THE GLOWINSKI–LE TALLEC SPLITTING METHOD REVISITED: A GENERAL CONVERGENCE AND CONVERGENCE RATE ANALYSIS

YAONAN MA AND LI-ZHI LIAO
Department of Mathematics, Hong Kong Baptist University
Kowloon Tong, Kowloon, Hong Kong SAR, China

(Communicated by Jie Sun)

ABSTRACT. In this paper, we focus on a splitting method called the $\theta$-scheme proposed by Glowinski and Le Tallec in [17, 20, 27]. First, we present an elaborative convergence analysis in a Hilbert space and propose a general convergent inexact $\theta$-scheme. Second, for unconstrained problems, we prove the convergence of the $\theta$-scheme and show a sublinear convergence rate in terms of objective value. Furthermore, a practical inexact $\theta$-scheme is derived to solve $l_2$-loss based problems and its convergence is proved. Third, for constrained problems, even though the convergence of the $\theta$-scheme is available in the literature, yet its sublinear convergence rate is unknown until we provide one via a variational reformulation of the solution set. Besides, in order to relax the condition imposed on the $\theta$-scheme, we propose a new variant and show its convergence. Finally, some preliminary numerical experiments demonstrate the efficiency of the $\theta$-scheme and our proposed methods.

1. Introduction. In this paper we consider a zero point problem

\[ 0 \in A(z) + B(z), \]

where $A$ and $B$ are two maximal monotone operators defined on a Hilbert space. Throughout this paper we assume that the solution set $G$ of the above problem is nonempty. To solve problem (1), we associate it with an initial value problem

\[ \frac{d}{dt} z + A(z) + B(z) \ni 0, \]

for some initial condition $z(0) = z_0$ with $z_0 \in H$ and we look for a steady state solution. Many operator splitting methods have been designed to capture the steady state solution of problem (2). In [23, 29], a thorough discussion and many applications of operator splitting methods are presented. In general, these methods can be classified into two categories.

The first one is the multiplicative operator splitting scheme, such as Lie’s scheme, Strang’s symmetrized operator splitting scheme and some variants. This kind of schemes are robust and easy to be implemented. However, the splitting errors of Lie’s scheme and Strang’s symmetrized scheme are $O(\Delta t)$ and $O(\Delta t^2)$, where

\[ 2010 \text{ Mathematics Subject Classification. Primary: 90C25; Secondary: 49M27.} \]

\[ \text{Key words and phrases. Splitting method, the } \theta\text{-scheme, convergence, convergence rate, inexactness criterion.} \]

The work of L.-Z. Liao was supported in part by grants from Hong Kong Baptist University (FRG) and General Research Fund (GRF) of Hong Kong.
\( \Delta t \) is the step size for the time discretization. Due to the splitting errors, these methods are asymptotically inconsistent which means that the iterative solutions only converge to some approximation of the steady state solution of problem (2), not the exact one.

The second one is the additive operator splitting scheme such as the Alternating Direction Implicit (ADI) type methods including the Douglas-Rachford Splitting Method (DRSM) and the Peaceman-Rachford Splitting Method (PRSM) which were first introduced for solving the elliptic and parabolic equations in [12, 31]. Given \( z^k \), DRSM and PRSM read as follows:

For DRSM:

\[
\begin{align*}
\frac{z^{k+1/2} - z^k}{\Delta t} + A(z^{k+1/2}) + B(z^k) &\geq 0, \\
\frac{z^{k+1} - z^k}{\Delta t} + A(z^{k+1}) + B(z^{k+1}) &\geq 0
\end{align*}
\]

And for PRSM:

\[
\begin{align*}
\frac{z^{k+1/2} - z^k}{\Delta t/2} + A(z^{k+1/2}) + B(z^k) &\geq 0, \\
\frac{z^{k+1} - z^{k+1/2}}{\Delta t/2} + A(z^{k+1/2}) + B(z^{k+1}) &\geq 0
\end{align*}
\]

Although both DRSM and PRSM are splitting error free, they are not well suited to simulate fast transient phenomena and capture the steady state solution of problem (2) efficiently if \( T \) is stiff, see [18, 23] for more details. To tackle the issues mentioned above, one may consider a variant of PRSM, called the \( \theta \)-scheme, which was firstly proposed by Glowinski and Le Tallec in [17, 20, 27]. To apply the \( \theta \)-scheme, we introduce a constant \( \theta \in (0, 1/2) \) and \( \Delta t \) as the step size for the time discretization, then the \( \theta \)-scheme applied to the initial value problem (2) reads as follows: Given \( z^k \), solve

For \( \theta \)-scheme:

\[
\begin{align*}
\frac{z^{k+\theta} - z^k}{\theta \Delta t} + A(z^{k+\theta}) + B(z^k) &\geq 0, \\
\frac{z^{k+1-\theta} - z^{k+\theta}}{(1 - 2\theta) \Delta t} + A(z^{k+\theta}) + B(z^{k+1-\theta}) &\geq 0, \\
\frac{z^{k+1} - z^{k+1-\theta}}{\theta \Delta t} + A(z^{k+1}) + B(z^{k+1-\theta}) &\geq 0
\end{align*}
\]

Among operator splitting methods, the \( \theta \)-scheme is the best suited to capture the steady state solution since it has better asymptotic properties as \( k \to +\infty \), see [17] for more details. Besides, the efficiency of the \( \theta \)-scheme has been verified in many areas such as equilibrium problems [36], linear eigenvalue problem, liquid crystal theory, viscoplasticity, elasto-viscoplasticity, and anisotropic Eikonal equation, see [20, 23] and the references therein. In particular, the \( \theta \)-scheme outperforms both DRSM and PRSM for the anisotropic Eikonal equation [21]. All these facts motivate us to study the \( \theta \)-scheme further.

The stability and convergence of the \( \theta \)-scheme has been discussed in [18, 19] for a special problem: \( A = \alpha_1 M \) and \( B = \alpha_2 M \) where \( M \) is a symmetric positive definite matrix and \( \alpha_1, \alpha_2 \) are two positive constants satisfying \( \alpha_1 + \alpha_2 = 1 \). Furthermore, if \( \theta \) is properly chosen, then the \( \theta \)-scheme is stiff A-stable not only for this particular case but for other more complex problems such as the unsteady incompressible viscous flow modeled by Navier-Stokes equations. As for a general quadratic optimization problem, the convergence of the \( \theta \)-scheme was proved in [20].
Moreover, [25] proved the convergence for problem (1) where the proof depends on Fejér inequality proposed in [35]. Recently, [36] improved the linear convergence rate obtained from [25]. It is noted that there is no more research focused on $\theta$-scheme’s general convergence and convergence rate. The purpose of this paper is to complement the study on the $\theta$-scheme in these aspects and show some useful variants.

The rest of this paper is organized as follows. First, some preliminaries are presented in Section 2. In Section 3, we show our main convergence result of the $\theta$-scheme in the Hilbert space. A general inexact $\theta$-scheme is proposed in Section 4. In Section 5, we analyze the convergence and convergence rate of the $\theta$-scheme for both unconstrained and constrained optimization problems. In addition, a practical inexact $\theta$-scheme is derived to solve a class of optimization problems and a new variant is proposed to relax the condition imposed on the $\theta$-scheme. Finally, two numerical experiments are provided to demonstrate the efficiency of all concerned algorithms in Section 6.

2. Preliminaries. We first present some results in operator theory.

2.1. Averaged nonexpansive operators.

**Definition 2.1.** Let $\alpha \in (0, 1)$. Then, an operator $T : \text{dom } T = \mathcal{H} \to \mathcal{H}$ is nonexpansive if
\[
\|Tx - Ty\| \leq \|x - y\|, \quad \forall (x, y) \in \mathcal{H}^2,
\]
and $T$ is $\alpha$-averaged if
\[
T = (1 - \alpha)I + \alpha S,
\]
for some nonexpansive operator $S : \text{dom } S = \mathcal{H} \to \mathcal{H}$. The class of $\alpha$-averaged operators is denoted by $\mathcal{A}(\alpha)$. In particular, $\mathcal{A}(1/2)$ is called the firmly nonexpansive operators.

In the case of $\alpha$-averaged operators in Definition 2.1, the following characterizations are due to [39].

**Lemma 2.2.** Let $T : \mathcal{H} \to \mathcal{H}$ and $\alpha \in (0, 1)$. Then, the following two characterizations are equivalent.

(i) $T \in \mathcal{A}(\alpha)$;
(ii) $\|Tx - Ty\|^2 \leq \|x - y\|^2 - ((1 - \alpha)/\alpha)\|(I - T)x - (I - T)y\|^2, \forall (x, y) \in \mathcal{H}^2$.

The following two lemmas show that under some conditions, operators $I - \beta B$ and $J_B = (I + B)^{-1}$ are $\alpha$-averaged.

**Lemma 2.3.** Let $B : \mathcal{H} \to \mathcal{H}$ and $\sigma \in (0, +\infty)$ such that $\sigma B \in \mathcal{A}(1/2)$, and let $\beta \in (0, 2\sigma)$. Then, $I - \beta B \in \mathcal{A}(\beta/2\sigma)$.

**Lemma 2.4.** Let $T : \mathcal{H} \to \mathcal{H}$. Then, $T \in \mathcal{A}(1/2)$ if and only if $T = J_B$ for some maximal monotone operator $B : \mathcal{H} \to 2^{\mathcal{H}}$.

2.2. A reformulation of the $\theta$-scheme. To analyze the convergence, we first reformulate the $\theta$-scheme (5) in the operator form as
\[
z^{k+1} = (I + \beta_1 A)^{-1}(I - \beta_1 B)(I + \beta_2 B)^{-1}(I - \beta_2 A)(I + \beta_1 A)^{-1}(I - \beta_1 B)z^k, \quad (6)
\]
where $\beta_1 = \theta \Delta t$ and $\beta_2 = (1 - 2\theta) \Delta t$. Then, by the definition of resolvent operator, (6) reduces to
\[
z^{k+1} = J_{\beta_1 A}(I - \beta_1 B)J_{\beta_2 B}(I - \beta_2 A)J_{\beta_1 A}(I - \beta_1 B)z^k.
\]
Moreover, noticing the identity
\[ I - \nu T = (\nu/\mu)(\gamma J_{\mu T} - I)(I + \mu T), \]  
where \( \gamma = 1 + \mu/\nu \) and \( \mu, \nu > 0 \), we can rewrite the iterate (7) in another form. Let 
\( T = A \) and \( \mu = \beta_1, \nu = \beta_2 \) in (8). Then, we obtain
\[ I - \beta_2 A = (\beta_2/\beta_1)(\gamma J_{\beta_1 A} - I)(I + \beta_1 A), \]
which implies that the \( \theta \)-scheme (7) reduces to
\[ z^{k+1} = J_{\beta_1 A}(I - \beta_1 B)J_{\beta_2 B}(\beta_2/\beta_1)(\gamma J_{\beta_1 A} - I)(I - \beta_1 B)z^k, \]  
where \( \gamma = 1 + \beta_1/\beta_2 \).

For the operator \( \gamma J_{\beta_1 A} - I \) in (9), an essential property was proved in [25].

Lemma 2.5. Let \( T \) be a maximal monotone operator defined on \( \mathcal{H} \), and let \( \mu \) and \( \nu \) be two positive constants. Set \( \gamma = 1 + \mu/\nu \). Then, the operator \( \gamma J_{\mu T} - I \) is Lipschitz continuous with constant \( L = \max\{1, \mu/\nu\} \).

3. Convergence analysis of the \( \theta \)-scheme. Now we are ready to prove the convergence of the \( \theta \)-scheme. In order to make our proof concise, we set
\[
\begin{align*}
T_1 &= J_{\beta_1 A}, \\
T_2 &= I - \beta_1 B, \\
T_3 &= J_{\beta_2 B}, \\
T_4 &= (\beta_2/\beta_1)(\gamma J_{\beta_1 A} - I),
\end{align*}
\]
then the iterate (9) can be rewritten as
\[ (\theta\text{-scheme}) \begin{cases} 
\nu^{k+1} = T_3T_4T_2z^k, \\
z^{k+1} = T_1T_2\nu^{k+1}.
\end{cases} \]  

Besides, it is proved that the solution set \( G \) is contained in the set of fixed points of both \( T_1T_2 \) and \( T_3T_4T_2 \), see [25, Theorem 2.1] for details.

Now we show our main convergence result.

Theorem 3.1. Suppose that \( B^{-1} \) is strongly monotone with modulus \( \sigma \in (0, +\infty) \) and \( 0 < \beta_2 \leq \beta_1 < 2\sigma \). Let \( \{(z^k, \nu^k)\}_{k \in \mathbb{N}} \) be generated by the iterate (11). Then, there exists an \( z^* \in G \) such that
\[
\begin{align*}
(i) & \sum_{k=0}^{\infty} \|z^{k+1} - \nu^{k+1}\| < +\infty; \\
(ii) & \text{the sequence} \{z^k\}_{k \in \mathbb{N}} \text{converges weakly to} z^*.
\end{align*}
\]

Proof. First, we know that \( T_1 \) and \( T_3 \) belong to \( A(1/2) \) by Lemma 2.4. Besides, as \( B^{-1} \) is strongly monotone, we have
\[ \langle x - y, Bx - By \rangle \geq \sigma \|Bx - By\|^2, \forall x, y \in \text{dom} \ B, \]
which implies \( \sigma B \in A(1/2) \). According to Lemma 2.3 and \( 0 < \beta_1 < 2\sigma \), we have \( T_2 \in A(\beta_1/2\sigma) \). From Lemma 2.5 and \( 0 < \beta_2 \leq \beta_1 \), we know that \( T_4 \) is nonexpansive.

Second, for any solution \( z^* \in G \), from Lemma 2.2 and the nonexpansivity of \( T_4 \), the following inequalities hold
\[ \|z^{k+1} - z^*\| \leq \|T_1T_2\nu^{k+1} - T_1T_2z^\infty\|^2 \]
\[ \leq \|T_2\nu^{k+1} - T_2z^\infty\|^2 - d^{(k)}_1 \]
Remark 1. Theorem 3.1 provides a novel convergence analysis of the \( \theta \)-scheme without assuming the finite dimension of \( \mathcal{H} \) which is necessary in [19, 20, 25, 36]. Besides, the proof fully utilizes the properties of all involved operators which is different from the proof in [25, Theorem 2.1] where the Fejér inequality proposed in [35] must be used. Furthermore, (i) is novel for the \( \theta \)-scheme as far as we know.

Remark 2. From Lemma 2.5 and Theorem 3.1, we know that if \( 0 < \beta_1 < \beta_2 \), then the operator \( \beta_2/\beta_1 (\gamma J_{\beta_1 A} - I) \) is Lipschitz continuous with \( \beta_2/\beta_1 \) (bigger than 1). Therefore the iterate (9) cannot be guaranteed to be nonexpansive if only \( B^{-1} \) is assumed to be strongly monotone. Some additional assumptions are needed to achieve nonexpansivity, see [25, Theorem 2.3] and [36, Theorem 3.3] for details.
4. A general inexact $\theta$-scheme. As for every operator in (9), it is possible that the output cannot be computed exactly. In the following, we consider a general inexact $\theta$-scheme

$$
\begin{align*}
\begin{cases}
\theta^{k+1} = z^k - \beta_{1,k}(Bz^k + e_{1,k}), \\
\theta^{k+1} = \beta_{2,k}/\beta_{1,k}(\gamma_k J_{\beta_{1,k}A} - I)v^{k+1} + e_{2,k}, \\
\theta^{k+1} = J_{\beta_{2,k}B}v^{k+1} + e_{3,k}, \\
\theta^{k+1} = v^{k+1} - \beta_{1,k}(Bv^{k+1} + e_{4,k}), \\
\theta^{k+1} = z^k + \alpha_k(J_{\beta_{1,k}A}z^{k+1} - z^k + e_{5,k}),
\end{cases}
\end{align*}
$$

(A general inexact $\theta$-scheme)

where $\gamma_k = 1+\beta_{1,k}/\beta_{2,k}$ and the last step can be regarded as a convex combination between the output of one step $\theta$-scheme and the previous point $z^k$. As far as we know, the above inexact scheme is novel for the $\theta$-scheme. A similar design for other splitting methods can refer to [10].

The assumption below is crucial to prove the convergence of the iterate (15).

Assumption 1.

\[
\begin{cases}
B^{-1} \text{ is strongly monotone with } \sigma \in (0, +\infty), \\
\sum_{k=0}^{\infty} \alpha_k \|e_i,k\| < +\infty, \forall i = 1, \ldots, 5, \\
0 < \inf_{k \geq 0} \alpha_k \leq \alpha_k \leq 1, \\
0 < \inf_{k \geq 0} \beta_{2,k} \leq \beta_{2,k} \leq \beta_{1,k} \leq \sup_{k \geq 0} \beta_{1,k} < 2\sigma.
\end{cases}
\]

If we let

\[
\begin{align*}
T_{1,k} &= J_{\beta_{1,k}A}, \quad T_{2,k} = I - \beta_{1,k}B, \\
T_{3,k} &= J_{\beta_{2,k}B}, \quad T_{4,k} = (\beta_{2,k}/\beta_{1,k})(\gamma_k J_{\beta_{1,k}A} - I),
\end{align*}
\]

and set

\[
\begin{align*}
v^k &= T_{3,k}T_{4,k}T_{2,k}z^k, \\
y^k &= T_{1,k}T_{2,k}v^k,
\end{align*}
\]

then the convergence of the iterate (15) can be established as follows.

Theorem 4.1. Suppose that Assumption 1 is satisfied. Let $\{z^k\}_{k \in \mathbb{N}}$ be generated by the iterate (15) and $\{(y^k, v^k)\}_{k \in \mathbb{N}}$ be generated by the iterate (16). Then, there exists an $z^* \in G$ such that

(i) $\sum_{k=0}^{\infty} \alpha_k(1 - \alpha_k)\|y^k - z^k\|^2 < +\infty$;

(ii) the sequence $\{z^k\}_{k \in \mathbb{N}}$ converges weakly to $z^*$.

Proof. First, setting $z^{k+1} = x^k + e_k$ where $x^k = z^k + \alpha_k(y^k - z^k)$ and considering the iterate (15), we obtain

$$e_k = \alpha_k(T_{1,k}(T_{2,k}(T_{3,k}(T_{4,k}z^k - \beta_{1,k}e_{1,k}) + e_{2,k}) + e_{3,k}) - \beta_{1,k}e_{4,k}) + e_{5,k} - y^k).$$

Recall the nonexpansivity of $(T_{i,k})_{1 \leq i \leq 4}$, we can get

$$\|e_k\| \leq \alpha_k(\beta_{1,k}\|e_{1,k}\| + \|e_{2,k}\| + \|e_{3,k}\| + \|e_{4,k}\| + \|e_{5,k}\|).$$

Combining this with Assumption 1, we have

$$\sum_{k=0}^{\infty} \|e_k\| < +\infty.$$ (17)
Second, for any solution \( z^\infty \) in \( G \), by the nonexpansivity of the operators \((T_{i,k})_{1 \leq i \leq 4} \), we know
\[
\| z^{k+1} - z^\infty \| \leq \| x^k - z^\infty \| + \| e_k \|
\leq (1 - \alpha_k)\| z^k - z^\infty \| + \alpha_k\| y^k - z^\infty \| + \| e_k \|
\leq \| z^k - z^\infty \| + \| e_k \|,
\]
which implies that the sequence \( \{ z^k \}_{k \in \mathbb{N}} \) is bounded. From (18), we can obtain
\[
\| z^{k+1} - z^\infty \|^2
\leq \|(1 - \alpha_k)(z^k - z^\infty) + \alpha_k(y^k - z^\infty)\|^2 + (2\| x^k - z^\infty \| + \| e_k \|)\| e_k \|
\leq (1 - \alpha_k)\| z^k - z^\infty \|^2 + \alpha_k\| y^k - z^\infty \|^2 - \alpha_k(1 - \alpha_k)\| y^k - z^k \|^2 + \| e_k \|
\leq \| z^k - z^\infty \|^2 - \alpha_k(e_{k1}^{(k)} + e_{k2}^{(k)} + d_{k1}^{(k)} + d_{k2}^{(k)}) - \alpha_k(1 - \alpha_k)\| y^k - z^k \|^2 + \| e_k \|,
\]
where \( \xi = \sup_{k \geq 0} \{ 2\| x^k - z^\infty \| + \| e_k \| \} < +\infty \) and
\[
\begin{align*}
d_{k1}^{(k)} &= \|(I - T_{1,k})T_{2,k}v^k - (I - T_{1,k})T_{2,k}z^\infty \|^2, \\
e_{k1}^{(k)} &= \frac{1 - \beta_{1,k}/2\sigma}{\beta_{1,k}/2\sigma} \| \beta_{1,k}Bv^k - \beta_{1,k}Bz^\infty \|^2, \\
d_{k2}^{(k)} &= \|(I - T_{3,k})T_{4,k}T_{2,k}z^k - (I - T_{3,k})T_{4,k}T_{2,k}z^\infty \|^2, \\
e_{k2}^{(k)} &= \frac{1 - \beta_{1,k}/2\sigma}{\beta_{1,k}/2\sigma} \| \beta_{1,k}Bz^k - \beta_{1,k}Bz^\infty \|^2.
\end{align*}
\]
Then, summing up (19) for all \( k \)'s, (i) is obtained immediately.

Finally, by Theorem 3.1, we have \( \| v^k - y^k \| \to 0 \). Besides, from (i), we know \( \alpha_k(1 - \alpha_k)\| y^k - z^k \| \to 0 \). If \( \sup_{k \geq 0} \alpha_k < 1 \), then by Assumption 1 and (i), we get \( \| z^k - y^k \| \to 0 \). If \( \sup_{k \geq 0} \alpha_k = 1 \), then by \( z^{k+1} = \alpha_k y^k + (1 - \alpha_k)z^k + e_k \) and (17), we get \( \| z^{k+1} - y^k \| \to 0 \). From the boundedness of the sequence \( \{ z^k \}_{k \in \mathbb{N}} \), we know that there exists an \( z^* \in \mathcal{H} \) such that \( z^k \rightharpoonup z^* \). Setting \( u^k = (v^k - y^k)/\beta_{1,k} - Bv^k \), we have \( u^k \in Ap^k \) and \( u^k \to -Bz^\infty \). If \( \| z^k - y^k \| \to 0 \), then we get \( y^k \rightharpoonup z^* \) and \( v^k \rightharpoonup z^* \). By \( Bv^k \to Bz^\infty \) and the weak-strong closeness of gra \( B \), we can have \( Bz^\infty = Bz^* \) which implies \( u^k \to -Bz^* \). Furthermore, due to the weak-strong closeness of gra \( A \) and \( u^k \in Ap^k \), we can obtain \( -Bz^* \in Az^* \); if \( \| z^{k+1} - y^k \| \to 0 \), then we get \( y^k \rightharpoonup z^* \) and \( v^k \rightharpoonup z^* \). Similarly, we can have \( Bz^\infty = Bz^* \). In addition, by \( u^k \rightharpoonup \in Ap^k \), we can obtain \( -Bz^* \in Az^* \). In summary, both cases indicate that \( z^* \) is a solution in \( G \). From (19) and [10, Lemma 2.8], the sequence \( \{ z^k \}_{k \in \mathbb{N}} \) converges weakly to \( z^* \).

**Remark 3.** The summable condition in Assumption 1 is not practical as it does not specify what precision is at every iterate. In the next section, we will propose a practical inexact \( \theta \)-scheme for \( l_2 \)-loss based convex optimization problems.

5. **Applications of the \( \theta \)-scheme.** In this section, we will focus on the convergence and convergence rate of the \( \theta \)-scheme for convex optimization problems. Obviously, some previous results can be easily extended to convex optimization problems while other results, such as convergence rate in terms of objective value, need careful scrutiny.

Let us first focus on the \( \theta \)-scheme for unconstrained convex optimization problems.
5.1. **Unconstrained convex optimization problems.** The following convex optimization problem is considered

\[
\min_{z \in \mathcal{H}} f(z) + g(z),
\]

where \( f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\} \) and \( g : \mathcal{H} \to \mathbb{R} \cup \{+\infty\} \) are two closed proper convex functions such that \( g \) is smooth with a \( 1/\sigma \)-Lipschitz continuous gradient for some \( \sigma \in (0, +\infty) \). It follows from the definition of \( G \) that

\[
G = \{ z \in \mathcal{H} \mid 0 \in \partial f(z) + \nabla g(z) \}.
\]

According to Theorem 3.1, we can immediately obtain the following convergence result.

**Theorem 5.1.** Suppose \( 0 < \beta_2 \leq \beta_1 < 2\sigma \). Let \( \{(z^k, v^k)\}_{k \in \mathbb{N}} \) be generated by

\[
\begin{align*}
v^{k+1} &= \text{prox}_{\beta_2 g}(\beta_2 / \beta_1)(\gamma \text{prox}_{\beta_1 f} I)(z^k - \beta_1 \nabla g(z^k)), \\
z^{k+1} &= \text{prox}_{\beta_1 f}(v^{k+1} - \beta_1 \nabla g(v^{k+1})),
\end{align*}
\]

where \( \gamma = 1 + \beta_1 / \beta_2 \) and \( \text{prox}_f(v) := \arg \min_z f(z) + \frac{1}{2}\|z - v\|^2 \). Then, there exists an \( z^* \in G \) such that

(i) \( \sum_{k=0}^{\infty} \|z^{k+1} - v^{k+1}\|^2 < +\infty \);

(ii) the sequence \( \{z^k\}_{k \in \mathbb{N}} \) converges weakly to \( z^* \).

**Proof.** Since \( f \) and \( g \) are two closed proper convex functions, [3, Theorem 3.1.11] asserts that both \( \partial f \) and \( \nabla g \) are maximal monotone. Besides, as \( g \) is smooth with \( 1/\sigma \), it follows from [38, Corollaire 10] that \( (\nabla g)^{-1} \) is strongly monotone with \( \sigma \). Then, the iterate (20) is a special case of the \( \theta \)-scheme (11) where \( \theta = \partial f \) and \( B = \nabla g \). Therefore, (i) and (ii) can be obtained by Theorem 3.1. \( \square \)

**Remark 4.** One can also consider an inexact form of the iterate (20) similar to the iterate (15) and obtain related convergence results. However, as the form is not practical, we omit it here for the sake of succinctness.

**Remark 5.** If some condition is imposed on the solution set \( G \) or function \( f \) or \( g \), for example, \( G \neq \emptyset \), then one can get strong convergence of the sequence \( \{z^k\}_{k \in \mathbb{N}} \) generated by the iterate (20), see [11] for more details.

Theorem 5.1 only provides the convergence of the iterate (20). Next we show a sublinear convergence rate which is firstly proposed as far as we know.

**Theorem 5.2.** Suppose \( 0 < \beta_2 \leq \beta_1 < \sigma \). Let \( \{(z^k, v^k)\}_{k \in \mathbb{N}} \) be generated by the iterate (20). Then, we have

\[
F(\tilde{z}^n) - F(z^*) \leq \frac{\|z^0 - z^*\|^2}{2n\beta_1}, \quad \forall z^* \in G, \quad (21)
\]

where \( F(z) = f(z) + g(z) \), \( \tilde{z}^n = (\sum_{k=0}^{n-1} z^k) / n \).

**Proof.** For any solution \( z^* \) in \( G \), from [6, Lemma 2.3] and the iterate (20), we have

\[
2\beta_1(F(z^*) - F(z^{k+1})) \geq \|z^{k+1} - v^{k+1}\|^2 + 2\langle v^{k+1} - z^*, z^{k+1} - v^{k+1} \rangle \\
= \|z^{k+1} - z^*\|^2 - \|v^{k+1} - z^*\|^2 \\
\geq \|z^{k+1} - z^*\|^2 - \|z^k - z^*\|^2, \quad (22)
\]
where the last inequality is derived from the proof of Theorem 5.1. Summing up (22) from \( k = 0 \) to \( k = n - 1 \), we have

\[
2\beta_1(nF(z^*) - \sum_{k=0}^{n-1} F(z^{k+1})) \geq \|z^n - z^*\|^2 - \|z^0 - z^*\|^2,
\]

and by the convexity of function \( F \), we can obtain inequality (21) immediately. \( \square \)

5.1.1. A practical inexact \( \theta \)-scheme. Now we are going to propose a practical inexact \( \theta \)-scheme for the following convex optimization problem

\[
\min_{z \in \mathbb{R}^n} f(z) + \frac{1}{2}\|Qz - q\|^2,
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) is a closed proper convex function and \( Q \in \mathbb{R}^{m \times n} \) is a matrix. In general, function \( f(z) \) can be regarded as a penalty, for example, if \( f(z) = \eta \|z\|_1 \) where \( \eta > 0 \) is a constant, then problem (23) reduces to the Least Absolute Shrinkage and Selection Operator (LASSO) [34]. Applying the \( \theta \)-scheme (20) to solve problem (23), we obtain

\[
v^k = (\beta_2/\beta_1)(\gamma \text{prox}_{\beta_1 f - I})(z^k - \beta_1 Q^T(Qz^k - q)),
\]

\[
w^{k+1} = \arg \min_{z \in \mathbb{R}^n} \frac{\beta_2}{2}\|Qz - q\|^2 + \frac{1}{2}\|z - v^k\|^2,
\]

\[
z^{k+1} = \text{prox}_{\beta_1 f}(w^{k+1} - \beta_1 Q^T(Qw^{k+1} - q)),
\]

where \( \gamma = 1 + \beta_1/\beta_2 \). For the updates (24) and (26), let us assume that they can be solved exactly and fast. However, the update (25) can be expressed as

\[
w^{k+1} = (\beta_2 Q^TQ + I)^{-1}(\beta_2 Q^Tw^k + v^k),
\]

and this linear equation usually cannot be solved exactly because of large dimension of \( Q^TQ \). Thus, we must consider solving (27) inexact which raises a question: under what scale should we solve (27) at every iterate? On the one hand, if (27) is solved not very accurately, then it could severely affect the sequent iterate such that the generated sequence could even be divergent. On the other hand, if (27) is solved too accurately, then it can make the total computing time very long and it is also unnecessary to solve (27) too accurately at first a few iterates.

In the following, we consider an inexact \( \theta \)-scheme similar to the iterate (15) but with a practical criterion to control the residual of linear equation (27) dynamically. The inexact \( \theta \)-scheme we propose is

\[
\begin{align*}
v^k &= (\beta_2/\beta_1)(\gamma \text{prox}_{\beta_1 f - I})(z^k - \beta_1 Q^T(Qz^k - q)), \\
w^{k+1} &= \mathcal{L}v^k, \\
z^{k+1} &= \text{prox}_{\beta_1 f}(w^{k+1} - \beta_1 Q^T(Qw^{k+1} - q)),
\end{align*}
\]

where \( \mathcal{L} \) is a linear mapping from \( \mathbb{R}^n \) to \( \mathbb{R}^n \) and we will specify it later.

For linear equation (27), as referred in [9, Section 4.2.4], if \( m \ll n \), then we can compute \( w^{k+1} \) via the following procedures

\[
\begin{align*}
(\beta_2 Q^TQ + I)\eta^k &= Q\beta_2(Q^Tq + v^k/\beta_2) := Q\beta_2 h^k, \\
w^{k+1} &= \beta_2(h^k - Q^T\eta^k),
\end{align*}
\]

where \( h^k = Q^Tq + v^k/\beta_2 \). It is noted that the dimension of the above linear equation is \( m \times m \), much smaller than the dimension of linear equation (27). However, if \( m \) is still large, then we should also consider solving linear equation in (29) inexact.
Similar to the inexactness criterion proposed in [37], we derive the following process to formulate the linear mapping \( \mathcal{L} \):

\[
\begin{cases}
\text{Choose } \eta^k \text{ such that } \|e_k(\eta^k)\| \leq \alpha\|e_{k-1}(\eta^{k-1})\|, \\
w^{k+1} = \beta_2(h^k - Q^T\eta^k),
\end{cases}
\]

(30)

where \(e_k(\eta) = Q\beta_2h^k - (\beta_2QQ^T + I)\eta\) and \(\alpha \in (0, 1)\). As matrix \(\beta_2QQ^T + I\) is symmetric positive definite, we can apply numerical linear algebra solvers, for example, Successive Over-Relaxation (SOR), Conjugate Gradient (CG) methods to achieve the inexactness criterion in (30).

In the following, we will prove the convergence of the iterate (28). In order to make our proof concise, we let

\[
\begin{align*}
T_1 &= \text{prox}_{\beta_1f}, \\
T_2 &= I - \beta_1Q^T(Q \cdot -q), \\
T_3 &= \text{prox}_{\beta_2\|Q^{-q}\|_2^2}, \\
T_4 &= (\beta_2/\beta_1)(\gamma\text{prox}_{\beta_1f} - I),
\end{align*}
\]

and set

\[
y^k = T_1T_2T_3T_4z^k.
\]

(31)

**Theorem 5.3.** Suppose \(0 < \beta_2 \leq \beta_1 < 2/\|Q^TQ\|\). Let \(\{z^k, v^k, w^k\}\) be generated by the iterate (28) where \(L\) is defined in (30) and \(\{y^k\}\) be generated by the iterate (31). Then, there exists an \(z^* \in G\) such that

(i) \(\sum\limits_{k=0}^{\infty} \| z^{k+1} - w^{k+1} \|^2 < +\infty\);

(ii) the sequence \(\{z^k\}\) converges to \(z^*\).

**Proof.** First, we show that the sequence \(\{z^k\}\) is bounded. It is noted that the only difference between the iterate (31) and the inexact \(\theta\)-scheme (28) is that \(T_3\) is replaced by the linear mapping \(\mathcal{L}\). Setting \(\tilde{w}^{k+1} = T_3v^k\), we have

\[
\|w^{k+1} - \tilde{w}^{k+1}\| = \|L(v^k - T_3v^k)\| = \|\beta_2(h^k - Q^T\eta^k)\| - (\beta_2QQ^T + I)^{-1}\beta_2h^k\|
\]

\[
\leq \|\beta_2QQ^T + I\|^{-1} \|\beta_2QQ^T + I\|\|\beta_2(h^k - Q^T\eta^k)\| - \beta_2h^k\|
\]

\[
\leq \|\beta_2QQ^T + I\|^{-1} \|\beta_2QQ^T\|\|\beta_2h^k - (\beta_2QQ^T + I)\eta^k\|
\]

\[
\leq \alpha^k\|\beta_2QQ^T + I\|^{-1}\|\beta_2QQ^T\|\|e_0(\eta^0)\|
\]

\[
= \alpha^kM\|e_0(\eta^0)\|,
\]

(32)

where \(M = ||(\beta_2QQ^T + I)^{-1}||\|\beta_2QQ^T\|\) and the last inequality is obtained by the inexactness criterion in (30).

For any solution \(z^\infty\) in \(G\), from (32) and \(0 < \beta_2 \leq \beta_1 < 2/\|Q^TQ\|\), we obtain

\[
\|z^{k+1} - z^\infty\| = \|T_1T_2w^{k+1} - T_1T_2z^\infty\| \leq \|w^{k+1} - z^\infty\|
\]

\[
\leq \|w^{k+1} - T_3v^k\| + \|T_3v^k - z^\infty\|
\]

\[
= \|L(v^k - T_3v^k)\| + \|T_3T_4z^k - T_3T_4z^\infty\|
\]

\[
\leq M\|e_0(\eta^0)\|\alpha^k + \|z^k - z^\infty\|
\]

\[
\leq \sum\limits_{n=0}^{k} M\|e_0(\eta^0)\|\alpha^n + \|z^0 - z^\infty\|
\]

\[
= M\|e_0(\eta^0)\|(1 - \alpha^{k+1})/(1 - \alpha) + \|z^0 - z^\infty\|.
\]

(33)

By (33) and \(\alpha \in (0, 1)\), we conclude that the sequence \(\{z^k\}\) is bounded.
Second, setting \( s_k := T_1 T_2 CT_2 T_2 z^k - y^k \), we formulate the inexact \( \theta \)-scheme (28) as \( z^{k+1} = y^k + s_k \). Note \( \| s_k \| \leq \| \mathcal{L} v^k - T_3 v^k \| \leq M \| c_0(q) \| \alpha^k \) which means \( \sum_{k=0}^{\infty} \| s_k \| < +\infty \). Then, we have
\[
\| z^{k+1} - z^\infty \|^2 = \| y^k - z^\infty + s_k \|^2
\leq \| y^k - z^\infty \|^2 + (2\| y^k - z^\infty \| + \| s_k \|) \| s_k \|
\leq \| z^k - z^\infty \|^2 - c_2^{(k)} - d_2^{(k)} - c_1^{(k)} - d_1^{(k)} + \zeta \| s_k \|,
\]
where the last inequality is obtained by Lemma 2.2 and
\[
\begin{align*}
\zeta &= \sup_{k \geq 0} \{ 2\| y^k - z^\infty \| + \| s_k \| \} < +\infty, \\
d_1^{(k)} &= \| (I - T_1) T_2 w^{k+1} - (I - T_1) T_2 z^\infty \|^2, \\
c_1^{(k)} &= \frac{1 - \beta_1 / 2\sigma}{\beta_1 / 2\sigma} \| \beta_1 Q^T Q(\hat{w}^{k+1} - z^\infty) \|^2, \\
d_2^{(k)} &= \| (I - T_3) T_4 T_2 z^k - (I - T_3) T_4 T_2 z^\infty \|^2, \\
c_2^{(k)} &= \frac{1 - \beta_1 / 2\sigma}{\beta_1 / 2\sigma} \| \beta_1 B z^k - \beta_1 B z^\infty \|^2,
\end{align*}
\]
where \( \sigma = 1/\| Q^T Q \| \). Summing up (34) for all \( k' \)'s, we obtain
\[
\sum_{k=0}^{+\infty} d_1^{(k)} < +\infty, \quad \sum_{k=0}^{+\infty} c_1^{(k)} < +\infty. \tag{35}
\]

By (32), we know
\[
\sum_{k=0}^{+\infty} \| w^{k+1} - \hat{w}^{k+1} \|^2 < +\infty.
\]

This and (35) yield
\[
\begin{align*}
\sum_{k=0}^{+\infty} &\| (I - T_1) T_2 w^{k+1} - (I - T_1) T_2 z^\infty \|^2 < +\infty, \\
\sum_{k=0}^{+\infty} &\| Q^T Q(\hat{w}^{k+1} - z^\infty) \|^2 < +\infty. \tag{36}
\end{align*}
\]

Then, we can bound \( \| z^{k+1} - w^{k+1} \|^2 \) as follows
\[
\| z^{k+1} - w^{k+1} \|^2
= \| T_1 T_2 w^{k+1} - w^{k+1} \|^2
\leq 2 \| (I - T_1) T_2 w^{k+1} - (I - T_1) T_2 z^\infty \|^2 + 2 \| Q^T Q(\hat{w}^{k+1} - z^\infty) \|^2.
\]

Combining this with (36), (i) is proved.

Finally, setting \( u^k = (w^k - z^k) / \beta_1 - Q^T (Q u^k - q) \), we obtain \( u^k \in \partial f(z^k) \). By (i) and (36), we have \( u^k \to -Q^T (Q z^\infty - q) \). Since the sequence \( \{ z^k \}_{k \in \mathbb{N}} \) is bounded, there exists an \( z^* \in \mathbb{R}^n \) such that \( z^j \to z^* \). By (i) and (36), we have \( u^k \to z^* \) and \( Q^T (Q u^k - q) \to Q^T (Q z^* - q) \), respectively. Since gra \( Q^T (Q \cdot - q) \) is sequentially strongly-strongly closed in \( \mathbb{R}^n \times \mathbb{R}^n \), we get \( Q^T (Q z^* - q) = Q^T (Q z^* - q) \) and \( u^k \to -Q^T (Q z^* - q) \). Similarly, by the strong-strong closeness of gra \( \partial f \) in
5.2. Constrained convex optimization problems. Now let us apply the \(\theta\)-scheme to solve the following constrained convex optimization problem

\[
\min_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} f(x) + g(y),
\]
\[
s.t. \quad C x + D y = b,
\]

where \(f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}\) and \(g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}\) are two closed proper convex functions, \(C\) is a \(d \times n\) matrix, \(D\) is a \(d \times m\) matrix, and \(b\) is a vector of \(\mathbb{R}^d\).

If the corresponding augmented Lagrange function is defined as

\[
L_\beta(x, y, \lambda) = f(x) + g(y) - \lambda^T(C x + D y - b) + \frac{\beta}{2} \|C x + D y - b\|^2,
\]

where \(\lambda \in \mathbb{R}^d\) and \(\beta \in (0, +\infty)\), then the \(\theta\)-scheme (7) applied to (37) becomes

\[
(\theta\text{-scheme}) \begin{cases} 
\text{Step 1.} \quad y^k = \arg \min_y L_0(x^k, y, \lambda^k), \\
\quad x^{k+\theta} = \arg \min_x L_{\beta_1}(x, y^k, \lambda^k), \\
\quad \lambda^{k+\theta} = \lambda^k - \beta_1(C x^{k+\theta} + D y^k - b), \\
\text{Step 2.} \quad y^{k+1-\theta} = \arg \min_y L_{\beta_2}(x^{k+\theta}, y, \lambda^{k+\theta}), \\
\quad \lambda^{k+1-\theta} = \lambda^{k+\theta} - \beta_2(C x^{k+\theta} + D y^{k+1-\theta} - b), \\
\text{Step 3.} \quad x^{k+1} = \arg \min_x L_{\beta_1}(x, y^{k+1-\theta}, \lambda^{k+1-\theta}), \\
\quad \lambda^{k+1} = \lambda^{k+1-\theta} - \beta_1(C x^{k+1} + D y^{k+1-\theta} - b), 
\end{cases}
\]

where \(\beta_1, \beta_2 \in (0, +\infty)\) and \(L_0(\cdot, \cdot, \cdot)\) is a Lagrange function while \(L_\beta(\cdot, \cdot, \cdot)\) is an augmented Lagrange function. To prove the convergence, [25] proposed the following assumption.

**Assumption 2.** (i) There exist \(x \in \text{ri(dom} f)\) and \(y \in \text{ri(dom} g)\) such that \(C x + D y = b\);
(ii) the function \(f(x) + \|C x\|^2\) attains its minimum;
(iii) the function \(g\) is strongly convex with modulus \(\sigma > 0\), that is,

\[
g(y) - g(y') \geq s^T(y - y') + \frac{\sigma}{2} \|y - y'\|^2, \quad \forall y, y' \in \mathbb{R}^m,
\]

where \(s \in \partial g(y')\).

The following results are proved in [25, Proposition 3.4].

**Lemma 5.4.** Suppose that Assumption 2 is satisfied. Then, all the sequences generated by the iterate (38) are well-defined. Assume in addition that

\[
0 < \beta_2 \leq \beta_1 < 2\sigma/\|D\|^2.
\]

Then, we have

(i) \(\lambda^k\) converges to \(\lambda^*\);
(ii) \(y^k\) converges to \(y^*\);
(iii) \(C x^{k+\theta}\) converges to \(b - Dy^*\),

where \((x^*, y^*)\) are the exact solution to problem (37) and \(\lambda^*\) is the corresponding Lagrange multiplier.
It is noted that [25, Proposition 3.4] also proved the linear convergence rate for the iterate (38) when \( g \) is strongly convex and \( C \partial f^* C^T \) is strongly monotone (or \( D D g^* D^T \) is strongly monotone). However, there is no any convergence rate result when only \( g \) is assumed to be strongly convex. In the following, we will show a sublinear convergence rate, which is weaker than the linear convergence rate, under two scenarios: only \( g \) is strongly convex; both \( g \) and \( f \) are strongly convex (a stronger result will be obtained).

5.2.1. One strongly convex objective function. First, we introduce an equivalent variational inequalities (VI) reformulation to characterize the solution set of problem (37), which reads as

Find a \( w^* = (x^*, y^*, \lambda^*) \in \Omega := \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \) such that

\[
\text{VI}(\Omega, F, r) : \quad r(u) - r(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega, \quad (40)
\]

where

\[
u = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad w = \begin{pmatrix} x' \\ y' \\ \lambda' \end{pmatrix}, \quad F(w) = \begin{pmatrix} -C^T \lambda \\ -D^T \lambda \\ Cx + Dy - b \end{pmatrix} \quad \text{and} \quad r(u) = f(x) + g(y).
\]

Now we show an \( O(1/n) \) convergence rate for the iterate (38).

**Theorem 5.5.** Suppose that Assumption 2 is satisfied and

\[
0 < \beta_2 \leq \beta_1 \leq \frac{\sigma}{\|D\|^2}. \quad (41)
\]

Let \( \{(x^k, y^k, \lambda^k)\}_{k \in \mathbb{N}} \) be generated by the iterate (38). Then, we have

\[
r(\tilde{u}^n) - r(u) + (\tilde{w}^n - w)^T F(w) \leq \frac{\|\lambda - \lambda^0\|^2}{2n \Delta t}, \quad \forall w \in \Omega, \quad (42)
\]

where \( \Delta t = 2\beta_1 + \beta_2, \quad \tilde{w}^n = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{w}^{k+1}, \)

\[
\tilde{w}^{k+1} = \frac{1}{\Delta t} \left( \beta_1 w^{k+\theta,k} + \beta_2 w^{k+\theta,k+1-\theta} + \beta_1 w^{k+1,k+1-\theta} \right),
\]

and

\[
w^{k+\theta,k} = \begin{pmatrix} x^{k+\theta,k} \\ y^{k+\theta,k} \\ \lambda^{k+\theta,k} \end{pmatrix}, \quad w^{k+\theta,k+1-\theta} = \begin{pmatrix} x^{k+\theta,k+1-\theta} \\ y^{k+\theta,k+1-\theta} \\ \lambda^{k+\theta,k+1-\theta} \end{pmatrix}, \quad w^{k+1,k+1-\theta} = \begin{pmatrix} x^{k+1,k+1-\theta} \\ y^{k+1,k+1-\theta} \\ \lambda^{k+1,k+1-\theta} \end{pmatrix}. \quad (43)
\]

**Proof.** First, the optimality conditions of Step 1 of the iterate (38) are

\[
\begin{aligned}
g(y) - g(y^k) + (y - y^k)^T (-D^T \lambda^{k+\theta} + D^T (\lambda^{k+\theta} - \lambda^k)) \geq \frac{\sigma}{2} \|y - y^k\|^2, \\
f(x) - f(x^{k+\theta}) + (x - x^{k+\theta})^T (C^T \lambda^{k+\theta}) \geq 0,
\end{aligned} \quad (44)
\]

and the dual update \( \lambda^{k+\theta} = \lambda^k - \beta_1 (C x^{k+\theta} + D y^k - b) \) is reformulated as

\[
(\lambda - \lambda^{k+\theta})^T (C x^{k+\theta} + D y^k - b + \frac{1}{\beta_1} (\lambda^{k+\theta} - \lambda^k)) \geq 0, \quad \forall \lambda \in \mathbb{R}^d. \quad (45)
\]

Combining (44) and (45), we have a compact reformulation

\[
r(u) - r(u^{k+\theta,k}) + (w - w^{k+\theta,k})^T F(w^{k+\theta,k})
\]

\[
\geq \frac{\sigma}{2} \|y - y^k\|^2 + (D y - D y^k)^T (\lambda^k - \lambda^{k+\theta}) + \frac{1}{\beta_1} (\lambda - \lambda^{k+\theta})^T (\lambda^k - \lambda^{k+\theta}). \quad (46)
\]
Furthermore, the right-hand side of (46) reduces to
\[
\frac{\sigma}{2} \| y - y^k \|^2 + \beta_1 (Dy - Dy^k)^T (Cx^{k+\theta} + Dy^k - b) + \frac{1}{\beta_1} (\lambda - \lambda^{k+\theta})^T (\lambda^k - \lambda^{k+\theta}) \\
= \frac{\sigma}{2} \| y - y^k \|^2 - \frac{\beta_1}{2} \| Dy - Dy^k \|^2 + \frac{\beta_1}{2} \| Cx^{k+\theta} + Dy - b \|^2 \\
+ \frac{1}{2\beta_1} \| \lambda - \lambda^{k+\theta} \|^2 - \frac{1}{2\beta_1} \| \lambda - \lambda^k \|^2.
\]
(47)

Substituting (47) into (46) and using the monotonicity of \( \mathcal{F} \), we get
\[
\beta_1 (r(u) - r(u^{k+\theta,k}) + (w - w^{k+\theta,k})^T \mathcal{F}(w)) \\
\geq \frac{\sigma\beta_1}{2} \| y - y^k \|^2 - \frac{\beta_1^2}{2} \| Dy - Dy^k \|^2 + \frac{\beta_1^2}{2} \| Cx^{k+\theta} + Dy - b \|^2 \\
+ \frac{1}{2} \| \lambda - \lambda^{k+\theta} \|^2 - \frac{1}{2} \| \lambda - \lambda^k \|^2.
\]
(48)

Second, we consider the following iterate steps
\[
\begin{align*}
x^{k+\theta} &= \arg \min_x L_{\beta_1}(x, y^k, \lambda^k), \\
\lambda^{k+\theta} &= \lambda - \beta_1 (C x^{k+\theta} + Dy^k - b), \\
y^{k+1-\theta} &= \arg \min_y L_{\beta_2}(x^{k+\theta}, y, \lambda^{k+\theta}), \\
\lambda^{k+1-\theta} &= \lambda^{k+\theta} - \beta_2 (C x^{k+\theta} + Dy^{k+1-\theta} - b).
\end{align*}
\]

The optimality conditions are
\[
f(x) - f(x^{k+\theta}) + (x - x^{k+\theta})^T (-C^T \lambda^{k+\theta}) \geq 0,
\]
(49)
and
\[
g(y) - g(y^{k+1-\theta}) + (y - y^{k+1-\theta})^T (-D^T \lambda^{k+\theta} + \beta_2 D^T (C x^{k+\theta} + Dy^{k+1-\theta} - b)) \geq 0,
\]
(50)
and the dual update \( \lambda^{k+1-\theta} = \lambda^{k+\theta} - \beta_2 (C x^{k+\theta} + Dy^{k+1-\theta} - b) \) is reformulated as
\[
(\lambda - \lambda^{k+\theta})^T (C x^{k+\theta} + Dy^{k+1-\theta} - b + \frac{1}{\beta_2} (\lambda^{k+1-\theta} - \lambda^{k+\theta})) \geq 0, \quad \forall \lambda \in \mathbb{R}^d.
\]
(51)

Combining (49), (50), and (51), we obtain
\[
r(u) - r(u^{k+\theta,k+1-\theta}) + (w - w^{k+\theta,k+1-\theta})^T \mathcal{F}(w^{k+\theta,k+1-\theta}) \\
\geq \beta_2 (D y^{k+1-\theta} - Dy)^T (C x^{k+\theta} + Dy^{k+1-\theta} - b) + \frac{1}{\beta_2} (\lambda - \lambda^{k+\theta})^T (\lambda^{k+\theta} - \lambda^{k+1-\theta}),
\]
(52)
and the right-hand side of (52) reduces to
\[
\frac{1}{2\beta_2} \| \lambda - \lambda^{k+1-\theta} \|^2 - \frac{1}{2\beta_2} \| \lambda - \lambda^{k+\theta} \|^2 + \frac{\beta_2}{2} \| Dy - Dy^{k+1-\theta} \|^2 - \frac{\beta_2}{2} \| C x^{k+\theta} + Dy - b \|^2.
\]
(53)

Substituting (53) into (52) and using the monotonicity of \( \mathcal{F} \), we get
\[
\beta_2 (r(u) - r(u^{k+\theta,k+1-\theta}) + (w - w^{k+\theta,k+1-\theta})^T \mathcal{F}(w)) \\
\geq \frac{\beta_2}{2} \| Dy - Dy^{k+1-\theta} \|^2 - \frac{\beta_2}{2} \| C x^{k+\theta} + Dy - b \|^2 \\
+ \frac{1}{2} \| \lambda - \lambda^{k+1-\theta} \|^2 - \frac{1}{2} \| \lambda - \lambda^{k+\theta} \|^2.
\]
(54)
Third, the optimality condition of Step 2 is
\[
g(y) - g(y^{k+1-\theta}) + (y - y^{k+1-\theta})^T(-D^T\lambda^{k+1} + D^T(\lambda^{k+1} - \lambda^{k+1-\theta})) \geq \frac{\sigma}{2}\|y - y^{k+1-\theta}\|^2, \tag{55}
\]
and the optimality condition of Step 3 is
\[
f(x) - f(x^{k+1}) + (x - x^{k+1})^T(-C^T\lambda^{k+1}) \geq 0, \tag{56}
\]
and the dual update \(\lambda^{k+1} = \lambda^{k+1-\theta} - \beta_1(Cx^{k+1} + Dy^{k+1-\theta} - b)\) is reformulated as
\[
(\lambda - \lambda^{k+1})^T(Cx^{k+1} + Dy^{k+1-\theta} - b + \frac{1}{\beta_1}(\lambda^{k+1} - \lambda^{k+1-\theta})) \geq 0, \quad \forall \lambda \in \mathbb{R}^d. \tag{57}
\]
Combining (55), (56), and (57), we obtain
\[
r(u) - r(u^{k+1,k+1-\theta}) + (w - w^{k+1,k+1-\theta})^T\mathcal{F}(w^{k+1,k+1-\theta}) \geq \frac{\sigma}{2}\|y - y^{k+1-\theta}\|^2 + (Dy - Dy^{k+1-\theta})^T(\lambda^{k+1-\theta} - \lambda^{k+1}) + \frac{1}{\beta_1}(\lambda - \lambda^{k+1})^T(\lambda^{k+1-\theta} - \lambda^{k+1}), \tag{58}
\]
and the right-hand side of (58) reduces to
\[
\frac{\sigma}{2}\|y - y^{k+1-\theta}\|^2 - \frac{\beta_1}{2}\|Dy - Dy^{k+1-\theta}\|^2 + \frac{\beta_1}{2}\|Cx^{k+1} + Dy - b\|^2 + \frac{1}{2\beta_1}\|\lambda - \lambda^{k+1}\|^2 - \frac{1}{2\beta_1}\|\lambda - \lambda^{k+1-\theta}\|^2. \tag{59}
\]
Substituting (59) into (58) and using the monotonicity of \(\mathcal{F}\), we obtain
\[
\beta_1(r(u) - r(u^{k+1,k+1-\theta}) + (w - w^{k+1,k+1-\theta})^T\mathcal{F}(w)) \geq \frac{\sigma\beta_1}{2}\|y - y^{k+1-\theta}\|^2 - \frac{\beta_2}{2}\|Dy - Dy^{k+1-\theta}\|^2 + \frac{\beta_2}{2}\|Cx^{k+1} + Dy - b\|^2 + \frac{1}{2}\|\lambda - \lambda^{k+1}\|^2 - \frac{1}{2}\|\lambda - \lambda^{k+1-\theta}\|^2. \tag{60}
\]
Finally, adding (48), (54), and (60) together, setting \(2\beta_1 + \beta_2 = \Delta t\) and using the condition (41), we have
\[
r(u) - \frac{1}{\Delta t}(\beta_1 r(u^{k+\theta,k}) + \beta_2 r(u^{k+\theta,k+1-\theta}) + \beta_1 r(u^{k+1,k+1-\theta})) + (w - \frac{1}{\Delta t}(\beta_1 w^{k+\theta,k} + \beta_2 w^{k+\theta,k+1-\theta} + \beta_1 w^{k+1,k+1-\theta}))^T\mathcal{F}(w) \geq \frac{1}{2\Delta t}\|\lambda - \lambda^{k+1}\|^2 - \frac{1}{2\Delta t}\|\lambda - \lambda^k\|^2,
\]
and by the convexity of \(r(u)\) and the definition of \(\tilde{w}^{k+1}\), the above inequality reduces to
\[
r(\tilde{w}^{k+1}) - r(u) + (\tilde{w}^{k+1} - w)^T\mathcal{F}(w) \leq \frac{1}{2\Delta t}\|\lambda - \lambda^{k+1}\|^2 - \frac{1}{2\Delta t}\|\lambda - \lambda^k\|^2. \tag{61}
\]
Summing up (61) for \(k = 0, \ldots, n - 1\), we get
\[
\sum_{k=0}^{n-1} r(\tilde{w}^{k+1}) - nr(u) + (\sum_{k=0}^{n-1} \tilde{w}^{k+1} - nw)^T\mathcal{F}(w) \leq \frac{1}{2\Delta t}\|\lambda - \lambda^0\|^2.
\]
Furthermore, by the convexity of \( r(u) \) and the definition of \( w_n \), we have
\[
r(\tilde{w}^n) - r(u) + (\tilde{w}^n - w)^T F(w) \leq \frac{1}{2n \Delta t} \| \lambda - \lambda^0 \|^2.
\]

**Remark 6.** The upper bound of \( \beta_1 \) in Theorem 5.5 is a half of the bound in Lemma 5.4, i.e., the convergence of the iterate (38) is established if \( \beta_1 < 2 \sigma / \| D \|^2 \) while the \( O(1/n) \) ergodic convergence rate can be ensured if \( \beta_1 < \sigma / \| D \|^2 \).

### 5.2.2. Two strongly convex objective functions

If functions \( g \) and \( f \) are both strongly convex, then we can show an \( O(1/n) \) ergodic convergence rate of the iterate (38) in terms of dual objective value. For the sake of simplicity, we consider the case \( b = 0 \) and it is known that the Lagrange dual problem of (37) is
\[
\min_{\lambda \in \mathbb{R}^d} f_1(\lambda) + g_1(\lambda),
\]
where \( f_1(\lambda) = f^*(C^T \lambda) \), \( g_1(\lambda) = g^*(D^T \lambda) \), and \( f^* \) and \( g^* \) are conjugate functions of \( f \) and \( g \), respectively. In order to make our proof clear, we state two well-known results in the literature. First, if \( f \) is a closed proper convex function, then we have
\[
\lambda \in \partial f(x) \iff x \in \partial f^*(\lambda),
\]
see [33, Chapter 23]. Second, if \( f \) is Lipschitz smooth with modulus \( L \in (0, +\infty) \), then the so-called “Descent Lemma” is
\[
f(x) \leq f(y) + (\nabla f(y))^T (x - y) + \frac{L}{2} \| x - y \|^2, \quad \forall x, y \in \text{dom } f,
\]
see [8, Appendix A]. Now we are ready to show the following convergence rate result.

**Theorem 5.6.** Suppose that Assumption 2 is satisfied and \( f \) is strongly convex with modulus \( \delta \in (0, +\infty) \) and
\[
0 < \beta_1 \leq 1/\sigma_1, \ 0 < \beta_2 \leq 1/\delta_1,
\]
where \( \sigma_1 = \| D \|^2 / \sigma \), \( \delta_1 = \| C \|^2 / \delta \). Let \( \{ (x^k, y^k, \lambda^k) \}_{k \in \mathbb{N}} \) be generated by the iterate (38). Then, we have
\[
F(\tilde{\lambda}^n) - F(\lambda) \leq \frac{\| \lambda - \lambda^0 \|^2}{2n \Delta t}, \quad \forall \lambda \in \mathbb{R}^d,
\]
where \( F(\lambda) = f_1(\lambda) + g_1(\lambda) \), \( \Delta t = 2 \beta_1 + \beta_2 \), \( \tilde{\lambda}^n = (\sum_{k=0}^{n-1} \tilde{\lambda}^{k+1}) / n \), and
\[
\tilde{\lambda}^{k+1} = \frac{1}{\Delta t} (\beta_1 \lambda^{k+\theta} + \beta_2 \lambda^{k+1-\theta} + \beta_1 \lambda^{k+1}).
\]

**Proof.** First, since \( f \) and \( g \) are both strongly convex, we have that \( f_1(\lambda) \) and \( g_1(\lambda) \) are Lipschitz smooth with modulus \( \delta_1 \) and \( \sigma_1 \), respectively, see [4, Chapter 18]. Besides, Step 1 in (38) can be regarded as the “Proximal Gradient Method” applied to the dual problem (62), see [7]. Therefore, by [6, Theorem 3.1], we have
\[
2 \beta_1 (F(\lambda) - F(\lambda^{k+\theta})) \geq \| \lambda - \lambda^{k+\theta} \|^2 - \| \lambda - \lambda^k \|^2,
\]
where \( \beta_1 \leq 1/\sigma_1 \).

Second, by the optimality conditions in Steps 1, 2, 3 and the dual updates in (38), we obtain
\[
\begin{align*}
C^T \lambda^{k+\theta} &\in \partial f(x^{k+\theta}), \\
D^T \lambda^{k+1-\theta} &\in \partial g(y^{k+1-\theta}), \\
C^T \lambda^{k+1} &\in \partial f(x^{k+1}),
\end{align*}
\]
and with (63), we can reformulate above relations as

\[
\begin{align*}
C_{x}^{k+\theta} &= \nabla f_1(\lambda^{k+\theta}), \\
Dy^{k+1-\theta} &= \nabla g_1(\lambda^{k+1-\theta}), \\
C_{x}^{k+1} &= \nabla f_1(\lambda^{k+1}).
\end{align*}
\]  

(67)

From (67), the convexity of \( f_1 \), and the descent lemma (64) for \( f_1 \), it is clear that

\[
\begin{align*}
f_1(\lambda) - f_1(\lambda^{k+1-\theta}) &= f_1(\lambda) - f_1(\lambda^{k+\theta}) + f_1(\lambda^{k+\theta}) - f_1(\lambda^{k+1-\theta}) \\
&\geq (C_{x}^{k+\theta} + Dy^{k+1-\theta})^T (\lambda - \lambda^{k+\theta} + \lambda^{k+\theta} - \lambda^{k+1-\theta}) - \frac{\delta_1}{2} \|\lambda^{k+\theta} - \lambda^{k+1-\theta}\|^2 \\
&= (C_{x}^{k+\theta})^T (\lambda - \lambda^{k+1-\theta}) - \frac{\delta_1}{2} \|\lambda^{k+\theta} - \lambda^{k+1-\theta}\|^2,
\end{align*}
\]

(68)

and with the convexity of \( g_1 \), we have

\[
g_1(\lambda) - g_1(\lambda^{k+1-\theta}) \geq (Dy^{k+1-\theta})^T (\lambda - \lambda^{k+1-\theta}).
\]

(69)

Adding (68) and (69) together, we get

\[
\begin{align*}
F(\lambda) - F(\lambda^{k+1-\theta}) &\geq (C_{x}^{k+\theta} + Dy^{k+1-\theta})^T (\lambda - \lambda^{k+1-\theta}) - \frac{\delta_1}{2} \|\lambda^{k+\theta} - \lambda^{k+1-\theta}\|^2 \\
&= \frac{1}{\beta_2} (\|\lambda - \lambda^{k+1-\theta}\|^2 - \|\lambda - \lambda^{k+\theta}\|^2) + \left( \frac{1}{2 \beta_2} - \frac{\delta_1}{2} \right) \|\lambda^{k+\theta} - \lambda^{k+1-\theta}\|^2.
\end{align*}
\]

(70)

Third, by (67), the convexity of \( g_1 \), and the descent lemma (64) for \( g_1 \), we obtain

\[
\begin{align*}
g_1(\lambda) - g_1(\lambda^{k+1}) &= g_1(\lambda) - g_1(\lambda^{k+1-\theta}) + g_1(\lambda^{k+1-\theta}) - g_1(\lambda^{k+1}) \\
&\geq (Dy^{k+1-\theta})^T (\lambda - \lambda^{k+1-\theta} + \lambda^{k+1-\theta} - \lambda^{k+1}) - \frac{\sigma_1}{2} \|\lambda^{k+1} - \lambda^{k+1-\theta}\|^2 \\
&= (Dy^{k+1-\theta})^T (\lambda - \lambda^{k+1}) - \frac{\sigma_1}{2} \|\lambda^{k+1} - \lambda^{k+1-\theta}\|^2,
\end{align*}
\]

(71)

and by the convexity of \( f_1 \), we have

\[
f_1(\lambda) - f_1(\lambda^{k+1}) \geq (C_{x}^{k+1})^T (\lambda - \lambda^{k+1}).
\]

(72)

Adding (71) and (72) together, we have

\[
\begin{align*}
F(\lambda) - F(\lambda^{k+1}) &\geq (C_{x}^{k+1} + Dy^{k+1-\theta})^T (\lambda - \lambda^{k+1}) - \frac{\sigma_1}{2} \|\lambda^{k+1} - \lambda^{k+1-\theta}\|^2 \\
&= \frac{1}{\beta_2} (\|\lambda - \lambda^{k+1}\|^2 - \|\lambda - \lambda^{k+1-\theta}\|^2) + \left( \frac{1}{2 \beta_2} - \frac{\sigma_1}{2} \right) \|\lambda^{k+1} - \lambda^{k+1-\theta}\|^2.
\end{align*}
\]

(73)

Finally, combining (66), (70), and (73), and noticing the condition (65), we obtain

\[
\begin{align*}
F(\lambda) - \frac{1}{\Delta t} \left( \beta_1 F(\lambda^{k+\theta}) + \beta_2 F(\lambda^{k+1-\theta}) + \beta_1 F(\lambda^{k+1}) \right) &\geq \frac{1}{\Delta t} (\|\lambda - \lambda^{k+1}\|^2 - \|\lambda - \lambda^{k}\|^2),
\end{align*}
\]

\[\text{THE GLOWINSKI–LE TALLEC SPLITTING METHOD 17}\]
and by the convexity of $F(\lambda)$ and the definition of $\hat{\lambda}^{k+1}$, we have
\[
F(\hat{\lambda}^{k+1}) - F(\lambda) \leq \frac{1}{2\Delta t} (\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2),
\]
which is followed by
\[
\sum_{k=0}^{n-1} F(\hat{\lambda}^{k+1}) - nF(\lambda) \leq \frac{1}{2\Delta t} \|\lambda - \lambda^0\|^2.
\]
Furthermore, by the convexity of $F(\lambda)$ and the definition of $\bar{\lambda}^n$, we can obtain
\[
F(\bar{\lambda}^n) - F(\lambda) \leq \|\lambda - \lambda^0\|^2 \frac{1}{2n\Delta t}.
\]

Remark 7. The assumptions on functions $f$ and $g$ imply that $(\nabla f_1)^{-1}$ and $(\nabla g_1)^{-1}$ are both strongly monotone. Therefore, it is easy to prove the convergence of the iterate (38) under the condition (65).

Remark 8. Compared to the condition (41) in Theorem 5.5, the relation between $\beta_1$ and $\beta_2$ is relaxed under the condition (65).

5.2.3. A new variant. The $\theta$-scheme (38) is naturally connected to other well-known algorithms, such as Alternating Minimization Algorithm (AMA) [35, 24], PRSM [31, 28] and Alternating Direction Method of Multipliers (ADMM) [22, 13, 9, 26]. Basically, Step 1 in (38) is AMA followed by Step 2 and Step 3 which can be regarded as a generalized PRSM (GPRSM) if $\beta_1 \neq \beta_2$. Therefore, the $\theta$-scheme (38) is a combination of AMA and GPRSM.

It is interesting to explore how we can remove the condition on the $\theta$-scheme that one objective function is strongly convex. The possible reason for this condition is that in the first minimization of Step 1 in (38), the objective consists of a Lagrange function, not an augmented Lagrange function. Thus, we consider adding an augmented Lagrange term in the first minimization of Step 1 and obtain a new algorithm which can be regarded as ADMM+GPRSM

\[
\begin{align*}
\text{Step 1.} & \quad y^k = \arg \min_y L_{\beta_1}(x^k, y, \lambda^k), \\
& \quad x^{k+\theta} = \arg \min_x L_{\beta_1}(x, y^k, \lambda^k), \\
& \quad \lambda^{k+\theta} = \lambda^k - \beta_1(Cx^{k+\theta} + Dy^k - b). \\
\text{(ADMM+GPRSM)} & \quad \text{Step 2.} \quad y^{k+1-\theta} = \arg \min_y L_{\beta_2}(x^{k+\theta}, y, \lambda^{k+\theta}), \\
& \quad \lambda^{k+1-\theta} = \lambda^{k+\theta} - \beta_2(Cx^{k+\theta} + Dy^{k+1-\theta} - b). \\
& \quad \text{Step 3.} \quad x^{k+1} = \arg \min_x L_{\beta_1}(x, y^{k+1-\theta}, \lambda^{k+1-\theta}), \\
& \quad \lambda^{k+1} = \lambda^{k+1-\theta} - \beta_1(Cx^{k+1} + Dy^{k+1-\theta} - b). 
\end{align*}
\]

In the following, we show its convergence and convergence rate.

Theorem 5.7. Suppose $\beta_1 = \beta_2 > 0$. Let $\{(x^k, y^k, \lambda^k)\}_{k \in \mathbb{N}}$ be generated by the iterate (74). Then, we have

(i) $\lambda^k$ converges to $\lambda^*$;
(ii) $Dy^k$ converges to $D^*$;
(iii) $Cx^{k+\theta}$ converges to $b - D^*$. 


where \((x^*, y^*, \lambda^*)\) is a solution of the variational inequality (40).

**Proof.** Since \(\beta_1 = \beta_2\), the iterate (74) can be regarded as ADMM+PRSM and can be obtained by applying a combination of DRSM [12] and PRSM [31] to problem (37). Therefore, we can prove its convergence by the fixed point theory as in [14]. For brevity, we omit the proof here, see [14] for more details.

Next, we adopt notations in Theorem 5.5 and show an \(O(1/n)\) ergodic convergence rate for the iterate (74).

**Theorem 5.8.** Suppose \(\beta_1 = \beta_2 > 0\). Let \(\{(x^k, y^k, \lambda^k)\}_{k \in \mathbb{N}}\) be generated by the iterate (74). Then, we have

\[
r(u) - r(u) + (\bar{w}^n - w)^T F(w) \leq \frac{1}{6n\beta_1} \| \lambda - \lambda^0 \|^2 + \frac{\beta_1}{6n} \| Cx^0 + Dy - b \|^2, \quad \forall w \in \Omega.
\]

**Proof.** First, the optimality conditions of Step 1 of the iterate (74) are

\[
g(y) - g(y^k) + (y - y^k)^T (-D^T (\lambda^{k+\theta} + \lambda^k - \lambda^{k+\theta}) + \beta_1 D^T (Cx^k + Dy^k - b)) \geq 0,
\]

and

\[
f(x) - f(x^{k+\theta}) + (x - x^{k+\theta})^T (-C^T \lambda^{k+\theta}) \geq 0.
\]

Combining these with

\[
(\lambda - \lambda^{k+\theta})^T (C x^{k+\theta} + Dy^{k} - b + \frac{1}{\beta_1} (\lambda^{k+\theta} - \lambda^k)) \geq 0, \quad \forall \lambda \in \mathbb{R}^d,
\]

we have

\[
\beta_1 (r(u) - r(u^{k+\theta,k}) + (w - w^{k+\theta,k})^T F(w)) \\
\geq \beta_1 (Dy - Dy^k)^T (\lambda^k - \lambda^{k+\theta} - \beta_1 (Cx^k + Dy^k - b)) + (\lambda - \lambda^{k+\theta})^T (\lambda^k - \lambda^{k+\theta}) \\
= \beta_1^2 (Dy - Dy^k)^T (-Cx^k + b - (Cx^{k+\theta} + b)) + (\lambda - \lambda^{k+\theta})^T (\lambda^k - \lambda^{k+\theta}) \\
= \frac{\beta_1^2}{2} (\|Cx^{k+\theta} + Dy - b\|^2 - \|Cx^k + Dy - b\|^2 + \|Cx^k + Dy^k - b\|^2) \\
+ \frac{1}{2} (\|\lambda - \lambda^{k+\theta}\|^2 - \|\lambda - \lambda^k\|^2).
\]

Second, without the strong convexity of function \(g\), (60) reduces to

\[
\beta_1 (r(u) - r(u^{k+1,k+1-\theta}) + (w - w^{k+1,k+1-\theta})^T F(w)) \\
\geq -\frac{\beta_1^2}{2} \|Dy - Dy^{k+1-\theta}\|^2 + \frac{\beta_1^2}{2} \|Cx^{k+1} + Dy - b\|^2 \\
+ \frac{1}{2} (\|\lambda - \lambda^{k+1-\theta}\|^2 - \|\lambda - \lambda^{k+1-\theta}\|^2).
\]

Finally, adding (75), (54), and (76) together, using the convexity of function \(r\) and \(\beta_1 = \beta_2\), we get

\[
r(\bar{u}^{k+1}) - r(u) + (\bar{w}^{k+1} - w)^T F(w) \\
\leq \frac{1}{6\beta_1} (\|\lambda - \lambda^k\|^2 - \|\lambda - \lambda^{k+1}\|^2) \\
+ \frac{\beta_1}{6} (\|Cx^k + Dy - b\|^2 - \|Cx^{k+1} + Dy - b\|^2 - \|Cx^k + Dy^k - b\|^2).
\]
Summing up (77) for $k = 0, \ldots, n - 1$, we have

$$\sum_{k=0}^{n-1} r(\tilde{u}^{k+1}) - nr(u) + (\sum_{k=0}^{n-1} \tilde{w}^{k+1} - nw)^T F(w) \leq \frac{1}{6n\beta_1} \|\lambda - X^0\|^2 + \frac{\beta_1}{6} (\|Cz^0 + Dy - b\|^2 - \sum_{k=0}^{n-1} \|Cx^k + Dy^k - b\|^2).$$

Furthermore, by the convexity of $r(u)$ and the definition of $\bar{w}$, we have

$$r(\bar{u}^n) - r(u) + (\bar{w}^n - w)^T F(w) \leq \frac{1}{6n\beta_1} \|\lambda - X^0\|^2 + \frac{\beta_1}{6n} \|Cz^0 + Dy - b\|^2.$$

**Remark 9.** It is noted that Theorem 5.7 and Theorem 5.8 are only established under the condition that $\beta_1 = \beta_2$. It is still unknown whether the iterate (74) is theoretically convergent without this condition, or with a more relaxed condition.

6. **Numerical experiments.** In this section, we first compare the $\theta$-scheme, our proposed ADMM+GPRSM (74), general ADMM (GADMM) [22, 13, 1], and Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [6, 5] for an image deblurring problem. Second, we verify the efficiency of our proposed inexact $\theta$-scheme (28) for the Least Absolute Shrinkage and Selection Operator (LASSO). In particular, we compare the inexact $\theta$-scheme (28) and inexact $\theta$-schemes using other three different criterions in the subproblem (details will be described precisely later). Our code was written on MATLAB 2017b. For image deblurring, our code was based on sources in Amir Beck’s homepage which are publicly available. All experiments were conducted in a laptop computer with a 2.9GHz i7 processor and an 8GB memory.

6.1. **Total variation based image deblurring.** Here we consider recovering an unknown image $z \in \mathbb{R}^n$ from an observation $q \in \mathbb{R}^m$. We focus on a linear inverse problem of the form

$$Qz = q + w,$$

where $Q \in \mathbb{R}^{m \times n}$ is a known blurring operator, $w$ is a vector standing for unknown noise. The vectors $z$ and $q$ are obtained by stacking the columns of their matrix form in lexicographical order.

Normally, it is not possible to recover $z$ from $q$ by solving equation (78) directly. Instead, various models are proposed to approach the true $z$. Here we consider solving the following convex non-smooth minimization problem

$$\min_z \|Qz - q\|^2 + 2\eta TV(z) \ (\eta > 0),$$

where isotropic $TV(\cdot)$ represents a discrete total variation semi-norm

$$TV(z) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(z_{i,j} - z_{i+1,j})^2 + (z_{i,j} - z_{i,j+1})^2}$$

$$+ \sum_{i=1}^{M-1} |z_{i,N} - z_{i+1,N}| + \sum_{j=1}^{N-1} |z_{M,j} - z_{M,j+1}|,$$
If \( f(z) = 2\eta TV(z), g(z) = \|Qz - q\|^2 \), then the \( \theta \)-scheme (20) applied to solve (79) becomes

\[
\begin{aligned}
v^{k+1} &= z^k - 2\beta_1 Q^T(Qz^k - q), \\
v'_{k+1} &= \arg\min_z 2\beta_1 \eta TV(z) + \frac{1}{2}\|z - v^{k+\frac{1}{2}}\|^2, \\
v_{k+1} &= \beta_2/\beta_1 (\gamma v^{k+\frac{1}{2}} - v^{k+\frac{1}{2}}), \\
v''_{k+1} &= \arg\min_v \beta_2 \|Qv - q\|^2 + \frac{1}{2}\|v - v^{k+1}\|^2, \\
v''_{k+1} &= v^{k+\frac{1}{2}} - 2\beta_1 Q^T(Qv^{k+\frac{1}{2}} - q), \\
\lambda_{k+1} &= \arg\min_z 2\beta_1 \eta TV(z) + \frac{1}{2}\|z - v^{k+1}\|^2,
\end{aligned}
\]

where \( \gamma = 1 + \beta_1/\beta_2 \) and \( \beta_1, \beta_2 \in (0, +\infty) \) are two positive constants. The first and last minimization problem above are TV-based image denoising which can be solved exactly by Fast Gradient-Based Projection (FGP) [5]. The second minimization problem above reduces to a linear equation

\[
(2\beta_2 Q^TQ + I)v^{k+\frac{1}{2}} = 2\beta_2 Q^Tq + v^{k+1},
\]

which can be solved exactly by Cosine Discrete Transform [2, 32].

For GADMM, we first reformulate unconstrained problem (79) as

\[
\begin{align*}
\min_{z, y} &\quad \|Qz - q\|^2 + 2\eta TV(y), \\
\text{s.t.} &\quad z = y,
\end{align*}
\]

then the related iterate is

\[
\begin{aligned}
z^{k+1} &= \arg\min_z \|Qz - q\|^2 - (\lambda^k)^T(z - y^k) + \frac{\beta}{2}\|z - y^k\|^2, \\
y^{k+1} &= \arg\min_y 2\eta TV(y) - (\lambda^k)^T(z^{k+1} - y) + \frac{\beta}{2}\|\rho z^{k+1} + (1 - \rho)y^k - y\|^2, \\
\lambda^{k+1} &= \lambda^k - \beta(\rho z^{k+1} + (1 - \rho)y^k - y^{k+1}),
\end{aligned}
\]

where \( \beta \in (0, +\infty) \) is a constant, \( \rho \in (0, 2) \) is a relaxation factor, and \( \lambda^k \) is Lagrange multiplier. Similarly, the above two minimization problems can be solved efficiently.

For ADMM+GPRSM (74), we have

\[
\begin{aligned}
z^{k+\frac{1}{2}} &= \arg\min_z \|Qz - q\|^2 - (\lambda^{k+\frac{1}{2}})^T(z - y^{k+\frac{1}{2}}) + \frac{\beta_1}{2}\|z - y^{k+\frac{1}{2}}\|^2, \\
y^{k+\frac{1}{2}} &= \arg\min_y 2\eta TV(y) - (\lambda^{k+\frac{1}{2}})^T(z^{k+\frac{1}{2}} - y) + \frac{\beta_1}{2}\|z^{k+\frac{1}{2}} - y\|^2, \\
\lambda^{k+\frac{1}{2}} &= \lambda^{k+\frac{1}{2}} - \beta_1 (z^{k+\frac{1}{2}} - y^{k+\frac{1}{2}}), \\
z^{k+1} &= \arg\min_z \|Qz - q\|^2 - (\lambda^{k+1})^T(z - y^{k+1}) + \frac{\beta_2}{2}\|z - y^{k+1}\|^2, \\
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \beta_2 (z^{k+1} - y^{k+1}), \\
y^{k+1} &= \arg\min_y 2\eta TV(y) - (\lambda^{k+1})^T(z^{k+1} - y) + \frac{\beta_1}{2}\|z^{k+1} - y\|^2, \\
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \beta_1 (z^{k+1} - y^{k+1}),
\end{aligned}
\]

where \( \beta_1, \beta_2 \in (0, +\infty) \) are two positive constants. Similar to the \( \theta \)-scheme, all minimization problems above can be solved efficiently.
For FISTA, the specific iterate is
\[
\begin{aligned}
    z^k &= \arg \min_z 2\eta TV(z) + \frac{L}{2} \|z - (y^k - \frac{2}{L} Q^T (Q y^k - q))\|^2, \\
t_{k+1} &= \frac{2}{1 + \sqrt{1 + 4t_k^2}} \\
y^{k+1} &= z^k + t_{k+1} \left( z^k - z^{k-1} \right),
\end{aligned}
\]
where \( t_0 = 1 \) and \( L \) is the Lipschitz constant of function \( 2Q^T (Qz - q) \). Similarly, the above minimization problem can be solved by FGP.

For fair comparison, we first run GADMM with \( \rho = 1.5 \) and \( \beta = 1 \) and the related stopping criterion is
\[
|{\text{Objective}_k - \text{Objective}_{k-1}}| < 10^{-10}.
\]
Then, for other three algorithms, we use the objective value obtained by GADMM to formulate the stopping criterion as
\[
\|Qz^k - q\|^2 + 2\eta TV(z^k) \leq \text{Objectve}_{\text{GADMM}},
\]
where \( z^k \) is obtained in the iteration.

We also apply two quantities to judge the quality of recovered image. The error of recovery is measured by the mean squared error (MSE)
\[
\text{MSE} = \frac{\|z_{\text{ori}} - z_{\text{rec}}\|^2}{m \times n},
\]
where \( z_{\text{ori}} \) is the original image and \( z_{\text{rec}} \) is the recovered image. The improved signal-to-noise ratio (ISNR) is defined as
\[
\text{ISNR} = 10 \log_{10} \left( \frac{\|z_{\text{ori}} - z_{\text{dis}}\|^2}{\|z_{\text{ori}} - z_{\text{rec}}\|^2} \right),
\]
where \( z_{\text{dis}} \) is the distorted image.

Now we consider a 256 \( \times \) 256 “cameraman” test image (whose pixels are scaled to be between 0 and 1) and set \( Q \) as a uniform \( 3 \times 3 \) blur with an additive zero-mean white Gaussian noise with standard deviation \( 10^{-4} \). Note that we use reflexive boundary conditions in the blurring operation. The regularization parameter \( \eta \) is chosen to be \( 10^{-4} \). It is computed that the Lipschitz constant \( L(2Q^T (Qz - q)) = 2 \). The maximum outer iterations are uniformly set as 3000. For all TV-based image denoising subproblems, we use FGP with a warm-start strategy and the stopping criterion of FGP is to judge whether the relative error of dual variable is less than a fixed value, see [5] for more details.

Since \( L(2Q^T (Qz - q)) = 2 \), we have \( 0 < \beta_2 \leq \beta_1 < 1 \) for the \( \theta \)-scheme. Therefore, it is necessary to do a sensitivity test on \( \beta_1 \) and \( \beta_2 \). Setting precision of FGP as \( 10^{-4} \), we record iterations and times obtained by the \( \theta \)-scheme on six pairs of \( \beta_1 \) and \( \beta_2 \). Table 1 shows that as both \( \beta_1 \) and \( \beta_2 \) get closer to 1, the \( \theta \)-schemes converge faster. So, we choose aggressively \( \beta_1 = \beta_2 = 1 \) for the \( \theta \)-scheme. As for ADMM+GPRSM, according to Theorem 5.7, we set \( \beta_1 = \beta_2 = 1 \).

Second, as the image denoising subproblem is not exactly solved by FGP, we must test influence of different precisions of FGP (inner precisions) for the entire algorithm. We use five inner precisions, which are \( 1e-02, 1e-04, 1e-06, 1e-8, \) and \( 1e-10 \), for all four algorithms. The results are reported in Table 2 where the \( \theta \)-scheme, ADMM+GPRSM, and GPRSM converge under all inner precisions while FISTA is not feasible for \( 1e-02, 1e-04, \) and \( 1e-6 \). Furthermore, the \( \theta \)-scheme is the
Table 1. A sensitivity test of the $\theta$-scheme on TV-based deblurring

| $\beta_1$ | $\beta_2$ | Iteration | Time (s) | Objective | MSE | ISNR |
|----------|----------|-----------|----------|-----------|-----|------|
| 0.1      | 0.1      | 3000      | 106.41   | 0.6259    | 1.6135e-04 | 11.9101 |
| 0.3      | 0.2      | 2852      | 101.04   | 0.6259    | 1.5999e-04 | 11.9470 |
| 0.7      | 0.5      | 1202      | 42.45    | 0.6259    | 1.5999e-04 | 11.9470 |
| 0.8      | 0.5      | 1088      | 38.57    | 0.6259    | 1.5999e-04 | 11.9470 |
| 0.9      | 0.8      | 879       | 31.01    | 0.6259    | 1.5999e-04 | 11.9470 |
| 0.9      | 0.9      | 847       | 29.82    | 0.6259    | 1.5999e-04 | 11.9470 |

Table 2. Comparison of $\theta$-scheme, ADMM+GPRSM, GADMM, and FISTA on TV-based deblurring under same inner precision

| Inner precision | Algorithm         | Iteration | Mean/Max FGP | Time (s) | Objective | MSE | ISNR |
|-----------------|-------------------|-----------|--------------|----------|-----------|-----|------|
| 1e-02           | $\theta$-scheme   | 763       | 4.00/4       | 29.00    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | ADMM+GPRSM        | 763       | 4.00/4       | 30.00    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | GADMM             | 1526      | 2.00/2       | 33.43    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | FISTA             | 3000      | 2.00/2       | 53.92    | 0.6261    | 1.5999e-04 | 11.9489 |
| 1e-04           | $\theta$-scheme   | 763       | 4.00/6       | 28.84    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | ADMM+GPRSM        | 763       | 4.00/6       | 29.39    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | GADMM             | 1526      | 2.00/3       | 33.13    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | FISTA             | 3000      | 2.00/3       | 53.70    | 0.6261    | 1.5999e-04 | 11.9487 |
| 1e-06           | $\theta$-scheme   | 763       | 4.00/20      | 29.28    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | ADMM+GPRSM        | 763       | 4.00/20      | 29.29    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | GADMM             | 1526      | 2.27/10      | 32.42    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | FISTA             | 3000      | 2.27/10      | 54.16    | 0.6259    | 1.5999e-04 | 11.9470 |
| 1e-08           | $\theta$-scheme   | 763       | 8.60/20      | 38.31    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | ADMM+GPRSM        | 763       | 8.60/20      | 39.30    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | GADMM             | 1526      | 4.32/10      | 42.45    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | FISTA             | 1277      | 7.62/10      | 44.74    | 0.6259    | 1.5997e-04 | 11.9476 |
| 1e-10           | $\theta$-scheme   | 763       | 16.61/20     | 58.43    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | ADMM+GPRSM        | 763       | 16.61/20     | 59.79    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | GADMM             | 1525      | 8.44/10      | 63.16    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | FISTA             | 1214      | 9.87/10      | 51.49    | 0.6259    | 1.5997e-04 | 11.9476 |

Table 3. Comparison of $\theta$-scheme, ADMM+GPRSM, GADMM, and FISTA on TV-based deblurring under different inner precisions.

| Inner precision | Algorithm         | Iteration | Mean/Max FGP | Time (s) | Objective | MSE | ISNR |
|-----------------|-------------------|-----------|--------------|----------|-----------|-----|------|
| 1e-02           | $\theta$-scheme   | 763       | 4.00/4       | 27.87    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | ADMM+GPRSM        | 763       | 4.00/4       | 30.00    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | GADMM             | 1526      | 2.00/3       | 33.01    | 0.6259    | 1.5999e-04 | 11.9470 |
| 1e-04           | GADMM             | 1526      | 2.00/3       | 33.01    | 0.6259    | 1.5999e-04 | 11.9470 |
|                 | FISTA             | 1277      | 7.62/10      | 47.88    | 0.6259    | 1.5997e-04 | 11.9476 |

fastest among all four algorithms under near all inner precisions (except 1e-10). In addition, ADMM+GPRSM is comparable to the $\theta$-scheme with similar objective value, MSE, and ISNR.

Next, we compare all four algorithms under different inner precisions. According to Table 2, we choose 1e-04 for GADMM, 1e-02 for the $\theta$-scheme and ADMM+GPRSM, and 1e-08 for FISTA. The results are reported in Table 3 where the $\theta$-scheme is again the fastest among all algorithms. The mean/max FGP of the $\theta$-scheme and ADMM+GPRSM are the same (4.00/4) while the $\theta$-scheme is near 8% faster than the second fastest ADMM+GPRSM which is faster than GADMM by around 10%. FISTA is apparently the slowest among all algorithms.
Finally, we present the original image, blurred image, and recovered image obtained by the $\theta$-scheme in Fig. 1 which demonstrate the efficiency of the $\theta$-scheme for image deblurring. Besides, we plot objective value, MSE, and ISNR for all four algorithms in the first 30 iterations in Fig. 2. It can be seen that the $\theta$-scheme has a lower objective value and MSE, a higher ISNR than FISTA and GADMM. As for the $\theta$-scheme and ADMM+GPRSM, these quantities are nearly the same.

6.2. Least absolute shrinkage and selection operator. In this example, we verify the efficiency of our proposed inexact $\theta$-scheme (28) for LASSO [34]

$$\min_{z \in \mathbb{R}^n} \frac{1}{2}\|Qz - q\|^2 + \eta \|z\|_1 \ (\eta > 0),$$

where $Q \in \mathbb{R}^{m \times n}$ is a feature matrix. The corresponding $\theta$-scheme is

$$v^k = (\beta_2 / \beta_1)(\gamma \text{prox}_{\beta_1 \eta \| \cdot \|_1} - I)(z^k - \beta_1 Q^T (Qz^k - q)),$$

$$w^{k+1} = (\beta_2 Q^T Q + I)^{-1} (\beta_2 Q^T q + v^k),$$

$$z^{k+1} = \text{prox}_{\beta_1 \eta \| \cdot \|_1} (w^{k+1} - \beta_1 Q^T (Qw^{k+1} - q)).$$
and it is known that
\[ \text{prox}_{\beta_1 \eta \| \cdot \|_1}(v) = (v - \beta_1 \eta)_+ - (-v - \beta_1 \eta)_+, \]
which means that both (80) and (82) can be solved efficiently. The point here is how to solve linear system (81) effectively, we show four techniques below.

First, we use LSQR [30] (with a fixed relative precision 1e-06) which is referred as lsqr in MATLAB to solve linear system (81) and denote the related algorithm by IN-\(\theta_{\text{LSQR}}\)
\[
\begin{align*}
\text{v}^k &= (\beta_2/\beta_1)(\gamma \text{prox}_{\beta_1 \eta \| \cdot \|_1} - I)(z^k - \beta_1 Q^T(Qz^k - q)), \\
\text{w}^{k+1} &= \beta_2(h^k - Q^T \eta^k), \\
\text{z}^{k+1} &= \text{prox}_{\beta_1 \eta \| \cdot \|_1}(w^{k+1} - \beta_1 Q^T(Qw^{k+1} - q)),
\end{align*}
\]
where \(h^k = Q^T q + v^k/\beta_2\) and the Cholesky decomposition is denoted by \text{chol} in MATLAB.

Second, according to (29), we use Cholesky decomposition to solve linear system (81) and the related algorithm is denoted by IN-\(\theta_{\text{Cholesky}}\)
\[
\begin{align*}
\text{v}^k &= (\beta_2/\beta_1)(\gamma \text{prox}_{\beta_1 \eta \| \cdot \|_1} - I)(z^k - \beta_1 Q^T(Qz^k - q)), \\
\text{w}^{k+1} &= \beta_2(h^k - Q^T \eta^k), \\
\text{z}^{k+1} &= \text{prox}_{\beta_1 \eta \| \cdot \|_1}(w^{k+1} - \beta_1 Q^T(Qw^{k+1} - q)),
\end{align*}
\]
where \(h^k = Q^T q + v^k/\beta_2\) and the Cholesky decomposition is denoted by \text{chol} in MATLAB.

Third, we adopt our proposed inexactness criterion (30) to solve linear system (81) and the related algorithm is denoted by IN-\(\theta\)
\[
\begin{align*}
\text{v}^k &= (\beta_2/\beta_1)(\gamma \text{prox}_{\beta_1 \eta \| \cdot \|_1} - I)(z^k - \beta_1 Q^T(Qz^k - q)), \\
\text{w}^{k+1} &= \beta_2(h^k - Q^T \eta^k), \\
\text{z}^{k+1} &= \text{prox}_{\beta_1 \eta \| \cdot \|_1}(w^{k+1} - \beta_1 Q^T(Qw^{k+1} - q)),
\end{align*}
\]
where \(e_k(\eta) := Q\beta_2 h^k - (\beta_2 QQ^T + I)\eta\). Throughout this subsection \(\alpha\) is fixed at 0.9.

Finally, we solve linear system (81) with a fixed relative precision and the related algorithm is denoted by IN-\(\theta_{1e^{-t}}\)
\[
\begin{align*}
\text{v}^k &= (\beta_2/\beta_1)(\gamma \text{prox}_{\beta_1 \eta \| \cdot \|_1} - I)(z^k - \beta_1 Q^T(Qz^k - q)), \\
\text{w}^{k+1} &= \beta_2(h^k - Q^T \eta^k), \\
\text{z}^{k+1} &= \text{prox}_{\beta_1 \eta \| \cdot \|_1}(w^{k+1} - \beta_1 Q^T(Qw^{k+1} - q)),
\end{align*}
\]
where \(t\) takes 2, 4, 6, 8 and 10. The inexactness criterions in IN-\(\theta\) and IN-\(\theta_{1e^{-t}}\) can be achieved by different linear solvers, we use Conjugate Gradient Method [15, 16] here and it is denoted by \text{cgs} in MATLAB.

Outer stopping criterion for IN-\(\theta_{\text{LSQR}}\) and IN-\(\theta_{\text{Cholesky}}\) is
\[ \|z^{k+1} - w^{k+1}\| + \|w^{k+1} - z^k\| < \varepsilon, \]
where $\varepsilon = 10^{-5}$, while for all other algorithms, the stopping criterion is
\[
\frac{1}{2} \| Qz^k - q \|^2 + \eta \| z^k \|_1 \leq \min\{ \text{Objective}_{\text{IN-}\theta_{\text{LSQR}}}, \text{Objective}_{\text{IN-}\theta_{\text{Cholesky}}} \},
\]
where $z^k$ is obtained in the iteration. In the following, we compare all algorithms on synthetic and real dataset, respectively.

6.2.1. Synthetic dataset. Here we generate six pairs of synthetic LASSO models and the difference mainly lies in the dimensionality and sparsity of $Q$ and $q$. The $Q$ is given by MATLAB function \texttt{sprand}(m,n,s) where $(m,n)$ and $s$ are the dimension and sparsity of $Q$, respectively. If a vector $x_0 \in \mathbb{R}^n$ is defined as \texttt{sprand}(n,1,p) where $p = 100/n$, then the vector $q$ is set as $Qx_0 + 0.1\varepsilon$ where $\varepsilon \in \mathbb{R}^m$ is a standard normally distributed random noise vector. Same as [9], the parameter $\eta$ is set to be $0.1\max(Q^Tq)$ for controlling the sparsity of solution.

The Lipschitz constant $L = \|Q^TQ\|$ and we use script \texttt{normest} to give its value. Because of memory restriction in MATLAB, if the dimension of $Q$ is equal to or larger than $10^5 \times 10^5$ scale, then $L$ is set to be $\|Q\|^2$. In addition, the parameters $\beta_1$ and $\beta_2$ are set to be equal and $\beta_1 = 2/L$. Table 4 lists all referred parameters where the synthetic dataset is generated with gradually increasing dimensionality and sparsity.

\begin{table}[h]
\centering
\caption{Values of $\eta$, $L$ and $\beta_1(\beta_2)$ for synthetic dataset}
\begin{tabular}{|c|c|c|c|}
\hline
$\text{(m,n,s)}$ & $\eta$ & $L$ & $\beta_1(\beta_2)$ \\
\hline
$(1 \times 10^4, 1.5 \times 10^4, 50\%)$ & 878.32 & 1.8416e+04 & 1.0860e-04 \\
$(1 \times 10^4, 1.5 \times 10^4, 10\%)$ & 293.15 & 4.4668e+03 & 4.4775e-04 \\
$(1.5 \times 10^4, 2 \times 10^4, 5\%)$ & 174.12 & 3.2264e+03 & 6.1989e-04 \\
$(2 \times 10^4, 3 \times 10^4, 1\%)$ & 67.90 & 9.4158e+02 & 2.1281e-03 \\
$(2 \times 10^4, 3 \times 10^4, 0.1\%)$ & 58.27 & 9.0143e+02 & 2.2187e-03 \\
$(3 \times 10^4, 2 \times 10^4, 0.01\%)$ & 9.59 & 3.6101e+02 & 5.5400e-03 \\
\hline
\end{tabular}
\end{table}

We set maximum outer iterations to be 500 and inner maximum iterations for LSQR and \texttt{cgs} to be 200 where the initial points in the inner loop are points obtained from the last step. The results are reported in Table 5 where we record iterations, mean/max Conjugate Gradient (CG) steps, times, and objective values.

First, it shows that our proposed IN-$\theta$ is the fastest among all algorithms and is $11\% \sim 44\%$ faster than the second fastest IN-$\theta_{1e-6}$. Besides, the performance of IN-$\theta$ indicates that the mean CG in $[0.89, 0.90]$ is enough to ensure the convergence for the entire algorithm.

Second, it can be seen that IN-$\theta_{1e-2}$ and IN-$\theta_{1e-4}$ do not stop in 500 iterations which implies that the mean CG (less than 0.1) is too small to ensure the convergence for the entire algorithm.

Third, all IN-$\theta_{1e-6}$, IN-$\theta_{1e-8}$ and IN-$\theta_{1e-10}$ are convergent. As the precision becomes more tight, the computing times become longer.

Finally, the results in Table 5 show that IN-$\theta_{\text{LSQR}}$ is slower than IN-$\theta_{1e-8}$ but faster than IN-$\theta_{1e-10}$ except the fifth example where IN-$\theta_{\text{LSQR}}$ is even faster than IN-$\theta_{1e-8}$ but slower than IN-$\theta_{1e-6}$. Because Cholesky decomposition is time consuming, IN-$\theta_{\text{Cholesky}}$ is the slowest among all algorithms. In addition, we did not implement IN-$\theta_{\text{Cholesky}}$ for the last two examples due to the large dimension.
Table 5. Comparison of IN-$\theta_{\text{LSQR}}$, IN-$\theta_{\text{Cholesky}}$, IN-$\theta_{1e-t}$ and IN-$\theta$ on synthetic dataset

| (m,n,s) | Algorithm     | Iteration | Mean/Max CG | Time (s) | Objective |
|---------|---------------|-----------|-------------|----------|-----------|
| $(1 \times 10^4, 1.5 \times 10^4, 50\%)$ | IN-$\theta_{\text{LSQR}}$ | 9 | $\sim / \sim$ | 24.72 | 5.8157e+04 |
|          | IN-$\theta_{\text{Cholesky}}$ | 9 | $\sim / \sim$ | 710.26 | 5.8157e+04 |
|          | IN-$\theta_{1e-10}$ | 10 | 7.00/10 | 27.00 | 5.8157e+04 |
|          | IN-$\theta_{1e-8}$ | 9 | 5.44/8 | 20.46 | 5.8157e+04 |
|          | IN-$\theta_{1e-6}$ | 9 | 3.44/6 | 15.51 | 5.8157e+04 |
|          | IN-$\theta_{1e-4}$ | 500 | 0.03/4 | 319.16 | 5.8157e+04 |
|          | IN-$\theta_{1e-2}$ | 500 | 0.01/2 | 314.86 | 5.8157e+04 |
|          | IN-$\theta$ | 10 | 0.90/1 | 22.91 | 5.8157e+04 |
| $(1 \times 10^4, 1.5 \times 10^4, 10\%)$ | IN-$\theta_{\text{LSQR}}$ | 9 | $\sim / \sim$ | 6.36 | 1.6892e+04 |
|          | IN-$\theta_{\text{Cholesky}}$ | 9 | $\sim / \sim$ | 61.89 | 1.6892e+04 |
|          | IN-$\theta_{1e-10}$ | 9 | 7.11/10 | 6.17 | 1.6892e+04 |
|          | IN-$\theta_{1e-8}$ | 9 | 5.44/8 | 5.11 | 1.6892e+04 |
|          | IN-$\theta_{1e-6}$ | 9 | 3.44/6 | 3.87 | 1.6892e+04 |
|          | IN-$\theta_{1e-4}$ | 500 | 0.03/4 | 77.57 | 1.6892e+04 |
|          | IN-$\theta_{1e-2}$ | 500 | 0.01/2 | 79.92 | 1.6892e+04 |
|          | IN-$\theta$ | 9 | 0.89/1 | 2.68 | 1.6892e+04 |
| $(1.5 \times 10^4, 2 \times 10^4, 5\%)$ | IN-$\theta_{\text{LSQR}}$ | 8 | $\sim / \sim$ | 5.98 | 1.0405e+04 |
|          | IN-$\theta_{\text{Cholesky}}$ | 8 | $\sim / \sim$ | 96.69 | 1.0405e+04 |
|          | IN-$\theta_{1e-10}$ | 9 | 7.11/10 | 6.57 | 1.0405e+04 |
|          | IN-$\theta_{1e-8}$ | 8 | 5.63/8 | 4.97 | 1.0405e+04 |
|          | IN-$\theta_{1e-6}$ | 8 | 3.63/6 | 3.85 | 1.0405e+04 |
|          | IN-$\theta_{1e-4}$ | 500 | 0.03/4 | 84.29 | 1.0405e+04 |
|          | IN-$\theta_{1e-2}$ | 500 | 0.01/2 | 83.01 | 1.0405e+04 |
|          | IN-$\theta$ | 9 | 0.89/1 | 2.81 | 1.0405e+04 |
| $(2 \times 10^4, 3 \times 10^4, 1\%)$ | IN-$\theta_{\text{LSQR}}$ | 10 | $\sim / \sim$ | 2.94 | 4.7993e+03 |
|          | IN-$\theta_{\text{Cholesky}}$ | 10 | $\sim / \sim$ | 232.74 | 4.7993e+03 |
|          | IN-$\theta_{1e-10}$ | 10 | 7.10/10 | 3.24 | 4.7993e+03 |
|          | IN-$\theta_{1e-8}$ | 11 | 4.82/8 | 2.73 | 4.7993e+03 |
|          | IN-$\theta_{1e-6}$ | 10 | 3.30/6 | 1.96 | 4.7993e+03 |
|          | IN-$\theta_{1e-4}$ | 500 | 0.03/4 | 36.20 | 4.7993e+03 |
|          | IN-$\theta_{1e-2}$ | 500 | 0.01/2 | 35.20 | 4.7993e+03 |
|          | IN-$\theta$ | 10 | 0.90/1 | 1.74 | 4.7993e+03 |
| $(2 \times 10^5, 3 \times 10^5, 0.1\%)$ | IN-$\theta_{\text{LSQR}}$ | 9 | $\sim / \sim$ | 38.52 | 4.9942e+03 |
|          | IN-$\theta_{1e-10}$ | 9 | 7.22/10 | 59.94 | 4.9942e+03 |
|          | IN-$\theta_{1e-8}$ | 9 | 5.44/8 | 41.94 | 4.9942e+03 |
|          | IN-$\theta_{1e-6}$ | 9 | 3.44/6 | 31.82 | 4.9942e+03 |
|          | IN-$\theta_{1e-4}$ | 500 | 0.03/4 | 644.25 | 4.9942e+03 |
|          | IN-$\theta_{1e-2}$ | 500 | 0.01/2 | 644.81 | 4.9942e+03 |
|          | IN-$\theta$ | 9 | 0.89/1 | 20.72 | 4.9942e+03 |
| $(3 \times 10^5, 2 \times 10^6, 0.01\%)$ | IN-$\theta_{\text{LSQR}}$ | 38 | $\sim / \sim$ | 159.14 | 2.1247e+03 |
|          | IN-$\theta_{1e-10}$ | 37 | 5.49/8 | 180.66 | 2.1247e+03 |
|          | IN-$\theta_{1e-8}$ | 37 | 3.95/7 | 146.45 | 2.1247e+03 |
|          | IN-$\theta_{1e-6}$ | 37 | 2.43/5 | 113.97 | 2.1247e+03 |
|          | IN-$\theta_{1e-4}$ | 500 | 0.08/4 | 707.33 | 2.1247e+03 |
|          | IN-$\theta_{1e-2}$ | 500 | 0.02/2 | 693.12 | 2.1247e+03 |
|          | IN-$\theta$ | 37 | 0.97/1 | 91.95 | 2.1247e+03 |

6.2.2. Real dataset. In this case, we verify the efficiency of our proposed IN-$\theta$ on two real dataset which can be found in https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/. It is seen that without any change, residual $\|e_k(n^k)\|$ obtained by IN-$\theta$ decreases so fast such that after a few iterations it is nearly zero which severely increases the computing time of sequent iterations of IN-$\theta$. Therefore,
we apply a “RESTART” technique to IN-θ (denoted by IN-θR) where we restart IN-θ after every fixed steps and this can get rid of rapid decreasing of \( \| e_k(\eta^k) \| \). In the following, IN-θ is restarted after every 3 steps.

The first dataset is called “news20.scale” where the dimension of the feature matrix \( Q \) is \( 15935 \times 62061 \). Then, we have that the Lipschitz constant \( L = 520.83, \) the parameters \( \beta_1 = \beta_2 = 2/L = 0.0038, \) and the parameter \( \eta = 1.1989e+03. \)

The results are reported in Table 6. Similarly, IN-θR is the fastest and faster than the second fastest IN-\( \theta_{1e-6} \) by around 23%; IN-\( \theta_{1e-8} \) and IN-\( \theta_{1e-10} \) are slower than IN-\( \theta_{1e-6} \) by near 29% and 53%, respectively; IN-\( \theta_{LSQR} \), IN-\( \theta_{1e-2} \), and IN-\( \theta_{1e-4} \) did not stop in 500 iterations. In Fig. 3, we plot the objective value with respect to the computing time and residual of linear system (81) with respect to the outer loop iteration for all four convergent algorithms. It can be seen that the objective value from IN-θR decreases fastest among all four algorithms and the residual from IN-θR gradually decreases while the residuals from other three algorithms almost remain unchanged after some iterations.

**Table 6. Comparison of IN-\( \theta_{LSQR} \), IN-\( \theta_{1e-t} \) and IN-\( \theta_{R} \) on “news20.scale”**

| Algorithm | Iteration | Mean/Max CG | Time (s) | Objective |
|-----------|-----------|-------------|----------|-----------|
| IN-θ_{LSQR} | 500 | \(~ / ~\) | 26.31 | 6.5859e+05 |
| IN-\( \theta_{1e-10} \) | 44 | 4.11/6 | 2.68 | 6.5859e+05 |
| IN-\( \theta_{1e-8} \) | 44 | 3.07/5 | 2.28 | 6.5859e+05 |
| IN-\( \theta_{1e-6} \) | 43 | 1.91/4 | 1.76 | 6.5859e+05 |
| IN-\( \theta_{1e-4} \) | 500 | 0.08/3 | 10.92 | 6.5859e+05 |
| IN-\( \theta_{1e-2} \) | 500 | 0.02/2 | 10.39 | 6.5860e+05 |
| IN-\( \theta_R \) | 46 | 0.61/2 | 1.36 | 6.5859e+05 |

**Figure 3. Objective value with respect to the computing time and residual of the linear equation (29) with respect to the outer loop iteration on “news20.scale”**

Second, we focus on another real dataset called “url_{combined}” where the dimension of \( Q \) is \( 2,396,130 \times 3,231,961 \). Because of this large dimension, we set precision \( \varepsilon = 10^{-2} \) for IN-\( \theta_{LSQR} \). The Lipschitz constant \( L = \| Q \|^2 = 1.6425e+08, \) the parameters \( \beta_1 = \beta_2 = 2/L = 1.2177e-08, \) and the parameter \( \eta = 1.4123e+05. \)
The results are presented in Table 7. It is seen that IN-$\theta_R$ is again the fastest among all algorithms while IN-$\theta_{1e-02}$ and IN-$\theta_{1e-04}$ converge unlike previous examples and IN-$\theta_{1e-02}$ is the second fastest slower than IN-$\theta_R$ by about 4%. All other algorithms converge but apparently slower than IN-$\theta_R$ and IN-$\theta_{1e-2}$. For relatively fast four algorithms, we also plot the objective value with respect to the computing time and residual of the linear equation (81) with respect to the outer loop iteration in Fig. 4. Similarly, the objective value from IN-$\theta_R$ decreases fastest among all four algorithms while the residual from IN-$\theta_R$ keeps in the range of $10^{-4}$ to $10^{-2}$.

### Table 7. Comparison of IN-$\theta_{LSQR}$, IN-$\theta_{1e-t}$ and IN-$\theta_R$ on “url_combined” dataset

| Algorithm | Iteration | Mean/Max CG | Time (s) | Objective |
|-----------|-----------|-------------|----------|-----------|
| IN-$\theta_{LSQR}$ | 29 | $\sim / \sim$ | 1794.95 | 6.1097e+05 |
| IN-$\theta_{1e-10}$ | 30 | 4.00/4 | 869.09 | 6.0859e+05 |
| IN-$\theta_{1e-8}$ | 30 | 3.13/4 | 757.60 | 6.0859e+05 |
| IN-$\theta_{1e-6}$ | 30 | 3.00/3 | 740.08 | 6.0859e+05 |
| IN-$\theta_{1e-4}$ | 29 | 2.00/2 | 592.07 | 6.1097e+05 |
| IN-$\theta_{1e-2}$ | 29 | 1.10/2 | 474.73 | 6.1074e+05 |
| IN-$\theta_R$ | 29 | 0.90/2 | 454.99 | 6.1045e+05 |

![Figure 4. Objective value with respect to the computing time and residual of the linear equation (29) with respect to the outer loop iteration on “url_combined”](image.png)

7. Conclusion. In this paper, we first provide a comprehensive convergence result of the $\theta$-scheme in the Hilbert space. Then, some useful variants are proposed in the operator form and algorithm form. Besides, three sublinear convergence rates of the $\theta$-scheme for three types of optimization problems are shown. Our numerical experiments clearly demonstrate the efficiency of the $\theta$-scheme and our proposed methods.

Acknowledgments. The authors would like to thank Prof. Xiaoming Yuan and Mr. Yongcun Song for helpful discussions. The authors are also in debt to the two anonymous referees for their constructive comments and suggestions on the earlier version of this paper.
REFERENCES

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, Fast image recovery using variable splitting and constrained optimization, IEEE Trans. Image Process., 19 (2010), 2345–2356.
[2] N. Ahmed, T. Natarajan, and K. R. Rao, Discrete cosine transform, IEEE Trans. Comput., C-23 (1974), 90–93.
[3] J. B. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés etn-cycliquement monotones, Israel J. Math., 26 (1977), 137–150.
[4] H. H. Bauschke and P. L. Combettes, et al., Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer, Cham, 2017.
[5] A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, IEEE Trans. Image Process., 18 (2009), 2419–2434.
[6] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci., 2 (2009), 183–202.
[7] J. Eckstein and W. Yao, Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results, RUTCOR Research Reports, 32 (2012).
[8] R. Glowinski, Variational Methods for the Numerical Solution of Nonlinear Elliptic Problems, CBMS-NSF Regional Conference Series in Applied Mathematics, 86, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.
[9] R. Glowinski, P. G. Ciarlet and J.-L. Lions, Numerical Methods for Fluids: Finite Element Methods for Incompressible Viscous Flow, North Holland, 2003.
[10] R. Glowinski and F. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, 9, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.
[11] R. Glowinski, S. Leung and J. L. Qian, Operator-splitting based fast sweeping methods for isotropic wave propagation in a moving fluid, SIAM J. Sci. Comput., 38 (2016), A1195–A1223.
[12] R. Glowinski and A. Marroco, Sur l’approximation, par éléments finis d’ordre un, et l’arésolution, par pénalisation-dualité d’une classe de problèmes de dirichlet non linéaires, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér., 9 (1975), 41–76.
[13] R. Glowinski, S. J. Osher and W. T. Yin, Splitting Methods in Communication, Imaging, Science, and Engineering, Scientific Computation, Springer, Cham, 2016.
[14] T. Goldstein, B. O’Donoghue, S. Setzer and R. Baraniuk, Fast alternating direction optimization methods, SIAM J. Imaging Sci., 7 (2014), 1588–1623.
[25] S. Haubruge, V. H. Nguyen and J. J. Strodiot, Convergence analysis and applications of the Glowinski–Le Tallec splitting method for finding a zero of the sum of two maximal monotone operators, *J. Optim. Theory Appl.*, 97 (1998), 645–673.

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

[27] P. Le Tallec, *Numerical Analysis of Viscoelastic Problems*, Research in Applied Mathematics, 15, Masson, Paris, 1990.

[28] P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, 16 (1979), 964–979.

[29] G. I. Marchuk, Splitting and alternating direction methods, in *Handbook of Numerical Analysis, Vol. I*, Handb. Numer. Anal., 1, North-Holland, Amsterdam, 1990, 197–462.

[30] C. C. Paige and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software*, 8 (1982), 43–71.

[31] D. W. Peaceman and H. H. Rachford Jr., The numerical solution of parabolic and elliptic differential equations, *J. Soc. Indust. Appl. Math.*, 3 (1955), 28–41.

[32] K. K. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, Inc., Boston, MA, 1990.

[33] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, 28, Princeton University Press, Princeton, NJ, 1970.

[34] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, 58 (1996), 267–288.

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

[36] P. T. Vuong and J. J. Strodiot, The Glowinski–Le Tallec splitting method revisited in the framework of equilibrium problems in Hilbert spaces, *J. Global Optim.*, 70 (2018), 477–495.

[37] H. R. Yue, Q. Z. Yang, X. F. Wang and X. M. Yuan, Implementing the alternating direction method of multipliers for big datasets: A case study of least absolute shrinkage and selection operator, *SIAM J. Sci. Comput.*, 40 (2018), A3121–A3156.

[38] C. Zalinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co., Inc., River Edge, NJ, 2002.

[39] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory. I: Projections on convex sets; II: Spectral theory*, in *Contributions to Nonlinear Functional Analysis*, Academic Press, New York, 1971, 237–424.

Received August 2019; revised October 2019.

E-mail address: 16482565@life.hkbu.edu.hk

E-mail address: liliao@hkbu.edu.hk