THE DUAL STEP SIZE OF THE ALTERNATING DIRECTION
METHOD CAN BE LARGER THAN 1.618 WHEN ONE
FUNCTION IS STRONGLY CONVEX

FENG MA*, JIANSHENG SHU, YAXIONG LI AND JIAN WU

High-Tech Institute of Xi’an
Xi’an 710025, China

(Communicated by Changjun Yu)

ABSTRACT. The alternating direction method of multipliers (ADMM) is one
of the most well-known optimization scheme for solving linearly constrained
separable convex problem. In the literature, Fortin and Glowinski proved that
the step size for updating the Lagrange multiplier of the ADMM can be chosen
in the open interval of zero to the golden ratio. But, it is still unknown whether
the dual step size can be larger than the golden ratio. In this paper, for the case
where one function term is strongly convex and the associate coefficient matrix
is full column rank, we present an affirmative answer to the above question.
We then derive an exact relationship between the modulus and the dual step
size. Our analysis deepens the understanding of the convergence properties of
the ADMM.

1. Introduction. In this paper, we restrict our discussion to the following convex
programming

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

where \( A \in \mathbb{R}^{m \times n_1}, B \in \mathbb{R}^{m \times n_2}, b \in \mathbb{R}^m, X \subseteq \mathbb{R}^{n_1} \) and \( Y \subseteq \mathbb{R}^{n_2} \) are closed convex sets, \( \theta_1 : \mathbb{R}^{n_1} \to \mathbb{R} \) and \( \theta_2 : \mathbb{R}^{n_2} \to \mathbb{R} \) are proper, closed and convex (not necessarily smooth) functions. Throughout, we assume that the solution set of (1) is nonempty and the matrix \( B \) is full column rank. The model (1) captures a broad range of problems arising in image sciences, signal processing and machine learning, including Lasso estimation, total variation denoising, stable principal component pursuit, etc. We refer the readers to the monographs [1, 14] and the references therein for more examples.

In the literature, many first-order optimization methods have been proposed to solve (1). Among them, the alternating direction method of multipliers (ADMM), which was introduced by Glowinski and Marroco [10] and Gabay and Mercier [11], and subsequently studied, e.g., in [6, 7, 8, 24] has regained much attention in recent
years, we refer to the surveys [2, 13] for an overview on its recent developments. For problem (1), the ADMM iterates as follows:

\[
x^{k+1} \in \arg \min \{ \theta_1(x) - (\lambda^k)^T A x + \frac{\beta}{2} \| A x + B y^k - b \|^2 \mid x \in \mathcal{X} \},
\]

\[
y^{k+1} \in \arg \min \{ \theta_2(y) - (\lambda^k)^T B y + \frac{\beta}{2} \| A x^{k+1} + B y - b \|^2 \mid y \in \mathcal{Y} \},
\]

\[
\lambda^{k+1} = \lambda^k - \beta (A x^{k+1} + B y^{k+1} - b),
\]

where \( \beta > 0 \) is a penalty parameter for violation of the linear constraints. Note that the ADMM decomposes the minimization problem (1) into two smaller subproblems so that each of the subproblems only involves one function \( \theta_i(\cdot) \). In this case, the generated subproblems could be much easier, especially when \( \theta_1(\cdot) \) and \( \theta_2(\cdot) \) have some special structures to be employed. In addition, since the variables are treated individually, the computations of ADMM can be implemented in distributed systems. Note that the ADMM enjoys global convergence property, and it can be derived from its relation to the Douglas-Rachford operator splitting method; see [6]. Furthermore, the convergence rate of the ADMM can be found in e.g., [5, 18, 19, 20].

There are many practical variants of ADMM. One famous variant proposed by Fortin and Glowinski in [9] appears in many guises in the literature. This variant, which suggests to attach a relaxation factor to the Lagrange-multiplier-updating step in ADMM, reads as follows:

\[
x^{k+1} \in \arg \min \{ \theta_1(x) - (\lambda^k)^T A x + \frac{\beta}{2} \| A x + B y^k - b \|^2 \mid x \in \mathcal{X} \},
\]

\[
y^{k+1} \in \arg \min \{ \theta_2(y) - (\lambda^k)^T B y + \frac{\beta}{2} \| A x^{k+1} + B y - b \|^2 \mid y \in \mathcal{Y} \},
\]

\[
\lambda^{k+1} = \lambda^k - \gamma (A x^{k+1} + B y^{k+1} - b),
\]

where the relaxation factor \( \gamma \) is allowed to be in the interval \((0, \frac{1+\sqrt{5}}{2})\). Note that the ADMM variant differs from the ADMM (2) only in that the step size for updating the Lagrange-multiplier is in the range \((0, \frac{1+\sqrt{5}}{2})\). Thus, the scheme (3) enjoys the same easiness as ADMM. It is worthwhile to mention that although the two algorithms are very similar, the convergence study of the ADMM variant is very different from ADMM. Indeed, as remarked in [8]. They are “actually two distinct families of ADMM algorithms, one derived from the operator-splitting framework and the other derived from Lagrangian splitting”. Numerically, taking a larger value of \( \gamma \) is usually benefit to speeding up the performance; and it has been observed empirically in many cases that the upper bound of the step size, i.e., \( \gamma = 1.618 \), is helpful to get the best performance. We refer to e.g., [12, 21, 23] for some numerical demonstrations. Some other variants and extensions of ADMM, together with their convergence properties, can be seen in e.g., [3, 4, 5, 16, 17].

Remember that \( \gamma \in (0, \frac{1+\sqrt{5}}{2}) \) is only a sufficient condition to ensure the convergence of the ADMM and larger value of \( \gamma \) can accelerate its convergence. So it is reasonable to ask whether \( \gamma \) in (3) can be larger than \( \frac{1+\sqrt{5}}{2} \). Indeed, this issue stands as an open question by Glowinski raised in [12]. In [10], Gabay and Mercier proved that ADMM with \( \gamma \in (0, 2) \) is convergent if one of the two functions in the objective is linear. Recently, the authors [22] showed the convergence of the ADMM scheme (3) with \( \gamma \in (0, 2) \) when both the functions \( \theta_1 \) and \( \theta_2 \) in (1) are quadratic. In this work, we are concerned with the case where one function term
is strongly convex. We prove that the dual step size can be larger than 1.618 for this case. We also derive an exact relationship between the modulus of the strongly convex function and the dual step size. This result partially answered Glowinski’s open question and advanced the understanding of the convergence properties of the ADMM.

The paper is organized as follows: in Section 2, we present the assumptions for future discussion and summarize some necessary background; in Section 3, we prove the main convergence results. Finally, we conclude the work in Section 4.

2. Preliminaries. Throughout this paper, we assume that \( \theta_2(y) \) is strongly convex with the modulus \( \tau > 0 \). That is, there is a positive constant \( \tau \) such that for any \( \eta \in \partial \theta_2(y) \), we have

\[
\theta_2(\tilde{y}) - \theta_2(y) \geq \langle \tilde{y} - y, \eta \rangle + \frac{\tau}{2} \| \tilde{y} - y \|^2, \quad \forall y, \ \tilde{y} \in \mathcal{Y}.
\]

(4)

Unless otherwise stated, the results in this paper are given under this assumption.

Remark 1. Since \( \tau \) is positive and the matrix \( B \) is assumed to be full column rank, there exists a positive constant \( s \) such that \( \tau = s \cdot \rho(B^T B) \), where \( \rho(\cdot) \) denotes the spectral radius of a matrix.

2.1. Optimality condition. Let the Lagrangian function of problem (1) be

\[
L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b),
\]

and it is defined on \( \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m \). Since the problem (1) is well defined and the solution set is nonempty, there exists a saddle point \( (x^*, y^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m \), such that

\[
L_{\lambda \in \mathbb{R}^m}(x^*, y^*, \lambda) \leq L(x^*, y^*, \lambda^*) \leq L_{x \in \mathcal{X}, y \in \mathcal{Y}}(x, y, \lambda^*).
\]

Let \( \partial \theta_1(x) \) and \( \partial \theta_2(y) \) denote the subdifferentials of \( \theta_1(x) \) and \( \theta_2(y) \), respectively. Then, finding a saddle point of (5) is equivalent to finding \( w^* \in \mathcal{W} \), \( x^* \in \partial \theta_1(x^*) \), and \( \eta^* \in \partial \theta_2(y^*) \) such that the following variational inequality (VI) holds

\[
w^* \in \Omega, \quad (w - w^*)^T T(w^*) \geq 0, \quad \forall w \in \Omega,
\]

(6a)

where

\[
w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad T(w^*) = \begin{pmatrix} \xi^* - A^T \lambda^* \\ \eta^* - B^T \lambda^* \\ Ax^* + By^* - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m.
\]

(6b)

Recall that \( \theta_1(\cdot) \) is convex and \( \theta_2(\cdot) \) is strongly convex, we have

\[
\theta_1(x) - \theta_1(x^*) \geq \langle x - x^*, \xi^* \rangle,
\]

(7)

and

\[
\theta_2(y) - \theta_2(y^*) - \frac{\tau}{2} \| y - y^* \|^2 \geq \langle y - y^*, \eta^* \rangle,
\]

(8)

respectively. From the optimality condition, we have the following equivalent form:

\[
\text{VI}(\Omega, F, \theta) \quad \theta_1(x) + \theta_2(y) - \theta_1(x^*) - \theta_2(y^*) - \frac{\tau}{2} \| y - y^* \|^2 + (w - w^*)^T F(w^*) \geq 0,
\]

\[
\forall w \in \Omega,
\]

(9a)
where
\[ F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}. \]

2.2. Some notations. To facilitate the analysis of ADMM, we introduce some basic notations. First, for the iterate \((x^{k+1}, y^{k+1}, \lambda^{k+1})\) generated by the ADMM (3), we define an auxiliary vector \(\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)\) as
\[ \tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \]
and
\[ \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b). \]

In page 14 of [2], it was commented that the algorithm state in ADMM consists of \(y^k\) and \(\lambda^k\), and \(x^k\) is an intermediate result which is not involved to execute the new iteration, thus \((y, \lambda)\) are essentially involved in the ADMM iteration. Following this idea, for \(w = (x, y, \lambda)\) and \(w^k = (x^k, y^k, \lambda^k)\) generated by (3), we use the notations
\[ v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad \text{and} \quad \bar{v}^k = \begin{pmatrix} \bar{y}^k \\ \bar{\lambda}^k \end{pmatrix} \]
to denote the essential parts of \(w\) and \(w^k\), respectively. We denote the essential part of \(w^*\) in \(\Omega^*\) by using \(v^* = (y^*, \lambda^*)\) and let \(V^*\) denote all the collection of \(v^*\).

Now, we give a lemma to show the relations among \(v^k, \bar{v}^k, \text{and} \bar{v}^{k+1}\).

**Lemma 2.1.** For given \(v^k = (y^k, \lambda^k)\), let \(w^{k+1}\) be generated by the ADMM (3) and \(\bar{w}^k\) be defined by (10). Then, we have
\[ v^{k+1} = v^k - M(v^k - \bar{v}^k), \]
where
\[ M = \begin{pmatrix} I & 0 \\ -\gamma B & \gamma I_m \end{pmatrix}. \]

**Proof.** It follows from (3) and (10) that
\[ \lambda^{k+1} = \lambda^k - \gamma [\beta(\bar{x}^k + By^k - b) - \beta B(y^k - \tilde{y}^k)] \]
\[ = \lambda^k - \gamma [\lambda^k - \tilde{\lambda}^k + \gamma B(y^k - \tilde{y}^k)]. \]

Together with \(y^{k+1} = \tilde{y}^k\), we have the following relationship
\[ \begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\gamma B & \gamma I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \]

The proof is complete. \(\square\)

3. Convergence analysis. In this section, we present convergence results of the ADMM with larger step size when one function is strongly convex. Now, we first establish several lemmas, which will be used to show our main convergence results.

**Lemma 3.1.** Let \(w^{k+1}\) be generated by the ADMM (3) and \(\tilde{w}^k\) be defined by (10). Then, we have
\[ \tilde{w}^k \in \Omega, \quad \theta_1(x) + \theta_2(y) - \theta_1(\tilde{x}^k) - \theta_2(\tilde{y}^k) - \frac{\tau}{2} ||y - \tilde{y}^k||^2 + (w - \tilde{w}^k)^TF(\tilde{w}^k) \geq (v - \bar{v}^k)^TQ(v - \bar{v}^k), \quad \forall w \in \Omega, \]
\[ \text{(12a)} \]
where

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

(12b)

**Proof.** Using the optimality condition of the \( x \)-subproblem in ADMM, we have

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

Recalling the auxiliary vector \( \tilde{x}^k \) defined in (10), we have

\[ \tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T (-A^T \tilde{\lambda}^k) \geq 0, \quad \forall x \in \mathcal{X}. \]  

(13a)

For the \( y \)-subproblem in ADMM, we get the optimality condition

\[ y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) - \frac{\tau}{2} \| y - y^{k+1} \|^2 + (y - y^{k+1})^T \left\{ -B^T \lambda^k + \beta B^T (Ax^{k+1} + By^{k+1} - b) \right\} \geq 0, \quad \forall y \in \mathcal{Y}. \]

Rearrange the terms in the above inequality, we obtain

\[ y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) - \frac{\tau}{2} \| y - \tilde{y}^k \|^2 + (y - \tilde{y}^k)^T \left\{ -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^{k+1} - y^k) \right\} \geq 0, \quad \forall y \in \mathcal{Y}. \]  

(13b)

For \( \tilde{\lambda}^k \) defined by (10), we have

\[ (Ax^{k+1} + By^{k+1} - b) - B(\tilde{y}^{k+1} - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0, \]

and it can be written as

\[ \tilde{\lambda}^k \in \mathbb{R}^m, \quad (\lambda - \tilde{\lambda}^k)^T \{(Ax^{k+1} + By^{k+1} - b) - B(\tilde{y}^{k+1} - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \quad \forall \lambda \in \mathbb{R}^m. \]  

(13c)

Combining (13a), (13b) and (13c), and also noting the definitions in (6), the assertion of this lemma is proved. \( \Box \)

In (12a), the right hand side term \((v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)\) plays a key role in evaluating the solution point, we now focus on investigating it. Recalling the correction step (11a), we have

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

Substituting it into the right-hand side of (12a), we have

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

where

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

\[ = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta^2} I_m \end{pmatrix}. \]  

(14)
Since the matrix $B$ is assumed to be full column rank and the parameters $\beta, \gamma$ are both positive, $H$ is positive definite.

Now, substituting (14) into (12a), we obtain

$$\tilde{w}^k \in \Omega, \quad \theta_1(x) + \theta_2(y) - \theta_1(\tilde{x}^k) - \theta_2(\tilde{y}^k) - \frac{\tau}{2}\|y - \tilde{y}^k\|^2 + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H (v^k - v^{k+1}), \quad \forall w \in \Omega. \quad (16)$$

With the above analysis, we now present the following lemma, which is the basis for proving the strict contraction property of the scheme (3). It is also useful for estimating the convergence rate for the sequence generated by (3).

**Lemma 3.2.** Let the sequence $\{w^k\}$ be generated by the ADMM (3) and $\tilde{w}^k$ be defined by (10). Then we have

$$\theta_1(x) + \theta_2(y) - \theta_1(\tilde{x}^k) - \theta_2(\tilde{y}^k) - \frac{\tau}{2}\|y - \tilde{y}^k\|^2 + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{2}\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2 + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega, \quad (17)$$

where $H$ is defined in (15) and

$$G = Q^T + Q - M^T H M. \quad (18)$$

**Proof.** Since $H$ is positive definite, 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 right-hand side in (14) with

$$a = v, \quad b = \tilde{v}^k, \quad c = v^k, \quad \text{and} \quad d = v^{k+1},$$

we obtain

$$(v - \tilde{v}^k)^T H (v^k - v^{k+1}) = \frac{1}{2}\{\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\} + \frac{1}{2}\{\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\}. \quad (19)$$

Substituting (19) into the right-hand side of (16), we obtain

$$\theta_1(x) + \theta_2(y) - \theta_1(\tilde{x}^k) - \theta_2(\tilde{y}^k) - \frac{\tau}{2}\|y - \tilde{y}^k\|^2 + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{2}\{\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\} + \frac{1}{2}\{\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\}, \quad \forall w \in \Omega. \quad (20)$$

For the last term on the right-hand side of (20), we have

$$\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \tag{11a}$$

$$= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \tag{11a}$$

$$= 2(v^k - \tilde{v}^k)^T H M (v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T H M (v^k - \tilde{v}^k) \tag{15}$$

Substituting (11a) into (20) and using the matrix $G$ defined by (18), the assertion of this theorem is proved. \qed
Now, let us make a further investigation of the matrix $G$ defined by (18). Since $HM = Q$ (see (15)), we have $M^T HM = M^T Q$. Note that

$$M^T Q = \begin{pmatrix} I & -\gamma \beta B^T \\ 0 & \gamma I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I_m \end{pmatrix} = \begin{pmatrix} (1 + \gamma) \beta B^T B & -\gamma B \\ -\gamma B & \frac{2}{\beta} I_m \end{pmatrix}.$$  

Using the above equation, we have

$$G = (Q^T + Q) - M^T HM$$

$$= \begin{pmatrix} 2\beta B^T B & -B \\ -B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} (1 + \gamma) \beta B^T B & -\gamma B \\ -\gamma B & \frac{2}{\beta} I_m \end{pmatrix}$$

$$= \begin{pmatrix} (1 - \gamma) \beta B^T B & -(1 - \gamma) B \\ -(1 - \gamma) B & \frac{2}{\beta}(2 - \gamma) I_m \end{pmatrix}.$$  

(22)

Obviously, when $\gamma > 1$ the matrix $G$ is symmetric but not necessarily positive definite. For this case, we slightly abuse the notation of $\|v\|^2_G$ to denote the term $v^T G v$. Then, we have the following theorem.

**Theorem 3.3.** For the sequence $\{w^k\}$ generated by the ADMM (3) and $\tilde{w}^k$ be defined by (10), we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|^2_G - 2\tau \|\tilde{y}^k - y^*\|^2, \quad \forall v^* \in \mathcal{V}.$$  

(23)

**Proof.** Setting $w = w^*$ in (17), we get

$$\theta_1(x^*) + \theta_2(y^*) - \theta_1(\tilde{x}^k) - \theta_2(\tilde{y}^k) - \frac{\tau}{2} \|y^* - \tilde{y}^k\|^2 + (w^* - \tilde{w}^k)^T F(\tilde{w}^k)$$

$$\geq \frac{1}{2} (\|v^* - v^{k+1}\|_H^2 - \|v^* - v^k\|_H^2) + \frac{1}{2} \|v^k - \tilde{v}^k\|^2_G.$$  

(24)

On the other hand, setting $w = \tilde{w}^k$ in (9a), we get

$$\theta_1(\tilde{x}^k) + \theta_2(\tilde{y}^k) - \theta_1(x^*) - \theta_2(y^*) - \frac{\tau}{2} \|\tilde{y}^k - y^*\|^2 + (\tilde{w}^k - w^*)^T F(w^*) \geq 0.$$  

(25)

Adding (24) and (25), we have

$$-\tau \|\tilde{y}^k - y^*\|^2 + (\tilde{w}^k - w^*)^T (F(w^*) - F(\tilde{w}^k))$$

$$\geq \frac{1}{2} (\|v^* - v^{k+1}\|_H^2 - \|v^* - v^k\|_H^2) + \frac{1}{2} \|v^k - \tilde{v}^k\|^2_G.$$  

(26)

Recall the fact $(\tilde{w}^k - w^*)^T (F(w^*) - F(\tilde{w}^k)) = 0$, we obtain the assertion. \hfill \Box

Recall (22), when $0 < \gamma \leq 1$, the matrix $G$ is positive definite. For this case, the inequality (23) indicates that the sequence $\{v^k\}$ generated by (2) is strictly contractive with respect to the solution set $\mathcal{V}^*$, and the convergence of the sequence can be easily established. However, when $\gamma > 1$, $G$ is not positive definite, and we need to further investigate the quadratic term $\|v^k - \tilde{v}^k\|^2_G$. This is given by the following lemma.

**Lemma 3.4.** Let the sequence $\{w^k\}$ be generated by the ADMM (3) and $\tilde{w}^k$ be defined by (10). Then we have

$$\|v^k - \tilde{v}^k\|^2_G = \beta \|B(y^k - y^{k+1})\|^2 + (2 - \gamma) \beta \|Ax^{k+1} + By^{k+1} - b\|^2$$

$$+ 2\beta (Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}).$$  

(27)
Proof. Because $G = \begin{pmatrix} (1 - \gamma)\beta B^T B - (1 - \gamma)B^T \\ -(1 - \gamma)B - \frac{1}{\beta}(2 - \gamma)I_m \end{pmatrix}$ (see (22)), $v = \begin{pmatrix} y \\ \lambda \end{pmatrix}$ and $\tilde{y}^k = y^{k+1}$, we have
\[
\|v^k - \tilde{v}^k\|_G^2 = (1 - \gamma)\beta \|B(y^k - y^{k+1})\|^2 - 2(1 - \gamma)(y^k - y^{k+1})^T B^T(\lambda^k - \tilde{\lambda}^k) + (2 - \gamma)\beta \|\lambda^k - \tilde{\lambda}^k\|^2. \tag{28}
\]
Notice that (see (10))
\[
\lambda^k - \tilde{\lambda}^k = \beta(Ax^{k+1} + By^k - b) = \beta\{(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\}.
\]
Then, we have
\[
\|\lambda^k - \tilde{\lambda}^k\|^2 = \beta^2\{(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\}^2
= \beta^2\|Ax^{k+1} + By^{k+1} - b\|^2 + 2\beta^2\{(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\}^2 + 2\beta\|B(y^k - y^{k+1})\|^2. \tag{29}
\]
Substituting (29) into the right-hand side of (28), we obtain
\[
\|v^k - \tilde{v}^k\|_G^2 = (1 - \gamma)\beta \|B(y^k - y^{k+1})\|^2 - 2(1 - \gamma)\beta(y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b) - 2(1 - \gamma)\beta \|B(y^k - y^{k+1})\|^2 + (2 - \gamma)\beta \|B(y^k - y^{k+1})\|^2 + 2(2 - \gamma)\beta(y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b) + 2\beta(y^k - y^{k+1})^T B^T (Ax^{k+1} + By^{k+1} - b).
\]
This completes the proof. \qed

In the following we will treat the crossing term in the right-hand side of (27), namely,
\[
\beta(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}).
\]
We want to find a lower bound for it.

Lemma 3.5. Let the sequence \{\omega^k\} be generated by the ADMM (3). Then, for any $\gamma > 0$, we have
\[
\beta(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \geq -\frac{(1 - \gamma)^2}{2(T - \gamma)} \beta \|B(y^{k+1} - y^k)\|^2 - \frac{T - \gamma}{2} \beta \|Ax^k + By^k - b\|^2 + \tau \|y^k - y^{k+1}\|^2, \tag{30}
\]
where $T \in (\gamma, 2)$ is a constant.

Proof. The optimality condition of the $y$-subproblem (3b) is
\[
y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) - \frac{\tau}{2} \|y - y^{k+1}\|^2 + (y - y^{k+1})^T \{ -B^T \lambda^k + \beta B^T (Ax^{k+1} + By^{k+1} - b) \} \geq 0, \quad \forall y \in \mathcal{Y}. \tag{31}
\]
Analogously, for the previous iteration, we have
\[
y^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^k) - \frac{\tau}{2} \|y - y^k\|^2 + (y - y^k)^T \{-B^T \lambda^{k-1} + \beta B^T (Ax^k + By^k - b)\} \geq 0, \quad \forall y \in \mathcal{Y}.
\] (32)

Setting \( y = y^k \) and \( y = y^{k+1} \) in (31) and (32), respectively, and then adding them, we get
\[
(y^k - y^{k+1})^T B \{\lambda^{k-1} - \lambda^k - \beta (Ax^k + By^k - b) + \beta (Ax^{k+1} + By^{k+1} - b)\}
\geq \tau \|y^k - y^{k+1}\|^2.
\] (33)

Note that in the \{k-1\}-th iteration (see (3c)), we have
\[
\lambda^k = \lambda^{k-1} - \gamma \beta (Ax^k + By^k - b).
\] (34)

Substituting (34) into (33) and with a simple manipulation, we have
\[
(y^k - y^{k+1})^T B \{\beta (Ax^{k+1} + By^{k+1} - b) - (1 - \gamma) \beta (Ax^k + By^k - b)\}
\geq \tau \|y^k - y^{k+1}\|^2.
\]

Then, we get
\[
(y^k - y^{k+1})^T B (Ax^{k+1} + By^{k+1} - b)
\geq (1 - \gamma)(y^k - y^{k+1})^T B (Ax^k + By^k - b) + \tau \|y^k - y^{k+1}\|^2.
\] (35)

Using the Cauchy-Schwarz inequality to the crossing terms in the right-hand side of (35), we have
\[
(y^k - y^{k+1})^T B (Ax^{k+1} + By^{k+1} - b)
\geq -\frac{(1 - \gamma)^2}{2(T - \gamma)} \beta \|B(y^{k+1} - y^k)\|^2 - \frac{T - \gamma}{2} \beta \|Ax^k + By^k - b\|^2
\]
\[
+ \tau \|y^k - y^{k+1}\|^2,
\] (36)

this lemma is proved.

Based on Lemma 3.4 and Lemma 3.5, we now establish the following results.

**Lemma 3.6.** Let the sequence \( \{w^k\} \) be generated by the ADMM (3) and \( \hat{w}^k \) be defined by (10). Then, for any \( T \in (\gamma, 2) \) we have
\[
\|v^k - \hat{v}^k\|^2 \geq \left[ 1 - \frac{(1 - \gamma)^2}{2T(T - \gamma)} \beta \|B(y^k - y^{k+1})\|^2 + (2 - T)\beta \|Ax^{k+1} + By^{k+1} - b\|^2
\]
\[
+ 2\tau \|y^k - y^{k+1}\|^2 + (T - \gamma)\beta \|Ax^{k+1} + By^{k+1} - b\|^2
\]
\[
- (T - \gamma)\beta \|Ax^k + By^k - b\|^2
\] (37)

**Proof.** Substituting (30) into (27), we obtain the assertion (37) immediately.

**Theorem 3.7.** Let the sequence \( \{w^k\} \) be generated by the ADMM (2). Then, for any \( T > \gamma \), we have
\[
\forall v^* \in \mathcal{Y}^*, \quad \|v^{k+1} - v^*\|^2_H + (T - \gamma)\beta \|Ax^{k+1} + By^{k+1} - b\|^2 + \tau \|y^{k+1} - y^*\|^2
\leq \left(\|v^k - v^*\|^2_H + (T - \gamma)\beta \|Ax^k + By^k - b\|^2 + \tau \|y^k - y^*\|^2\right)
\]
\[
- \beta \left\{ (1 - \gamma)^2 \|B(y^k - y^{k+1})\|^2 + \frac{5\tau}{2} \|y^k - y^{k+1}\|^2 \right\}
\]
\[
+ (2 - T)\|Ax^{k+1} + By^{k+1} - b\|^2.
\] (38)
Proof. Substituting (37) into (23) and with a simple manipulation, we have

\[ \forall v^* \in \mathcal{V}^*, \|v^{k+1} - v^*\|^2_H + (T - \gamma)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 + \tau\|y^{k+1} - y^*\|^2 \leq (\|v^k - v^*\|^2_H + (T - \gamma)\beta\|Ax^k + By^k - b\|^2 + \tau\|y^k - y^*\|^2) - \beta \left\{1 - \frac{(1 - \gamma)^2}{T - \gamma}\right\} \|B(y^k - y^{k+1})\|^2 + 2\tau\|y^k - y^{k+1}\|^2 + (2 - T)\|Ax^{k+1} + By^{k+1} - b\|^2 + \tau\|y^k - y^*\|^2 + \tau\|y^{k+1} - y^*\|^2\]. \quad (39) \]

Note that

\[ \tau\|y^k - y^*\|^2 + \tau\|y^{k+1} - y^*\|^2 \geq \frac{\tau}{2}\|y^k - y^{k+1}\|^2, \quad (40) \]

we obtain the assertion directly. \qed

According to our assumption that \( \tau = s \cdot \rho(B^T B) \), we have

\[ \left(1 - \frac{(1 - \gamma)^2}{T - \gamma}\right) \|B(y^k - y^{k+1})\|^2 + \frac{5\tau}{2}\|y^k - y^{k+1}\|^2 \geq \left(1 + \frac{5s}{2} - \frac{(1 - \gamma)^2}{T - \gamma}\right) \|B(y^k - y^{k+1})\|^2. \]

So if the following two inequalities hold

\[
\begin{cases}
1 + \frac{5s}{2} - \frac{(1 - \gamma)^2}{T - \gamma} > 0, \\
2 - T > 0,
\end{cases} \quad (41a) \]

the last term on the right hand side in (38) is in definite quadratic form.

Now, by solving the two inequalities, we have

\[ 0 < \gamma < \frac{1 - \frac{5}{2}s + \sqrt{(\frac{5}{2}s + 3)^2 - 4}}{2}. \quad (42) \]

With the above analysis, we are now ready to state the convergence result of the algorithm.

**Theorem 3.8.** Suppose that \( \theta_2(\cdot) \) is \( \tau \)-strongly convex and \( \tau = s \cdot \rho(B^T B) \). Then, for the step size

\[ 0 < \gamma < \frac{1 - \frac{5}{2}s + \sqrt{(\frac{5}{2}s + 3)^2 - 4}}{2}, \quad (43) \]

the sequence \( \{v^k\} \) generated by the ADMM (3) converges to a solution point \( v^\infty \in \mathcal{V}^* \).

**Proof.** Let \( (y^0, \lambda^0) \) be the initial point and \( x^1 \) be the computational result of \( (y^0, \lambda^0) \). Using (38), we have

\[ \sum_{k=1}^{\infty} \beta \left\{\left(1 - \frac{(1 - \gamma)^2}{T - \gamma}\right) \|B(y^k - y^{k+1})\|^2 + \frac{5\tau}{2}\|y^k - y^{k+1}\|^2\right\} + (2 - T)\|Ax^{k+1} + By^{k+1} - b\|^2 \leq (\|v^1 - v^*\|^2_H + (T - \gamma)\beta\|Ax^1 + By^1 - b\|^2 + \tau\|y^1 - y^*\|^2). \]

Then from (41), we have

\[ \lim_{k \to \infty} (\|y^k - y^{k+1}\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2) = 0. \quad (44) \]
Furthermore, we know from (38) that the sequence \( \{v^k\} \) is bounded. Thus it must have a finite cluster point \( v^\infty \), and there exists a subsequence \( \{v^{k_j}\} \) convergent to this point. We now prove that \( v^\infty \) is a solution point of (6). Let \( \hat{x}^\infty \) be induced by (3) with given \( (y^\infty, \lambda^\infty) \). Recall the matrix \( B \) is assumed to be full column rank. From (44), we immediately have
\[
y^\infty - \tilde{y}^\infty = 0 \quad \text{and} \quad \lambda^\infty - \tilde{\lambda}^\infty = 0.
\]
Then, it follows from (12a) that
\[
\tilde{w}^\infty \in \Omega, \quad \theta_1(x) + \theta_2(y) - \theta_1(\hat{x}^\infty) - \theta_2(\tilde{y}^\infty) - \frac{T}{2} \|y - \tilde{y}^\infty\|^2
\]
\[
+ (w - \tilde{w}^\infty)^T F(\tilde{w}^\infty) \geq 0, \quad \forall w \in \Omega.
\]
We obtain that \( \tilde{w}^\infty = w^\infty \) is a solution point of (6). On the other hand, since (38) holds for any solution point of (6) and \( w^\infty \in \Omega^* \), we have
\[
\|v^{k+1} - v^\infty\|^2_H + (T - \gamma)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 + \tau\|y^{k+1} - y^\infty\|^2
\leq \|v^k - v^\infty\|^2_H + (T - \gamma)\beta\|Ax^k + By^k - b\|^2 + \tau\|y^k - y^\infty\|^2.
\]
Because \( \lim_{k \to \infty} \|Ax^k + By^k - b\|^2 = 0 \) and \( \lim_{k \to \infty} \|y^k - y^{k-1}\|^2 = 0 \), the sequence \( \{v^k\} \) cannot have another cluster point and thus it converges to a solution point \( v^* = v^\infty \in \mathcal{V}^* \).

**Remark 2.** To the best of our knowledge, it is the first time that such a formulation (43) is presented. As we can see, when \( s = 0 \), the parameter \( \gamma \) lies in \((0, 1+\sqrt{5}/2)\); while \( s \to +\infty \), we have
\[
\lim_{s \to +\infty} \frac{1 - \frac{5}{2}s + \sqrt{(\frac{5}{2}s + 3)^2 - 4}}{2} = 2,
\]
then \( \gamma \) can be chosen in the interval \((0, 2)\), which is very similar with the relaxation factor for the ALM. Note that the function \( \frac{1 - \frac{5}{2}s + \sqrt{(\frac{5}{2}s + 3)^2 - 4}}{2} \) is monotonically increasing regards of \( s \), so the step size can be larger than \( 1+\sqrt{5}/2 \) when \( s \) is positive. Thus the step size is further enlarged compared to the golden step size.

**Remark 3.** Although the formula (43) regards of \( s \), we can easily estimate the feasible range of \( \gamma \) with given \( \tau \) and matrix \( B \) by using remark 1. This can be realized as follows: we first find a value of \( s \) such that \( \tau = s \cdot \rho(B^TB) \), then substitute \( s \) into (43) to derive the feasible range of \( \gamma \).

4. **Conclusions.** In this paper, we focus on the convergence of the alternating direction method of multipliers (ADMM). We show that the dual step size of ADMM can be larger than the golden ration when one function term is strongly convex, which is not provided in the existing ADMM literatures. We also derive an exact relationship between the modulus and the dual step size. Since the step size considered here is larger than Glowinski’s step size, the convergence analysis of this paper covers ADMM and Glowinski’s ADMM as special cases.

**Acknowledgments.** We would like to thank the editor and the anonymous referees for their valuable suggestions which helped us dramatically improve the presentation of the manuscript.
REFERENCES

[1] F. Bach, R. Jenatton, J. Mairal and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Mach. Learn.*, 4 (2012), 1–106.
[2] 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. Learn.*, 3 (2010), 1–122.
[3] X. J. Cai, D. Han and X. M. Yuan, On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function, *Comput. Optim. Appl.*, 66 (2017), 39–73.
[4] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, *Math. Program.*, 155 (2016), 57–79.
[5] W. Deng and W. T. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, *J. Sci. Comput.*, 66 (2016), 889–916.
[6] J. Eckstein and D. F. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.*, 55 (1992), 293–318.
[7] J. Eckstein and M. Fukushima, Some reformulation and applications of the alternating direction method of multipliers, *Large Scale Optimization, Kluwer Acad. Publ., Dordrecht*, (1994), 115–134.
[8] J. Eckstein and W. Yao, Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives, *Pacific J. Optim.*, 11 (2015), 619–644.
[9] M. Fortin and R. Glowinski *Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems*, Studies in Mathematics and its Applications, 15. North-Holland Publishing Co., 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), 17–40.
[11] R. Glowinski and A. Marrocco, Sur l’approximation par éléments finis d’ordre un et la résolution par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires, *Revue Fr. Autom. Inform. Rech. Opér.*, *Anal. Numér.*, 9 (1975), 41–76.
[12] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer Series in Computational Physics, Springer-Verlag, New York, 1984.
[13] R. Glowinski, On alternating direction methods of multipliers: A historical perspective, *Modeling, Simulation and Optimization for Science and Technology, Computational Methods in Applied Sciences, Springer, Dordrecht*, 34 (2014), 59–82.
[14] R. Glowinski, S. J. Osher and W. T. Yin *Splitting Methods in Communication, Imaging, Science, and Engineering*, Scientific Computation. Springer, Cham, 2016.
[15] D. Han and X. M. Yuan, A note on the alternating direction method of multiplier, *J. Optim. Theory Appl.*, 155 (2012), 227–238.
[16] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM J. Optim.*, 24 (2014), 1011–1040.
[17] B. S. He, F. Ma and X. M. Yuan, Convergence study on the symmetric version of ADMM with larger step sizes, *SIAM J. Imaging Sci.*, 9 (2016), 1467–1501.
[18] B. S. He and X. M. Yuan, On the O(1/n) convergence rate of the Douglas-Rachford alternating direction method, *SIAM J. Num. Anal.*, 50 (2012), 700–709.
[19] B. S. He and X. M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers, *Numer. Math.*, 130 (2015), 567–577.
[20] M. Y. Hong and Z.-Q. Luo, On the linear convergence of the alternating direction method of multipliers, *Math. Program.*, 162 (2017), 165–199.
[21] Y. Shen, Z. W. Wen and Y. Zhang, Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization, *Optim. Methods Softw.*, 29 (2014), 239–263.
[22] M. Tao and X. M. Yuan, On Glowinski’s open question on the alternating direction method of multipliers, *J. Optim. Theory Appl.*, 179 (2018), 163–196.
[23] Z. W. Wen, D. Goldfarb and W. T. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, *Math. Program. Comput.*, 2 (2010), 203–230.
[24] J. F. Yang, Y. Zhang and W. T. Yin, A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data, IEEE J. Selc. Topics Signal Proc., 4 (2010), 288–297.

Received January 2019; revised June 2019.

E-mail address: mafengnju@gmail.com
E-mail address: s20130720@126.com
E-mail address: 13571996716@139.com
E-mail address: wujian6029@163.com