, to appear.
- [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, 2009.
- [7] E. Esser, X. Q. Zhang, T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imag. Sci. 3(2010), pp. 1015-1046.
- [8] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vol. I. Springer Series in Operations Research, Springer Verlag, New York, 2003.
- [9] M. Fukushima, *Application of the alternating directions method of multipliers to separable convex programming problems*, Comput. Optim. Appl., 2 (1992), pp. 93-111.
- [10] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems*, Stud. Math. Appl. 15, NorthCHolland, Amsterdam, 1983.
- [11] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.
- [12] R. Glowinski, *Numerical methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.
- [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, edited by Y. Kuznetsov, P. Neittaanmaki and O. Pironneau, Barcelona, 2003.
- [14] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Non-linear Mechanics*, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.
- [15] T. Goldstein and S. Osher, *The split Bregman method for l_1 regularized problems*, SIAM J. Imag. Sci., 2(2) (2008), pp. 323-343.
- [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. S. He, M. H. Xu and X. M. Yuan, *Solving large-scale least squares covariance matrix problems by alternating direction methods*, SIAM J. Matrix Analy. Appl., 32(1) (2011), pp. 136-152.
- [18] S. Kontogiorgis and R. R. Meyer, *A variable-penalty alternating directions method for convex optimization*, Math. Program., 83 (1998), pp. 29-53.
- [19] R. D. C. Monteiro and B. F. Svaiter, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, manuscript, 2010.
- [20] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim. 15 (2005), pp. 229-251.
- [21] Y. E. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$* , Doklady AN SSSR, 269 (1983), pp. 543-547.
- [22] Y. E. Nesterov, *Smooth minimization of nonsmooth functions*, Math. Program., 103 (2005), pp.127-152.
- [23] M. K. Ng, P.A. Weiss and X.M. Yuan, *Solving constrained total-variation problems via alternating direction methods*, SIAM J. Sci. Comput., 32(5) (2010), pp. 2710-2736.
- [24] S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, Lecture Notes in Comput. Sci., 5567 (2009), pp.464–476.
- [25] 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.
- [26] M. Tao and X.M. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57-81.
- [27] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Con. Optim., 29 (1991), pp. 119-138.
- [28] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, manuscript, University of Washington, USA, 2008.
- [29] X. M. Yuan, *Alternating direction methods for covariance selection models*, J. of Sci. Computing, to appear.
- [30] S. Zhang, J. Ang and J. Sun, *An alternating direction method for solving convex nonlinear semidefinite programming problem*, Optimization, to appear.
- [31] X. Q. Zhang, M. Burger, X. Bresson and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imag. Sci., 3(3) (2010), pp. 253-276.
- [32] X. Q. Zhang, M. Burger and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Computing, 46(1)(2010), pp. 20-46.