SIAM J. OPTIM.
Vol. 22, No. 4, pp. 1431–1448

ON THE $O(1/t)$ CONVERGENCE RATE OF ALTERNATING DIRECTION METHOD WITH LOGARITHMIC-QUADRATIC PROXIMAL REGULARIZATION*

MIN TAO† AND XIAOMING YUAN‡

Abstract. It was shown recently that the Douglas–Rachford alternating direction method of multipliers can be combined with the logarithmic-quadratic proximal regularization for solving a class of variational inequalities with separable structures. This paper further shows a worst-case $O(1/t)$ convergence rate for this algorithm where a general Glowinski relaxation factor is used.

Key words. alternating direction method of multipliers, logarithmic-quadratic proximal regularization, convergence rate, Glowinski's relaxation factor, variational inequality

AMS subject classifications. 90C25, 90C33, 65K05

DOI. 10.1137/110847639

1. Introduction.

The Douglas–Rachford alternating direction method of multipliers (ADM for short) was proposed originally in [19] (see also [16]), and it has been well studied in the literature. We refer to [6, 13, 15, 17, 18] for some early articles on partial differential equations; to [9, 14, 21, 27, 33] for some studies on convex programming and variational inequalities; and to [4, 5, 10, 11, 22, 29, 28, 30, 31, 32, 34, 35, 38] and the references therein for some recent novel applications in various areas such as image processing, statistical learning, and semidefinite programming.

In this paper, we concentrate our discussion of ADM on a class of variational inequalities with block-separable structure: Find $u^* \in \Omega$ such that

$$(u - u^*)^T U(u^*) \geq 0 \quad \forall u \in \Omega, \tag{1.1}$$

with

$$u = \begin{pmatrix} x \\ y \end{pmatrix} \qquad U(u) = \begin{pmatrix} f(x) \\ g(y) \end{pmatrix}, \tag{1.2}$$

and

$$\Omega = \left\{(x, y) \mid Ax + By = b,\ x \in \mathcal{R}^n_+,\ y \in \mathcal{R}^m_+\right\}, \tag{1.3}$$

where $A \in \mathcal{R}^{l\times n}$ and $B \in \mathcal{R}^{l\times m}$ are given matrices; $b \in \mathcal{R}^l$ is a given vector; and $f : \mathcal{R}^n_+ \to \mathcal{R}^n$ and $g : \mathcal{R}^m_+ \to \mathcal{R}^m$ are continuous and monotone operators. The variational inequality (1.1)–(1.3) subsumes a class of convex programming problems with block-separable structures and thus finds a broad spectrum of applications. We denote by $\mathrm{VI}(\Omega, U)$ the variational inequality (1.1)–(1.3). In addition, we denote by Ω^* the solution set of $\mathrm{VI}(\Omega, U)$ and assume it to be nonempty.

*Received by the editors September 12, 2011; accepted for publication (in revised form) July 17, 2012; published electronically October 31, 2012.
http://www.siam.org/journals/siopt/22-4/84763.html

†School of Science, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210093, China (taomin0903@gmail.com). This author's research was supported by the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (NY210049).

‡Corresponding author. Department of Mathematics, Hong Kong Baptist University, Hong Kong, China (xmyuan@hkbu.edu.hk). This author's research was supported by the Hong Kong General Research Fund (HKBU 203311) and the HKBU FRG Fund (FRG2/11-12/120).

For solving $\mathrm{VI}(\Omega, U)$, ADM generates the new iterate $w^{k+1} := (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \mathcal{R}^n_+ \times \mathcal{R}^m_+ \times \mathcal{R}^l$ via solving the subproblems

$$0 \le x^{k+1} \perp \left\{ f(x^{k+1}) - A^T[\lambda^k - H(Ax^{k+1} + By^k - b)] \right\} \ge 0, \tag{1.4}$$

$$0 \le y^{k+1} \perp \left\{ g(y^{k+1}) - B^T[\lambda^k - H(Ax^{k+1} + By^{k+1} - b)] \right\} \ge 0, \tag{1.5}$$

$$\lambda^{k+1} = \lambda^k - \gamma H(Ax^{k+1} + By^{k+1} - b), \tag{1.6}$$

where $\lambda^k \in \mathcal{R}^l$ is the Lagrange multiplier, $H \in \mathcal{R}^{l\times l}$ is a matrix of penalty parameters, and $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is the Glowinski relaxation factor proposed in [17]. We refer to [8, 15] for the relationship between ADM and the Douglas–Rachford splitting method [7]. In the literature, H is often chosen as $\beta \cdot I$ (where I is the identity matrix in $\mathcal{R}^{l\times l}$ and $\beta > 0$ is a scalar) and $\gamma = 1$. With these choices, the convergence of ADM can be easily established. However, here we consider a more general case where H is a generic symmetric positive definite matrix. This general discussion is particularly inspired by the fact that preconditioning techniques become allowable for solving the ADM subproblems with a generic H, which are useful for many applications such as image processing problems. In addition, because of the effectiveness verified numerically in the literature (e.g., [17, 18]), we consider a general Glowinski relaxation factor γ in the step (1.6), instead of simply taking $\gamma = 1$. With this Glowinski relaxation factor, the resulting convergence analysis becomes more demanding technically.

According to (1.4)–(1.6), ADM makes it possible to exploit the properties of $f(x)$ and $g(y)$ individually, and the decomposed ADM subproblems are often simple enough to have closed-form solutions or can be easily solved up to high precision for many concrete applications of $\mathrm{VI}(\Omega, U)$. This feature contributes much to the recent burst of ADM's applications in various fields. However, for a general case where the subproblems (1.4) and (1.5) do not have closed-form solutions or it is not easy to solve them up to high precision, inner iterative procedures are required to find approximate solutions of these subproblems, and the efficiency of ADM relies critically on how efficiently these subproblems can be solved. In [21], it was suggested to regularize the subproblems (1.4) and (1.5) with quadratic proximal terms, yielding the following scheme of ADM with quadratic proximal regularization:

$$0 \le x^{k+1} \perp \left\{ f(x^{k+1}) - A^T[\lambda^k - H(Ax^{k+1} + By^k - b)] + R(x^{k+1} - x^k) \right\} \ge 0, \tag{1.7}$$

$$0 \le y^{k+1} \perp \left\{ g(y^{k+1}) - B^T[\lambda^k - H(Ax^{k+1} + By^{k+1} - b)] + S(y^{k+1} - y^k) \right\} \ge 0, \tag{1.8}$$

$$\lambda^{k+1} = \lambda^k - H(Ax^{k+1} + By^{k+1} - b), \tag{1.9}$$

where $R \in \mathcal{R}^{n\times n}$ and $S \in \mathcal{R}^{m\times m}$ are symmetric positive definite matrices and are called proximal matrices, and $R(x^{k+1} - x^k)$ and $S(y^{k+1} - y^k)$ are quadratic proximal regularization terms. At each iteration, ADM with quadratic proximal regularization requires solving two strongly monotone complementarity problems. The general scheme (1.7)–(1.9) turns out to be very useful when some customized choices of the proximal matrices R and S are chosen for particular applications. For example, if the proximal matrices are chosen as $R = rI - A^T HA$ and $S = sI - B^T HB$, where r and s are positive scalars and I is the identity matrix with appropriate dimensionality, the splitting Bregman iteration [20] and the split inexact Uzawa method in [39, 40], which have proved to be very efficient for image processing problems, are recovered.

Inspired by some nice properties of the logarithmic-quadratic proximal (LQP) method proposed in [1, 3] (see also [25, 26, 37] for some algorithmic developments), it was recommended in [36] to regularize the ADM subproblems (1.4) and (1.5) with the LQP regularization. With the LQP regularization, the subproblems (1.4) and (1.5) reduce to two systems of equations which are in general easier than the complementarity problems (1.4)–(1.5) and (1.7)–(1.8). More specifically, the ADM with LQP regularization in [36] has the following iterative scheme:

$$f(x^{k+1})-A^T[\lambda^k-H(Ax^{k+1}+By^k-b)]+R[(x^{k+1}-x^k)+\mu(x^k-X_k^2(x^{k+1})^{-1})]=0, \tag{1.10}$$

$$g(y^{k+1})-B^T[\lambda^k-H(Ax^{k+1}+By^{k+1}-b)]+S[(y^{k+1}-y^k)+\mu(y^k-Y_k^2(y^{k+1})^{-1})]=0, \tag{1.11}$$

$$\lambda^{k+1}=\lambda^k-\gamma H(Ax^{k+1}+By^{k+1}-b), \tag{1.12}$$

where $\mu\in(0,1)$ is a given constant (usually $\mu\in(0,0.2)$ is preferable, as demonstrated numerically in [25, 26, 37]), $X_k=\operatorname{diag}(x_1^k,x_2^k,\dots,x_n^k)$, $(x^{k+1})^{-1}\in\mathcal{R}^n$ is a vector whose jth element is $1/x_j^{k+1}$, $Y_k=\operatorname{diag}(y_1^k,y_2^k,\dots,y_n^k)$, $(y^{k+1})^{-1}\in\mathcal{R}^n$ is a vector whose jth element is $1/y_j^{k+1}$, and R and S are defined as in (1.7)–(1.8). The scheme (1.10)–(1.12) is thus a blend of ADM and LQP. Note that γ in (1.12) was taken as 1 in [36], but here we consider the case where the Glowinski relaxation factor $\gamma\in(0,\frac{1+\sqrt{5}}{2})$ is adopted.

This paper is a further study on the convergence rate of ADM with LQP regularization. We show that after t iterations of the scheme (1.10)–(1.12), it is possible to find an approximate solution of $\mathrm{VI}(\Omega,U)$ with an accuracy of $O(1/t)$. Hence, a worst-case $O(1/t)$ convergence rate of the scheme (1.10)–(1.12) is derived. We thus show that the blend of ADM and LQP enjoys the same convergence rate as ADM with quadratic proximal regularization which was established in [23] (and in [24] for some projection methods). We also refer the reader to [2, 24] for some convergence rate results of projection methods with LQP regularization.

The rest of this paper is organized as follows. In section 2, we provide some preliminaries for later analysis. Then, in section 3 we derive a worst-case $O(1/t)$ convergence rate for the scheme (1.10)–(1.12). We treat the special case where $\gamma=1$ in section 4. Finally, some conclusions are made in section 5.

2. Preliminaries.

In this section we summarize some existing results and show some trivial results which are useful for later analysis.

2.1. A variational reformulation.

First, we note that $\mathrm{VI}(\Omega,U)$ can be rewritten in the following compact form:

$$(x-x^*)^Tf(x^*)+(y-y^*)^Tg(y^*)+(w-w^*)^TF(w^*)\ge 0 \quad \forall w\in\mathcal{W}, \tag{2.1}$$

where

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\ F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}, \quad \text{and} \quad \mathcal{W}:=\mathcal{R}_+^n\times\mathcal{R}_+^m\times\mathcal{R}^l. \tag{2.2}$$

We denote by $\mathrm{VI}(\mathcal{W},F)$ the variational inequality (2.1)–(2.2). Note that the mapping $F(w)$ is monotone since it is affine with a skew-symmetric matrix. We denote by $\mathcal{W}^*$ the solution set of $\mathrm{VI}(\mathcal{W},F)$. Then, for all $(x^*,y^*)\in\Omega^*$, there exists a vector

$\lambda^* \in \mathcal{R}^l$ such that $w^* := (x^*, y^*, \lambda^*)$ is a solution point of VI$(\mathcal{W}, F)$. Hence, $\mathcal{W}^*$ is also nonempty under the nonempty assumption on Ω^*.

Inspired by Theorem 2.3.5 in [12], we show a characterization of $\mathcal{W}^*$ which later plays a crucial role in convergence rate analysis. Since its proof is similar to those of Theorem 2.3.5 in [12] and Theorem 2.1 in [23], we omit it.

THEOREM 2.1. *The solution set of VI$(\mathcal{W}, F)$ is convex and can be characterized as*

$$\mathcal{W}^* = \bigcap_{w\in\mathcal{W}} \left\{\tilde{w} \in \mathcal{W} : (x-\tilde{x})^T f(x) + (y-\tilde{y})^T g(y) + (w-\tilde{w})^T F(w) \geq 0\right\}. \tag{2.3}$$

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

$$(\tilde{x}-x)^T f(x) + (\tilde{y}-y)^T g(y) + (\tilde{w}-w)^T F(w) \leq \epsilon \quad \forall w \in \mathcal{W}.$$

In the rest of the paper, our purpose is to show that after t iterations of the scheme (1.10)–(1.12), we can find $\tilde{w}_t \in \mathcal{W}$ such that

$$\sup_{w\in\mathcal{W}} \{(\tilde{x}_t - x)^T f(x) + (\tilde{y}_t - y)^T g(y) + (\tilde{w}_t - w)^T F(w)\} \leq \epsilon, \tag{2.4}$$

where $\epsilon = O(1/t)$. A worst-case $O(1/t)$ convergence rate of ADM with LQP regularization is thus implied by (2.4).

2.2. Some results in [36]. We summarize some results in [36], which will be used later.

LEMMA 2.2. *Let $p = (p_1, p_2, \ldots, p_N) \in \mathcal{R}^N$ be a positive vector (i.e., $p_i > 0$ for all i's), and let $P := \mathrm{diag}(p_1, p_2, \ldots, p_N) \in \mathcal{R}^{N\times N}$. Let $q(\ldots)$ be a monotone mapping defined in $\mathcal{R}^N_+$. For a given positive vector a^k in $\mathcal{R}^N$, let $A_k := \mathrm{diag}(a^k_1, a^k_2, \ldots, a^k_N)$. Let $\mu \in (0, 1)$. Then the equation*

$$q(a) + P[(a - a^k) + \mu(a^k - A_k^2 a^{-1})] = 0 \tag{2.5}$$

has a unique positive solution. Moreover, let the solution of (2.5) *be denoted by a. Then, for any vector $b \in \mathcal{R}^N_+$, we have*

$$(b-a)^T q(a) \geq \tfrac{1+\mu}{2}\left(\|a-b\|_P^2 - \|a^k - b\|_P^2\right) + \tfrac{1-\mu}{2}\|a^k - a\|_P^2. \tag{2.6}$$

LEMMA 2.3. *Let $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ be generated by* (1.10)–(1.12) *from given $w^k = (x^k, y^k, \lambda^k)$. Then, for any $w = (x, y, \lambda) \in \mathcal{W}$, we have*

$$(x-x^{k+1})^T f_k(x^{k+1}) \geq \frac{1+\mu}{2}(\|x^{k+1}-x\|_R^2 - \|x^k - x\|_R^2) + \frac{1-\mu}{2}\|x^k - x^{k+1}\|_R^2 \tag{2.7}$$

and

$$(y-y^{k+1})^T g_k(y^{k+1}) \geq \frac{1+\mu}{2}(\|y^{k+1}-y\|_S^2 - \|y^k - y\|_S^2) + \frac{1-\mu}{2}\|y^k - y^{k+1}\|_S^2, \tag{2.8}$$

where

$$f_k(x^{k+1}) = f(x^{k+1}) - A^T[\lambda^k - H(Ax^{k+1} + By^k - b)] \tag{2.9}$$

and

$$g_k(y^{k+1}) = g(y^{k+1}) - B^T[\lambda^k - H(Ax^{k+1} + By^{k+1} - b)]. \tag{2.10}$$

2.3. Some useful results. Finally, we prove some simple but useful results. The first two lemmas are elementary; we thus omit their proofs.

LEMMA 2.4. *Let $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ and $T := \frac{1}{3}(\gamma^2 - \gamma + 5)$. Then, we have the following:*

1. $1 + \gamma - \gamma^2 > 0$.
2. $T - \gamma = \frac{1}{3}(\gamma^2 - 4\gamma + 5) = \frac{1}{3}[(\gamma - 2)^2 + 1] > \frac{1}{3}$.
3. $1 - \frac{(1-\gamma)^2}{T-\gamma} > 0$.
4. $2 - T = \frac{1}{3}(1 + \gamma - \gamma^2) > 0$.

LEMMA 2.5. *Let $G \in \mathcal{R}^{N\times N}$ be positive semidefinite and symmetric, and let $\|x\|_G := \sqrt{x^TGx}$. Then it holds that*

(2.11)
$$(a-b)^TG(c-d) = \frac{1}{2}\big(\|a-d\|_G^2 - \|a-c\|_G^2\big) + \frac{1}{2}\big(\|c-b\|_G^2 - \|d-b\|_G^2\big) \forall a, b, c, d \in \mathcal{R}^N.$$

LEMMA 2.6. *Let $\mu \in (0, 0.2)$. Then there exist scalars $\eta \in (0, \frac{1-\mu}{2})$ and $\rho > \mu$ such that*

$$(1+\eta)a^2 - \mu b^2 - (1+\mu)ab \geq \rho(a^2 - b^2) \quad \forall a, b \in \mathcal{R}. \tag{2.12}$$

Proof. The conclusion is obvious if $b = 0$. Thus, we now assume that $b \neq 0$. By dividing b^2 in both sides of (2.12), it is easy to see that (2.12) is equivalent to

$$(1+\eta-\rho)t^2 - (1+\mu)t + (\rho - \mu) \geq 0 \quad \forall t \in \mathcal{R}.$$

For any $\eta \in (0, \frac{1-\mu}{2})$ we have $1 + \eta > \mu$. Thus, for any $\rho \in (\mu, 1+\eta)$, we have $1 + \eta - \rho > 0$. To prove the existence of ρ and η such that the above inequality is satisfied, we introduce the quadratic function

$$h_1(t) := (1+\eta-\rho)t^2 - (1+\mu)t + (\rho-\mu)$$

and show that $h_1(t) \geq 0$ for any t. Equivalently, we need only prove that the minimal value of $h_1(t)$ is nonnegative.

Obviously, $h_1(t)$ attains the minimal value

$$h_1(t_{min}) := \rho - \mu - \frac{(1+\mu)^2}{4(1+\eta-\rho)}$$

at the point $t_{min} := \frac{1+\mu}{2(1+\eta-\rho)}$. When $\mu \in (0, 0.2)$, we have $2\mu < \frac{1-\mu}{2}$, and we can choose $2\mu < \eta < \frac{1-\mu}{2}$. Recall we require that $\rho \in (\mu, 1+\eta)$. Thus, we can represent ρ by $\rho = \mu + \delta$ with $\delta > 0$ and seek an appropriate δ to ensure the nonnegativeness of $h_1(t)$. Obviously, $h_1(t_{min}) \geq 0$ whenever the positive scalar δ can ensure that

$$4\delta^2 - 4(1+\eta-\mu)\delta + (1+\mu)^2 \leq 0.$$

Defining

$$h_2(\delta) := 4\delta^2 - 4(1+\eta-\mu)\delta + (1+\mu)^2,$$

$h_2(\delta)$ attains its minimal value

$$h_2(\delta_{min}) := (1+\mu)^2 - (1+\eta-\mu)^2 < 0 \tag{2.13}$$

at the point $\delta_{min} := \frac{1+\eta-\mu}{2} > 0$, where the inequality (2.13) follows from the fact that $\eta > 2\mu$, i.e., $1 + \eta - \mu > 1 + \mu$. Notice that $h_2(0) = (1+\mu)^2 > 0$. Thus, there exists a positive scalar δ in $(0, \frac{1+\eta-\mu}{2})$ such that $h_2(\delta) < 0$. In other words, there exists $\rho > \mu$ such that the minimal value of $h_1(t)$ is nonnegative. The assertion (2.12) is proved. ∎

3. Convergence rate of ADM with LQP. Now we start to derive a worst-case $O(1/t)$ convergence rate for the scheme (1.10)–(1.12) where the Glowinski relaxation factor $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is adopted.

3.1. More notation. To analyze the convergence rate of ADM with LQP, it is better to introduce some more notation in order to present the proof more succinctly. First, we define some matrices. Let

$$P = \begin{pmatrix} R & 0 \\ 0 & S \end{pmatrix}, \quad Q = \begin{pmatrix} B^T HB & 0 \\ -B & H^{-1} \end{pmatrix}, \tag{3.1}$$

$$M_\gamma = \begin{pmatrix} I_m & 0 \\ -\gamma HB & \gamma I_l \end{pmatrix}, \quad \text{and} \quad D_\gamma = \begin{pmatrix} B^T HB & 0 \\ 0 & \frac{1}{\gamma} H^{-1} \end{pmatrix}, \tag{3.2}$$

where I_m and I_l are the identity matrices in $\mathcal{R}^m$ and $\mathcal{R}^l$, respectively. By elementary manipulation, we have the relationship

$$Q = D_\gamma M_\gamma. \tag{3.3}$$

Then, as in [23], with the sequence $\{w^k\}$ generated by the scheme (1.10)–(1.12), our analysis needs a new sequence defined by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - H(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{3.4}$$

and we call $\{\tilde{w}^k\}$ the accompanying sequence of $\{w^k\}$. In fact, our convergence rate results (see Theorem 3.4 and 4.2) are carried out for the sequence $\{\tilde{w}^k\}$. Obviously, when $\gamma = 1$, $\tilde{w}^k$ defined in (3.4) differs from w^{k+1} generated by (1.10)–(1.12) only in the update of the Lagrange multiplier. More precisely, we have $\|\tilde{w}^k - w^{k+1}\| = \|HB(y^{k+1} - y^k)\|$. Thus, based on the fact that $\lim_{k\to\infty} \|y^k - y^{k+1}\| = 0$ (see (4.1) in [36]) and the convergence of $\{w^k\}$ proved in [36], we can immediately derive the convergence of the sequence $\{\tilde{w}^k\}$.

We also need the notation $u^k = (x^k, y^k)$, $v^k = (y^k, \lambda^k)$, $\tilde{v}^k = (\tilde{y}^k, \tilde{\lambda}^k)$, $\mathcal{V} = \mathcal{Y} \times \Re^m$, and $\mathcal{V}^* := \{v^* = (y^*, \lambda^*) \,|\, w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*\}$. Based on (3.1) and (3.4), we easily have a relationship

$$v^{k+1} = v^k - M_\gamma(v^k - \tilde{v}^k), \tag{3.5}$$

which will be used often later.

3.2. Some lemmas. We reiterate that the convergence rate will be established for the sequence $\{\tilde{w}^k\}$. Therefore, because of Theorem 2.1, it is important to estimate a unified lower bound of the term $(x - \tilde{x}^k)^T f(\tilde{x}^k) + (y - \tilde{y}^k)^T g(\tilde{y}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)$ for all $w \in \mathcal{W}$. On the other hand, note that the convergence rate to be established is in an ergodic sense (as in [23]). Thus, it is convenient to represent this unified lower bound in the form of some quadratic functions, without any crossed term involving both the current iterate w^{k+1} and the previous iterate w^k (see (3.9)). Certainly, the accompanying sequence $\{\tilde{w}^k\}$ should not appear in this lower bound when we estimate the convergence rate for ADM with LQP. In short, the roadmap for deriving a worst-case $O(1/t)$ convergence rate for ADM with LQP can be summarized as follows:

(1) Estimate a unified lower bound of the term $(x-\tilde{x}^k)^T f(\tilde{x}^k)+(y-\tilde{y}^k)^T g(\tilde{y}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)$ for all $w \in \mathcal{W}$. This is completed by Lemma 3.1.

(2) Refine the lower bound found in (1) by removing its crossed terms and $\tilde{w}^k$-related terms. A unified lower bound in the form of only quadratic terms is thus found. This is completed by Lemma 3.3.

(3) Derive a worst-case $O(1/t)$ convergence rate in an ergodic sense for ADM with LQP based on the characterization stated in Theorem 2.1. This is completed by Theorem 3.4.

In this subsection we complete the first two tasks, and in the next subsection the third task is completed.

Lemma 3.1. *Let the sequence* $\{w^k\}$ *be generated by the scheme* (1.10)–(1.12)*, let the accompanying sequence* $\{\tilde{w}^k\}$ *be defined by* (3.4)*, and let the matrices* P*,* D_γ *be given in* (3.1)–(3.2)*. Then we have*

$$\begin{aligned}
&(x-\tilde{x}^k)^T f(\tilde{x}^k)+(y-\tilde{y}^k)^T g(\tilde{y}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)\\
&\quad\geq \frac{1+\mu}{2}\left(\|u-u^{k+1}\|_P^2-\|u-u^k\|_P^2\right)+\frac{1-\mu}{2}\|u^k-u^{k+1}\|_P^2\\
&\qquad+\frac{1}{2}\left(\|v-v^{k+1}\|_{D_\gamma}^2-\|v-v^k\|_{D_\gamma}^2\right)\\
&\qquad+\frac{1}{2}\|B(y^k-\tilde{y}^k)\|_H^2+(By^k-By^{k+1})^T H(Ax^{k+1}+By^{k+1}-b)\\
&\qquad+\frac{1}{2}(2-\gamma)\|Ax^{k+1}+By^{k+1}-b\|_H^2 \quad \forall w\in\mathcal{W}.
\end{aligned} \tag{3.6}$$

Proof. See Appendix A. ∎

In (3.6), a unified lower bound of the term $(x-\tilde{x}^k)^T f(\tilde{x}^k)+(y-\tilde{y}^k)^T g(\tilde{y}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)$ for all $w \in \mathcal{W}$ is founded. When we analyze the convergence rate of $\tilde{w}^k$ in an ergodic sense, the first several quadratic terms in the right-hand side of (3.6) can be easily handled. Meanwhile, we should further relax the $\tilde{w}^k$-related term $\|B(y^k-\tilde{y}^k)\|_H^2$ and the crossed term $(By^k-By^{k+1})^T H(Ax^{k+1}+By^{k+1}-b)$ in terms of some quadratic terms. Below, we show that the summation of the last three terms of the right-hand side of (3.6) (denoted by Δ_γ^k), i.e.,

$$\begin{aligned}
\Delta_\gamma^k &:= \frac{1}{2}\|B(y^k-\tilde{y}^k)\|_H^2+(By^k-By^{k+1})^T H(Ax^{k+1}+By^{k+1}-b)\\
&\quad+\frac{1}{2}(2-\gamma)\|Ax^{k+1}+By^{k+1}-b\|_H^2,
\end{aligned} \tag{3.7}$$

can be further bounded by some quadratic terms in the absence of $\tilde{y}^k$ and crossed terms. The relationship (3.5) is crucial for this possibility.

Lemma 3.2. *Let* $\{w^k\}$ *be generated by the scheme* (1.10)–(1.12)*, and let* Δ_γ^k *be defined in* (3.7)*. Then, when* $\mu \in (0, 0.2)$*, there exist* $\eta \in (0, \frac{1-\mu}{2})$ *and* $\rho > \mu$ *such that*

$$\begin{aligned}
\Delta_\gamma^k &\geq \rho\{\|y^k-y^{k+1}\|_S^2-\|y^{k-1}-y^k\|_S^2\}-\eta\|y^k-y^{k+1}\|_S^2\\
&\quad+\frac{(T-\gamma)}{2}\left[\|Ax^{k+1}+By^{k+1}-b\|_H^2-\|Ax^k+By^k-b\|_H^2\right]\\
&\quad+\frac{1}{2}\left[\left(1-\frac{(1-\gamma)^2}{T-\gamma}\right)\|B(y^k-y^{k+1})\|_H^2+(2-T)\|Ax^{k+1}+By^{k+1}-b\|_H^2\right],
\end{aligned} \tag{3.8}$$

where $T=\frac{1}{3}(\gamma^2-\gamma+5)$*.*

Proof. See Appendix B. ∎

Now, with Lemmas 3.1 and 3.2, we can refine the lower bound founded in (3.6) in terms of only quadratic terms without the appearance of $\tilde{y}^k$. This nice form of the lower bound provides the possibility of analyzing the convergence rate for ADM with LQP in an ergodic sense.

LEMMA 3.3. *Let $\{w^k\}$ be generated by the scheme* (1.10)–(1.12), *and let* $T = \frac{1}{3}(\gamma^2 - \gamma + 5)$. *Then, when $\mu \in (0, 0.2)$, there exist $\eta \in (0, \frac{1-\mu}{2})$ and $\rho > \mu$ such that*

$$\begin{aligned}
&(x-\tilde{x}^k)^T f(\tilde{x}^k) + (y-\tilde{y}^k)^T g(\tilde{y}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k) \\
&\quad \geq \rho(\|y^k - y^{k+1}\|_S^2 - \|y^{k-1} - y^k\|_S^2) + \frac{1+\mu}{2}\left(\|u^{k+1}-u\|_P^2 - \|u^k - u\|_P^2\right) \\
&\qquad + \frac{1}{2}\Big[\|v - v^{k+1}\|_{D_\gamma} - \|v - v^k\|_{D_\gamma}\Big] \\
&\qquad + \frac{1}{2}(T-\gamma)\left(\|Ax^{k+1} + By^{k+1} - b\|_H^2 - \|Ax^k + By^k - b\|_H^2\right) \quad \forall w \in \mathcal{W},
\end{aligned} \tag{3.9}$$

Proof. Substituting (3.8) into (3.6) and rearranging the resulting terms, we have that

$$\begin{aligned}
&(x-\tilde{x}^k)^T f(\tilde{x}^k) + (y-\tilde{y}^k)^T g(\tilde{y}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k) \\
&\quad \geq \rho(\|y^k - y^{k+1}\|_S^2 - \|y^{k-1} - y^k\|_S^2) + \frac{1+\mu}{2}\left(\|u^{k+1}-u\|_P^2 - \|u^k - u\|_P^2\right) \\
&\qquad + \frac{1}{2}\Big[\|v - v^{k+1}\|_{D_\gamma} - \|v - v^k\|_{D_\gamma}\Big] \\
&\qquad + \frac{1}{2}(T-\gamma)\left(\|Ax^{k+1} + By^{k+1} - b\|_H^2 - \|Ax^k + By^k - b\|_H^2\right) \\
&\qquad + \frac{1-\mu}{2}\|x^k - x^{k+1}\|_R^2 + \Big(\frac{1-\mu}{2} - \eta\Big)\|y^k - y^{k+1}\|_S^2 \\
&\qquad + \frac{1}{2}\Big[\Big(1 - \frac{(1-\gamma)^2}{T-\gamma}\Big)\|B(y^k - y^{k+1})\|_H^2 + (2-T)\|Ax^{k+1} + By^{k+1} - b\|_H^2\Big]
\end{aligned} \tag{3.10}$$

holds for any $w \in \mathcal{W}$. Recalling Lemma 2.4 and the fact $\eta \in (0, \frac{1-\mu}{2})$, the numbers $(\frac{1-\mu}{2} - \eta)$ and $(1 - \frac{(1-\gamma)^2}{T-\gamma})$ are both positive. Thus, we have that

$$\begin{aligned}
0 &\leq \frac{1-\mu}{2}\|x^k - x^{k+1}\|_R^2 + \Big(\frac{1-\mu}{2} - \eta\Big)\|y^k - y^{k+1}\|_S^2 \\
&\quad + \frac{1}{2}\Big[\Big(1 - \frac{(1-\gamma)^2}{T-\gamma}\Big)\|B(y^k - y^{k+1})\|_H^2 + (2-T)\|Ax^{k+1} + By^{k+1} - b\|_H^2\Big]
\end{aligned} \tag{3.11}$$

holds for any $w \in \mathcal{W}$. Then, substituting (3.11) into (3.10), the assertion (3.9) is proved. ∎

Note that the unified lower bound found in (3.9) are fully representable by quadratic terms in a recursive fashion. Thus, by an ergodic operation it is possible to find a unified bound for

$$\sup_{w \in \mathcal{W}} \{(\tilde{x}_t - x)^T f(x) + (\tilde{y}_t - y)^T g(y) + (\tilde{w}_t - w)^T F(w)\},$$

from which a worst-case convergence rate of ADM with LQP can be derived.

3.3. Convergence rate. With the results proved in section 3.2, we are able to derive a worst-case $O(1/t)$ convergence rate in an ergodic sense for the scheme (1.10)–(1.11).

THEOREM 3.4. *Assume that the two functions $f(x)$ and $g(y)$ in the variational inequality (1.1)–(1.2) are continuous and monotone. Let $\{w^k\}$ be the sequence generated by the scheme (1.10)–(1.12), and let the accompanying sequence $\{\tilde{w}^k\}$ be defined in (3.4), the matrices P, Q and D_γ be given in (3.1)–(3.2), and $T = \frac{1}{3}(\gamma^2 - 4\gamma + 5)$. For any integer number $t > 0$, let $\tilde{w}_t$ be defined by*

$$\tilde{w}_t = \frac{1}{t}\sum_{k=1}^{t} \tilde{w}^k. \tag{3.12}$$

Then, when $\mu \in (0, 0.2)$, there exists $\rho > \mu$ such that $\tilde{w}_t \in \Omega$ and

$$\begin{aligned}&(\tilde{x}_t - x)^T f(x) + (\tilde{y}_t - y)^T g(y) + (\tilde{w}_t - w)^T F(w)\\ &\quad\le \frac{1}{2t}\{(T-\gamma)\|Ax^1 + By^1 - b\|_H^2 + 2\rho\|y^0 - y^1\|_S^2\\ &\quad\quad + (1+\mu)\|u - u^0\|_P^2 + \|v - v^1\|_{D_\gamma}\} \quad \forall w \in \mathcal{W}.\end{aligned} \tag{3.13}$$

Proof. First, using the fact (3.9) and the monotonicity of f and g, we get

$$\begin{aligned}&(x - \tilde{x}^k)^T f(x) + (y - \tilde{y}^k)^T g(y) + (w - \tilde{w}^k)^T F(w)\\ &\quad\ge \frac{1}{2}(T-\gamma)\left(\|Ax^{k+1} + By^{k+1} - b\|_H^2 - \|Ax^k + By^k - b\|_H^2\right)\\ &\quad\quad + \rho(\|y^k - y^{k+1}\|_S^2 - \|y^{k-1} - y^k\|_S^2) + \frac{1+\mu}{2}\left(\|u^{k+1} - u\|_P^2 - \|u^k - u\|_P^2\right)\\ &\quad\quad + \frac{1}{2}\Big(\|v - v^{k+1}\|_{D_\gamma} - \|v - v^k\|_{D_\gamma}\Big) \quad \forall w \in \mathcal{W}.\end{aligned} \tag{3.14}$$

Summing the inequality (3.14) over $k = 1, \ldots, t$, we obtain

$$\begin{aligned}&\left(tx - \sum_{k=1}^{t}\tilde{x}^k\right)^T f(x) + \left(ty - \sum_{k=1}^{t}\tilde{y}^k\right)^T g(y) + \left(tw - \sum_{k=1}^{t}\tilde{w}^k\right)^T F(w)\\ &\quad\ge -\frac{1}{2}\Big((T-\gamma)\|Ax^1 + By^1 - b\|_H^2 + 2\rho\|y^0 - y^1\|_S^2\\ &\quad\quad + (1+\mu)\|u^1 - u\|_P^2 + \|v - v^1\|_{D_\gamma}^2\Big) \quad \forall w \in \mathcal{W},\end{aligned} \tag{3.15}$$

which implies the assertion (3.13) immediately. The proof is complete. ∎

For any given compact set $\mathcal{W}^0 \subset \mathcal{W}$, let

$$d_\gamma := \sup_{w\in\mathcal{W}^0}\{(T-\gamma)\|Ax^1 + By^1 - b\|_H^2 + 2\rho\|y^0 - y^1\|_S^2 + (1+\mu)\|u - u^0\|_P^2 + \|v - v^1\|_{D_\gamma}\}.$$

Then, after t iterations of the scheme (1.10)–(1.12), the point $\tilde{w}_t$ defined in (3.12) satisfies

$$\sup_{w\in\mathcal{W}^0}\{(\tilde{x}_t - x)^T f(x) + (\tilde{y}_t - y)^T g(y) + (\tilde{w}_t - w)^T F(w)\} \le \frac{d_\gamma}{2t},$$

which means $\tilde{w}_t$ is an approximate solution of VI$(\mathcal{W}, F)$ with accuracy $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate for (1.10)–(1.12) is established in an ergodic sense.

4. The special case where $\gamma = 1$. In section 3, we establish a worst-case $O(1/t)$ convergence rate for the general case of ADM with LQP where $\gamma \in (0, \frac{1+\sqrt{5}}{2})$. The restriction $\mu \in (0, 0.2)$ is required in order to ensure the derived $O(1/t)$ convergence rate. In this section, we show that for the special case of ADM with LQP where $\gamma = 1$ (which is commonly used in the literature), the restriction of μ can be relaxed to $\mu \in (0, 1)$ and the establishment of a worst-case $O(1/t)$ convergence rate is much easier for this special case.

We first show a unified bound of the term $(x - \tilde{x}^k)^T f(\tilde{x}^k) + (y - \tilde{y}^k)^T g(\tilde{y}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)$ for all $w \in \mathcal{W}$. Because of the simplicity of $\gamma = 1$, this lower bound can be found easily.

LEMMA 4.1. *Let the sequence* $\{w^k\}$ *be generated by the scheme* (1.10)–(1.12), *where* $\gamma = 1$ *and* $\mu \in (0, 1)$, *let the accompanying sequence* $\{\tilde{w}^k\}$ *be defined by* (3.4), *and let the matrices* P, Q *and* D_1 *be given in* (3.1)–(3.2). *Then, we have*

$$
\begin{aligned}
&(x - \tilde{x}^k)^T f(\tilde{x}^k) + (y - \tilde{y}^k)^T g(\tilde{y}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\
&\quad \geq \frac{1+\mu}{2} \left(\|u - u^{k+1}\|_P^2 - \|u - u^k\|_P^2 \right) \\
&\quad + \frac{1}{2} \left(\|v - v^{k+1}\|_{D_1}^2 - \|v - v^k\|_{D_1}^2 \right) \quad \forall w \in \mathcal{W}.
\end{aligned} \tag{4.1}
$$

Proof. Similarly to the proof of Lemma 3.1, we can derive

$$
\begin{aligned}
&(x - \tilde{x}^k)^T f(\tilde{x}^k) + (y - \tilde{y}^k)^T g(\tilde{y}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\
&\quad \geq \frac{1+\mu}{2} (\|u^{k+1} - u\|_P^2 - \|u^k - u\|_P^2) + \frac{1-\mu}{2} \|u^k - u^{k+1}\|_P^2 \\
&\quad + \frac{1}{2} (\|v - v^{k+1}\|_{D_1}^2 - \|v - v^k\|_{D_1}^2) \\
&\quad + \frac{1}{2} (\|v^k - \tilde{v}^k\|_{D_1}^2 - \|v^{k+1} - \tilde{v}^k\|_{D_1}^2) \quad \forall w \in \mathcal{W}.
\end{aligned} \tag{4.2}
$$

Since $\gamma = 1$, we have $\lambda^k - \tilde{\lambda}^k = H(Ax^{k+1} + By^k - b)$ (see (3.4)). Setting $\gamma = 1$ in (A.7) (see Appendix A), we have

$$
\begin{aligned}
&\|v^k - \tilde{v}^k\|_{D_1}^2 - \|v^{k+1} - \tilde{v}^k\|_{D_1}^2 \\
&\quad = \|B(y^k - \tilde{y}^k)\|_H^2 + 2(By^k - By^{k+1})^T H(Ax^{k+1} + By^{k+1} - b) \\
&\quad\quad + \|Ax^{k+1} + By^{k+1} - b\|_H^2 \\
&\quad = \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \\
&\quad \geq 0.
\end{aligned} \tag{4.3}
$$

Substituting (4.3) into (4.2) and ignoring the positive term $\frac{1-\mu}{2}\|u^k - u^{k+1}\|_P^2$ (because $\mu \in (0, 1)$), the assertion (4.1) is obtained. ∎

Based on the fact (4.1), a worst-case $O(1/t)$ convergence rate can be established for ADM with LQP, where $\gamma = 1$ and $\mu \in (0, 1)$.

THEOREM 4.2. *Assume that the two functions* $f(x)$ *and* $g(y)$ *in the variational inequality* (1.1)–(1.2) *are continuous and monotone. Let* $\{w^k\}$ *be the sequence generated by the scheme* (1.10)–(1.12), *where* $\gamma = 1$ *and* $\mu \in (0, 1)$, *let the accompanying sequence* $\{\tilde{w}^k\}$ *be defined in* (3.4), *and let the matrices* P *and* D_1 *be given in* (3.1)–(3.2), *respectively. For any integer number* $t > 0$, *let* $\tilde{w}_t$ *be defined by*

$$
\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \tag{4.4}
$$

Then, $\tilde{w}_t \in \mathcal{W}$ *and*

$$(\tilde{x}_t - x)^T f(x) + (\tilde{y}_t - y)^T g(y) + (\tilde{w}_t - w)^T F(w) \\ \leq \frac{1}{2(t+1)}(\|v - v^0\|_{D_1}^2 + (1+\mu)\|u - u^0\|_P^2) \quad \forall w \in \mathcal{W}. \tag{4.5}$$

Proof. First, because of (3.4) and $w^k \in \mathcal{W}$, it holds that $\tilde{w}^k \in \mathcal{W}$ for all $k \geq 0$. Thus, together with the convexity of $\mathcal{R}_+^n$ and $\mathcal{R}_+^m$, (4.4) implies that $\tilde{w}_t \in \mathcal{W}$.

Second, inequality (4.1) implies that

$$(x - \tilde{x}^k)^T f(x) + (y - \tilde{y}^k)^T g(y) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_{D_1}^2 + \frac{1+\mu}{2}\|u - u^k\|_P^2 \\ \geq \frac{1}{2}\|v - v^{k+1}\|_D^2 + \frac{1+\mu}{2}\|u - u^{k+1}\|_P^2 \quad \forall w \in \mathcal{W}. \tag{4.6}$$

Summing inequality (4.6) over $k = 0, 1, \ldots, t$, for any $w \in \mathcal{W}$ it holds that

$$\left((t+1)x - \sum_{k=0}^{t} \tilde{x}^k\right)^T f(x) + \left((t+1)y - \sum_{k=0}^{t} \tilde{y}^k\right)^T g(y) + \left((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\right)^T F(w) \\ \geq -\left(\frac{1}{2}\|v - v^0\|_D^2 + \frac{1+\mu}{2}\|u - u^0\|_P^2\right).$$

Use the notation of $\tilde{w}_t$ in (4.4), this can be written as

$$(\tilde{x}_t - x)^T f(x) + (\tilde{y}_t - y)^T g(y) + (\tilde{w}_t - w)^T F(w) \\ \leq \frac{1}{2(t+1)}\left(\|v - v^0\|_D^2 + (1+\mu)\|u - u^0\|_P^2\right) \quad \forall w \in \mathcal{W}.$$

Thus, the assertion (4.5) is proved, and the proof is complete. ▯

5. Conclusions. This paper derives a worst-case $O(1/t)$ convergence rate in an ergodic sense for the blend of Douglas–Rachford alternating direction method of multipliers (ADM) and the logarithmic-quadratic proximal (LQP) regularization where the Glowinski relaxation factor is adopted, in the context of a class of variational inequalities. This is a theoretical support to the combination of ADM and LQP, and it is interesting to apply the analytic framework in this paper to analyze the convergence rate for some other methods with interior-point-oriented regularization. Interesting research in the future also will include deriving the $O(1/t)$ convergence rate in nonergodic senses for ADM with LQP.

Appendix A. Proof of Lemma 3.1.

Proof. Applying Lemma 2.3 and the notation in (3.4), inequalities (2.7) and (2.8) can be respectively written as

$$(x - x^{k+1})^T f(x^{k+1}) + (x - \tilde{x}^k)^T \{-A^T \tilde{\lambda}^k\} \\ \geq \frac{1+\mu}{2}(\|x^{k+1} - x\|_R^2 - \|x^k - x\|_R^2) + \frac{1-\mu}{2}\|x^k - x^{k+1}\|_R^2 \quad \forall x \in \mathcal{R}_+^n \tag{A.1}$$

and

$$(y - y^{k+1})^T g(y^{k+1}) + (y - \tilde{y}^k)^T \{-B^T \tilde{\lambda}^k + B^T H B(\tilde{y}^k - y^k)\} \\ \geq \frac{1+\mu}{2}(\|y^{k+1} - y\|_S^2 - \|y^k - y\|_S^2) + \frac{1-\mu}{2}\|y^k - y^{k+1}\|_S^2 \quad \forall y \in \mathcal{R}_+^m. \tag{A.2}$$

In addition, it follows from (3.4) that

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + H^{-1}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{A.3}$$

Combining (A.1), (A.2), and (A.3), and by simple manipulation, we get

$$\begin{aligned}&(x-\tilde{x}^k)^T f(\tilde{x}^k) + (y-\tilde{y}^k)^T g(\tilde{y}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k)\\ &\quad \geq \frac{1+\mu}{2}(\|u^{k+1}-u\|_P^2 - \|u^k-u\|_P^2) + \frac{1-\mu}{2}\|u^k-u^{k+1}\|_P^2\\ &\quad + (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) \quad \forall w \in \mathcal{W}.\end{aligned} \tag{A.4}$$

On the other hand, recall $Q = D_\gamma M_\gamma$ (see (3.3)) and $M_\gamma(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (see (3.5)). By using Lemma 2.5, we get

$$\begin{aligned}(v-\tilde{v}^k)^T D_\gamma(v^k - v^{k+1}) &= \frac{1}{2}\big(\|v-v^{k+1}\|_{D_\gamma}^2 - \|v-v^k\|_{D_\gamma}^2\big)\\ &+\frac{1}{2}(\|v^k-\tilde{v}^k\|_{D_\gamma}^2 - \|v^{k+1}-\tilde{v}^k\|_{D_\gamma}^2) \quad \forall v \in \mathcal{V}.\end{aligned} \tag{A.5}$$

Then we obtain

$$\begin{aligned}&(x-\tilde{x}^k)^T f(\tilde{x}^k) + (y-\tilde{y}^k)^T g(\tilde{y}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k)\\ &\quad \geq \frac{1+\mu}{2}\left(\|u-u^{k+1}\|_P^2 - \|u-u^k\|_P^2\right) + \frac{1-\mu}{2}\|u^k-u^{k+1}\|_P^2\\ &\quad +\frac{1}{2}\left(\|v-v^{k+1}\|_{D_\gamma}^2 - \|v-v^k\|_{D_\gamma}^2\right)\\ &\quad +\frac{1}{2}\left(\|\tilde{v}^k-v^k\|_{D_\gamma}^2 - \|\tilde{v}^k-v^{k+1}\|_{D_\gamma}^2\right), \quad w \in \mathcal{W}.\end{aligned} \tag{A.6}$$

On the other hand, by using (3.5) and the definitions of D_γ and M_γ in (3.2), we have

$$\begin{aligned}&\|\tilde{v}^k - v^k\|_{D_\gamma}^2 - \|\tilde{v}^k - v^{k+1}\|_{D_\gamma}^2\\ &\quad = \|v^k-\tilde{v}^k\|_{D_\gamma}^2 - \|(v^k-\tilde{v}^k) - M_\gamma(v^k-\tilde{v}^k)\|_{D_\gamma}^2\\ &\quad = 2(v^k-\tilde{v}^k)^T D_\gamma M_\gamma(v^k-\tilde{v}^k) - \|M_\gamma(v^k-\tilde{v}^k)\|_{D_\gamma}^2\\ &\quad = 2\begin{pmatrix} y^k-\tilde{y}^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}^T \begin{pmatrix} B^THB & 0\\ -B & H^{-1}\end{pmatrix}\begin{pmatrix} y^k-\tilde{y}^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}\\ &\qquad -\left\|\begin{pmatrix} H^{1/2}B & 0\\ -\sqrt{\gamma}H^{1/2}B & \sqrt{\gamma}H^{-1/2}\end{pmatrix}\begin{pmatrix} y^k-\tilde{y}^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}\right\|^2\\ &\quad = \|B(y^k-\tilde{y}^k)\|_H^2 + 2(\lambda^k-\tilde{\lambda}^k)^T H^{-1}\{(\lambda^k-\tilde{\lambda}^k) - HB(y^k-\tilde{y}^k)\}\\ &\qquad -\gamma\|(\lambda^k-\tilde{\lambda}^k) - HB(y^k-\tilde{y}^k)\|_{H^{-1}}^2.\end{aligned} \tag{A.7}$$

In the right-hand-side of (A.7), by substituting $\tilde{y}^k = y^{k+1}$, $\lambda^k - \tilde{\lambda}^k = H(Ax^{k+1} + By^k - b)$, and

$$(\lambda^k - \tilde{\lambda}^k) - HB(y^k - \tilde{y}^k) = H(Ax^{k+1} + By^{k+1} - b),$$

we obtain

$$\begin{aligned}&\|v^k-\tilde v^k\|_{D_\gamma}^2-\|v^{k+1}-\tilde v^k\|_{D_\gamma}^2\\&\quad=\|B(y^k-y^{k+1})\|_H^2+2(Ax^{k+1}+By^k-b)^TH(Ax^{k+1}+By^{k+1}-b)\\&\qquad-\gamma\|Ax^{k+1}+By^{k+1}-b\|_H^2\\&\quad=\|B(y^k-y^{k+1})\|_H^2+2(By^k-By^{k+1})^TH(Ax^{k+1}+By^{k+1}-b)\\&\qquad+(2-\gamma)\|Ax^{k+1}+By^{k+1}-b\|_H^2.\end{aligned}$$

Thus, we have

$$\begin{aligned}&\|v^k-\tilde v^k\|_{D_\gamma}^2-\|v^{k+1}-\tilde v^k\|_{D_\gamma}^2\\&\quad=\|B(y^k-y^{k+1})\|_H^2+2(By^k-By^{k+1})^TH(Ax^{k+1}+By^{k+1}-b)\\&\qquad+(2-\gamma)\|Ax^{k+1}+By^{k+1}-b\|_H^2.\end{aligned}\tag{A.8}$$

Substituting (A.8) into (A.6), the assertion (3.6) is proved. □

Appendix B. Proof of Lemma 3.2.

Proof. Let $\{y^k\}$ and $\{u^k\}$ be generated by the scheme (1.10)–(1.12). First, it follows from Lemma 2.6 that when $\mu\in(0,0.2)$, there exist $\eta\in(0,\frac{1-\mu}{2})$ and $\rho>\mu$ such that

$$\begin{aligned}&\|y^k-y^{k+1}\|_S^2-\mu\|y^{k-1}-y^k\|_S^2-(1+\mu)(y^k-y^{k+1})^TS(y^{k-1}-y^k)\\&\quad\ge\rho\|y^k-y^{k+1}\|_S^2-\rho\|y^{k-1}-y^k\|_S^2-\eta\|y^k-y^{k+1}\|_S^2.\end{aligned}\tag{B.1}$$

Second, applying Lemma 2.2 to (1.11) (namely, setting $P=S$, $a^k=y^k$, $a=y^{k+1}$, $q(\cdot)=g_k(\cdot)$, and $b=y^k$ in (2.6)) and using

$$g_k(y^{k+1})\overset{(2.10)}{=}g(y^{k+1})-B^T[\lambda^k-H(Ax^{k+1}+By^{k+1}-b)],$$

we obtain

$$(y^k-y^{k+1})^T\left\{g(y^{k+1})-B^T\left[\lambda^k-H(Ax^{k+1}+By^{k+1}-b)\right]\right\}\ge\|y^k-y^{k+1}\|_S^2.\tag{B.2}$$

Similarly, applying Lemma 2.2 to (1.11) at the (k−1)th iteration (namely, by setting $P=S$, $a^k=y^{k-1}$, $a=y^k$, $q(\cdot)=g_{k-1}(\cdot)$, and $b=y^{k+1}$ in (2.6)) with

$$g_{k-1}(y^k)\overset{(2.10)}{=}g(y^k)-B^T[\lambda^{k-1}-H(Ax^k+By^k-b)],$$

we have

$$\begin{aligned}&(y^{k+1}-y^k)^T\left\{g(y^k)-B^T\left[\lambda^{k-1}-H(Ax^k+By^k-b)\right]\right\}\\&\quad\ge\tfrac{1+\mu}{2}\left(\|y^k-y^{k+1}\|_S^2-\|y^{k-1}-y^{k+1}\|_S^2\right)+\tfrac{1-\mu}{2}\|y^{k-1}-y^k\|_S^2.\end{aligned}\tag{B.3}$$

Adding (B.2) and (B.3), and using the monotonicity of the operator g, we get

$$\begin{aligned}&(y^k-y^{k+1})^T\left\{B^T\left[(\lambda^{k-1}-\lambda^k)-H(Ax^k+By^k-b)\right.\right.\\&\qquad\left.\left.+H(Ax^{k+1}+By^{k+1}-b)\right]\right\}\\&\quad\ge\tfrac{3+\mu}{2}\|y^k-y^{k+1}\|_S^2-\tfrac{1+\mu}{2}\|y^{k-1}-y^{k+1}\|_S^2+\tfrac{1-\mu}{2}\|y^{k-1}-y^k\|_S^2.\end{aligned}\tag{B.4}$$

Recall $\lambda^{k-1} - \lambda^k = \gamma H(Ax^k + By^k - b)$ (see (3.4)). We thus obtain the assertion

$$\begin{aligned} &B(y^k - y^{k+1})^T H(Ax^{k+1} + By^{k+1} - b) \\ &\quad \geq (1-\gamma)(Ax^k + By^k - b)^T HB(y^k - y^{k+1}) + \|y^k - y^{k+1}\|_S^2 \\ &\qquad - \mu\|y^{k-1} - y^k\|_S^2 - (1+\mu)(y^k - y^{k+1})^T S(y^{k-1} - y^k). \end{aligned} \tag{B.5}$$

Then, by combining (B.1) and (B.5), we get

$$\begin{aligned} &(Ax^{k+1} + By^{k+1} - b)^T HB(y^k - y^{k+1}) \\ &\quad \geq (1-\gamma)(Ax^k + By^k - b)^T HB(y^k - y^{k+1}) + \|y^k - y^{k+1}\|_S^2 \\ &\qquad - \mu\|y^{k-1} - y^k\|_S^2 - (1+\mu)(y^k - y^{k+1})^T S(y^{k-1} - y^k) \\ &\quad \geq (1-\gamma)(Ax^k + By^k - b)^T HB(y^k - y^{k+1}) + \rho\|y^k \\ &\qquad - y^{k+1}\|_S^2 - \rho\|y^{k-1} - y^k\|_S^2 - \eta\|y^k - y^{k+1}\|_S^2. \end{aligned} \tag{B.6}$$

Thus, we have the conclusion

$$\begin{aligned} \Delta_\gamma^k \geq &\frac{1}{2}\left[\|B(y^k - y^{k+1})\|_H^2 + 2(1-\gamma)(Ax^k + By^k - b)^T HB(y^k - y^{k+1})\right] \\ &+ \frac{1}{2}(2-\gamma)\|Ax^{k+1} + By^{k+1} - b\|_H^2 + \rho\{\|y^k - y^{k+1}\|_S^2 - \|y^{k-1} - y^k\|_S^2\} \\ &- \eta\|y^k - y^{k+1}\|_S^2. \end{aligned} \tag{B.7}$$

On the other hand, recall the assertions in Lemma 2.4. In view of the Cauchy–Schwarz inequality, we instantly have

$$\begin{aligned} &2(1-\gamma)(Ax^k + By^k - b)^T H(By^k - By^{k+1}) \\ &\quad \geq -(T-\gamma)\|Ax^k + By^k - b\|_H^2 - \frac{(1-\gamma)^2}{T-\gamma}\|B(y^k - y^{k+1})\|_H^2. \end{aligned} \tag{B.8}$$

Combining (B.8) with the first term in the left-hand-side of (B.6) and by simple manipulation, we obtain

$$\begin{aligned} \|B(y^k - y^{k+1})\|_H^2 &+ 2(1-\gamma)(Ax^k + By^k - b)^T HB(y^k - y^{k+1}) \\ &+ (2-\gamma)\|Ax^{k+1} + By^{k+1} - b\|_H^2 \\ &\quad \geq \left(1 - \frac{(1-\gamma)^2}{T-\gamma}\right)\|B(y^k - y^{k+1})\|_H^2 + (2-T)\|Ax^{k+1} + By^{k+1} - b\|_H^2 \\ &+ (T-\gamma)\big(\|Ax^{k+1} + By^{k+1} - b\|_H^2 - \|Ax^k + By^k - b\|_H^2\big). \end{aligned} \tag{B.9}$$

Substituting (B.9) into (B.7), the assertion (3.8) is derived. The proof is complete. ∎

Appendix C. Global convergence of ADM with LQP with a Glowinski relaxation factor.

Since the convergence of ADM with LQP regularization is established in [36] only for the case where $\gamma = 1$ and its extension to the case with a general Glowinski relaxation factor $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is not trivial, we provide here the complete proof for the global convergence of (1.10)–(1.12). A key tool for the proof is inequality (3.9) in Lemma 3.3. We start the analysis with a theorem.

THEOREM C.1. *Let $\{w^k\}$ be generated by the scheme (1.10)–(1.12), let the matrix D_γ be defined in (3.2), and let $T = \frac{1}{3}(\gamma^2 - \gamma + 5)$. Then, when the positive scalar $\mu \in (0, 0.2)$, there exist scalars $\eta \in (0, \frac{1-\mu}{2})$ and $\rho > \mu$ such that*

$$\begin{aligned} &\|w^{k+1} - w^*\|_{G_\gamma}^2 + 2\rho\|y^k - y^{k+1}\|_S^2 \\ &\quad \le \|w^k - w^*\|_{G_\gamma}^2 + 2\rho\|y^{k-1} - y^k\|_S^2 - \Big(1 - \frac{(1-\gamma)^2}{T-\gamma}\Big)\|B(y^k - y^{k+1})\|_H^2 \\ &\qquad - (2-T)\|Ax^{k+1} + By^{k+1} - b\|_H^2 - (1-\mu)\|x^k - x^{k+1}\|_R^2 \\ &\qquad - (1-\mu-2\eta)\|y^k - y^{k+1}\|_S^2 \quad \forall w^* \in \mathcal{W}^*, \end{aligned} \tag{C.1}$$

where

$$G_\gamma = \begin{pmatrix} (1+\mu)R & & \\ & (1+\mu)S + B^T H B & \\ & & (\gamma H)^{-1} \end{pmatrix} + (T-\gamma)\begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} H(A, B, 0). \tag{C.2}$$

Proof. Setting $x = x^*$, $y = y^*$, and $w = w^*$ in (3.10), we get

$$\begin{aligned} &(x^* - \tilde{x}^k)^T f(\tilde{x}^k) + (y^* - \tilde{y}^k)^T g(\tilde{y}^k) + (w^* - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\quad \ge \frac{1}{2}(T-\gamma)\left(\|Ax^{k+1} + By^{k+1} - b\|_H^2 - \|Ax^k + By^k - b\|_H^2\right) \\ &\qquad + \rho(\|y^k - y^{k+1}\|_S^2 - \|y^{k-1} - y^k\|_S^2) + \frac{1+\mu}{2}\left(\|u^{k+1} - u^*\|_P^2 - \|u^k - u^*\|_P^2\right) \\ &\qquad + \frac{1-\mu}{2}\|x^k - x^{k+1}\|_R^2 + \Big(\frac{1-\mu}{2} - \eta\Big)\|y^k - y^{k+1}\|_S^2 \\ &\qquad + \frac{1}{2}\Big[\|v^* - v^{k+1}\|_{D_\gamma} - \|v^* - v^k\|_{D_\gamma}\Big] \\ &\qquad + \frac{1}{2}\left[\Big(1 - \frac{(1-\gamma)^2}{T-\gamma}\Big)\|B(y^k - y^{k+1})\|_H^2 + (2-T)\|Ax^{k+1} + By^{k+1} - b\|_H^2\right]. \end{aligned}$$

On the other hand, it follows from (2.1) that

$$(x^* - \tilde{x}^k)^T f(\tilde{x}^k) + (y^* - \tilde{y}^k)^T g(\tilde{y}^k) + (w^* - \tilde{w}^k)^T F(\tilde{w}^k) \le 0.$$

Therefore, the above two inequalities imply the following assertion:

$$\begin{aligned} &\|v^{k+1} - v^*\|_{D_\gamma}^2 + (1+\mu)\|u^{k+1} - u^*\|_P^2 \\ &\qquad + (T-\gamma)\|Ax^{k+1} + By^{k+1} - b\|_H^2 + 2\rho\|y^k - y^{k+1}\|_S^2 \\ &\quad \le \|v^k - v^*\|_{D_\gamma}^2 + (1+\mu)\|u^k - u^*\|_P^2 + (T-\gamma)\|Ax^k + By^k - b\|_H^2 \\ &\qquad + 2\rho\|y^{k-1} - y^k\|_S^2 - \Big(1 - \frac{(1-\gamma)^2}{T-\gamma}\Big)\|B(y^k - y^{k+1})\|_H^2 \\ &\qquad - (2-T)\|Ax^{k+1} + By^{k+1} - b\|_H^2 \\ &\qquad - (1-\mu)\|x^k - x^{k+1}\|_R^2 - (1-\mu-2\eta)\|y^k - y^{k+1}\|_S^2 \quad \forall w^* \in \mathcal{W}^*. \end{aligned} \tag{C.3}$$

Taking into account the definition of G_γ in (C.2), the assertion (C.3) can rewritten as the compact form (C.1). The proof is complete. ∎

Remark C.1. Note that the assertions in Lemma 2.4 ensure the positive definiteness of G_γ (thus, the notation $\|\cdot\|_{G_\gamma}$ makes sense) and the positiveness of the coefficients in the right-hand side of (C.1).

Now we are ready to show the global convergence of the scheme (1.10)–(1.12). The global convergence of the accompanying sequence $\{\tilde{w}^k\}$ is also derived.

THEOREM C.2. *Assume that the two functions $f(x)$ and $g(y)$ in the variational inequality* (1.1)–(1.2) *are continuous and monotone. The sequence $\{w^k\}$ generated by the scheme* (1.10)–(1.12) *converges globally to a solution point of VI$(\mathcal{W}, F)$.*

Proof. Given (C.1) and $1-\mu-2\eta>0$, $R\succ 0$, $S\succ 0$, $H\succ 0$, we have

$$\lim_{k\to\infty}\|x^k-x^{k+1}\|=0, \qquad \lim_{k\to\infty}\|y^k-y^{k+1}\|=0 \tag{C.4}$$

and

$$\lim_{k\to\infty}\|B(y^k-y^{k+1})\|=0, \qquad \lim_{k\to\infty}\|Ax^{k+1}+By^{k+1}-b\|=0. \tag{C.5}$$

Note that

$$\|\frac{1}{\gamma}H^{-1}(\lambda^k-\lambda^{k+1})\|=\|Ax^{k+1}+By^{k+1}-b\|\to 0, \;\; k\to\infty. \tag{C.6}$$

It thus follows from (3.4) that $\lim_{k\to\infty}\|w^k-w^{k+1}\|=0$.

Since

$$f_k(x^{k+1})=f(x^{k+1})-A^T\lambda^{k+1}+\tfrac{1-\gamma}{\gamma}A^T(\lambda^k-\lambda^{k+1})+A^THB(y^k-y^{k+1})$$

and

$$g_k(y^{k+1})=g(y^{k+1})-B^T\lambda^{k+1}+\tfrac{1-\gamma}{\gamma}B^T(\lambda^k-\lambda^{k+1}),$$

it follows from (2.7)–(2.8) and (C.4)–(C.6) that

$$\begin{cases}\lim\limits_{k\to\infty}\inf(x-x^{k+1})^T\left[f(x^{k+1})-A^T\lambda^{k+1}\right]\ge 0 & \forall x\in\mathcal{R}^n_+,\\ \lim\limits_{k\to\infty}\inf(y-y^{k+1})^T\left[g(y^{k+1})-B^T\lambda^{k+1}\right]\ge 0 & \forall y\in\mathcal{R}^m_+.\end{cases} \tag{C.7}$$

On the other hand, (C.1) shows that the sequence $\{w^k\}$ is bounded. Thus, there exists at least one cluster point. Let w^∞ be a cluster point of $\{w^k\}$, and let $\{w^{k_j}\}$ be the subsequence converging to w^∞. It follows from (C.6) and (C.7) that

$$\begin{cases}\lim\limits_{j\to\infty}\inf(x-x^{k_j})^T\left[f(x^{k_j})-A^T\lambda^{k_j}\right]\ge 0 & \forall x\in\mathcal{R}^n_+,\\ \lim\limits_{j\to\infty}\inf(y-y^{k_j})^T\left[g(y^{k_j})-B^T\lambda^{k_j}\right]\ge 0 & \forall y\in\mathcal{R}^m_+,\\ \lim\limits_{j\to\infty}(Ax^{k_j}+By^{k_j}-b)=0,\end{cases}$$

and consequently

$$\begin{cases}(x-x^\infty)^T\left[f(x^\infty)-A^T\lambda^\infty\right]\ge 0 & \forall x\in\mathcal{R}^n_+,\\ (y-y^\infty)^T\left[g(y^\infty)-B^T\lambda^\infty\right]\ge 0 & \forall y\in\mathcal{R}^m_+,\\ Ax^\infty+By^\infty-b=0.\end{cases}$$

That is, w^∞ is a solution point of VI$(\mathcal{W},F)$.

Note that inequality (C.1) is true for an arbitrary solution point of VI($\mathcal{W}, F$). Hence, there exists $\rho > \mu$ such that

$$\|w^{k+1} - w^\infty\|_{G_\gamma}^2 + 2\rho\|y^k - y^{k+1}\|_S^2 \le \|w^k - w^\infty\|_{G_\gamma}^2 + 2\rho\|y^{k-1} - y^k\|_S^2. \tag{C.8}$$

Since $w^{k_j} \to w^\infty$ and $\|y^{k_j-1} - y^{k_j}\| \to 0$ $(j \to \infty)$, for any given $\varepsilon > 0$, there exists an $l > 0$ such that

$$\|w^{k_l} - w^\infty\|_{G_\gamma}^2 + 2\rho\|y^{k_l-1} - y^{k_l}\|_S^2 \le \varepsilon. \tag{C.9}$$

Therefore, for any $k \ge k_l$, it follows from (C.8) and (C.9) that

$$\begin{aligned}\|w^k - w^\infty\|_{G_\gamma}^2 &\le \|w^k - w^\infty\|_{G_\gamma}^2 + 2\rho^0\|y^{k-1} - y^k\|_S^2\\ &\le \|w^{k_l} - w^\infty\|_{G_\gamma}^2 + 2\rho^0\|y^{k_l-1} - y^{k_l}\|_S^2\\ &\le \varepsilon.\end{aligned}$$

This implies that with an arbitrary initial iterate, the sequence $\{w^k\}$ converges to w^∞, which is a solution of VI($\mathcal{W}, F$).

With the proved convergence of $\{w^k\}$, the global convergence of the accompanying sequence $\{\tilde{w}^k\}$ is trivial based on the fact (3.5). The proof is complete. ∎

Acknowledgments. The authors are grateful to the associate editor and two anonymous referees for their valuable comments which have helped us greatly improve the presentation of this paper.

REFERENCES

[1] A. Auslender and M. Teboulle, *Lagrangian duality and related multiplier methods for variational inequality problems*, SIAM J. Optim., 10 (2000), pp. 1097–1115.

[2] A. Auslender and M. Teboulle, *Interior projection-like methods for monotone variational inequalities*, Math. Program., 104 (2005), pp. 39–68.

[3] A. Auslender, M. Teboulle, and S. Ben-Tiba, *A logarithmic-quadratic proximal method for variational inequalities*, Comput. Optim. Appl., 12 (1999), pp. 31–40.

[4] 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 (2011), pp. 1–122.

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

[6] 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, Stanford, CA, 1978.

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

[8] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal points algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–318.

[9] J. Eckstein and M. Fukushima, *Some reformulations and applications of the alternating direction method of multipliers*, in Large Scale Optimization: State of the Art, W. W. Hager, D. W. Hearn, and P. M. Pardalos, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994, pp. 115–134.

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

[11] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci. 3 (2010), pp. 1015–1046.

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

[13] M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary Valued Problem*, North-Holland, Amsterdam, 1983.

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

[15] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrange Methods: Applications to the Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, pp. 299–331.

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

[17] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, 1984.

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

[19] R. Glowinski and A. Marrocco, *Sur l'approximation par éléments finis d'ordre un, etla résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires*, RAIRO Anal. Numér., 9 (1975), pp. 41–76.

[20] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[21] B. S. He, X. M. Yuan and X. M. Xu, *A new inexact alternating directions method for monotone variational inequalities*, Math. Program., 92 (2002), pp. 103–118.

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

[23] B. S. He, X. M. Yuan and X. M. Xu, *On the $O(1/n)$ convergence rate of Douglas–Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[24] B. S. He, X. M. Yuan and X. M. Xu, *On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators*, submitted.

[25] B. S. He, X. M. Yuan and X. M. Xu, *A LQP based interior prediction-correction method for nonlinear complementarity problems*, J. Comput. Math. 24 (2006), pp. 33–44.

[26] B. S. He, X. M. Yuan and X. M. Xu, *A logarithmic-quadratic proximal prediction-correction method for structured monotone variational inequalities*, Comput. Optim. Appl., 35 (2006), pp. 19–46.

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

[28] M. K. Ng, F. Wang, and X. Yuan, *Inexact alternating direction methods for image recovery*, SIAM J. Sci. Comput., 33 (2011), pp. 1643–1668.

[29] M. K. Ng, P. Weiss, and X. Yuan, *Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods*, SIAM J. Sci. Comput., 32 (2010), pp. 2710–2736.

[30] S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, in Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision, Lecture Notes in Comput. Sci. 5567, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 464–476.

[31] 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.

[32] M. Tao and X. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57–81.

[33] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.

[34] Z. Wen, D. Goldfarb, and W. Yin, *Alternating direction augmented Lagrangian methods for semidefinite programming*, Math. Program. Comput., 2 (2010), pp. 203–230.

[35] X. M. Yuan, *Alternating direction methods of multipliers for covariance selection models*, J. Sci. Comput., 51 (2012), pp. 261–273.

[36] X. Yuan and M. Li, *An LQP-based decomposition method for solving a class of variational inequalities*, SIAM J. Optim., 21 (2011), pp. 1309–1318.

[37] X. M. Yuan, *The prediction-correction approach to nonlinear complementarity problems*, European J. Oper. Res., 176 (2007), pp. 1357–1370.

[38] S. Zhang, J. Ang, and J. Sun, *An alternating direction method for solving convex nonlinear semidefinite programming problem*, Optimization, to appear.

[39] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276.

[40] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.