SIAM J. IMAGING SCIENCES
Vol. 9, No. 3, pp. 1467–1501

Convergence Study on the Symmetric Version of ADMM with Larger Step Sizes*

Bingsheng He†, Feng Ma‡, and Xiaoming Yuan§

Abstract. The alternating direction method of multipliers (ADMM), also well known as a special split Bregman algorithm in imaging, is being popularly used in many areas including the image processing field. One useful modification is the symmetric version of the original ADMM, which updates the Lagrange multiplier twice at each iteration and thus the variables are treated in a symmetric manner. The symmetric version of ADMM, however, is not necessarily convergent. It was recently found that the convergence of symmetric ADMM can be sufficiently ensured if both the step sizes for updating the Lagrange multiplier are shrunk conservatively. Despite the theoretical significance in ensuring convergence, however, smaller step sizes should be strongly avoided in practice. On the contrary, we actually have the desire of seeking larger step sizes whenever possible in order to accelerate the numerical performance. Another technique leading to numerical acceleration of ADMM is enlarging its step size by a constant originally proposed by Fortin and Glowinski. These two numerically favorable techniques are commonly but usually separately used in ADMM literature, and intuitively they seem to be incompatible in combination with the symmetric ADMM due to the conflict between the theoretical role in ensuring the convergence with smaller step sizes and the empirical necessity in accelerating numerical performance with larger step sizes. It is thus open whether the ADMM scheme in combination with these two techniques simultaneously is convergent. We answer this question affirmatively in this paper and rigorously show the convergence of the symmetric version of ADMM with step sizes that can be enlarged by Fortin and Glowinski's constant. We thus move forward to the counterintuitive understanding that shrinking both the step sizes is not necessary for the symmetric ADMM. We conduct the convergence analysis by specifying a step size domain that can ensure the convergence of symmetric ADMM; some known results in the ADMM literature turn out to be special cases of our discussion. Since the step sizes can be enlarged by constants that are problem-independent and the strategy is applicable to the general iterative scheme when the generic setting of the model is considered, our theoretical study provides an easily implementable strategy to accelerate the ADMM numerically which can be immediately applied to a variety of applications including some standard image processing tasks.

Key words. alternating direction method of multipliers, split Bregman, image reconstruction, convex programming, large step size, convergence analysis

AMS subject classifications. 90C25, 94A08

DOI. 10.1137/15M1044448

*Received by the editors October 19, 2015; accepted for publication (in revised form) June 30, 2016; published electronically September 22, 2016.

http://www.siam.org/journals/siims/9-3/M104444.html

Funding: The work of the first author was supported by NFSC grants 11471156 and 91530115. The work of the second author was supported by NSFC grant 71401176. The work of the third author was supported by the General Research Fund of the Hong Kong Research Grants Council: HKBU 12302514; and the NSFC/RGC Joint Research Scheme: N_PolyU504/14.

†Department of Mathematics, South University of Science and Technology of China, Nanshan District, Shenzhen City, China, and Department of Mathematics, Nanjing University, Nanjing, China (hebma@nju.edu.cn).

‡High-Tech Institute of Xi'an, Xi'an, 710025, China (mafengnju@gmail.com).

§Department of Mathematics, Hong Kong Baptist University, Hong Kong, China (xmyuan@hkbu.edu.hk).

1. Introduction. Our discussion starts with the canonical convex minimization model with separable structure in the generic setting:

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

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$, and $\mathcal{Y} \subset \Re^{n_2}$ are closed convex sets, and $\theta_1 : \Re^{n_1} \to \Re$ and $\theta_2 : \Re^{n_2} \to \Re$ are convex (not necessarily smooth) functions. Throughout, the solution set of (1.1) is assumed to be nonempty, and the sets $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be simple.

The augmented Lagrangian function of (1.1) can be written as

$$\mathcal{L}_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{1.2}$$

where λ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter for the linear constraints. Thus, applying directly the augmented Lagrangian method in [34, 39] to (1.1), we obtain the iterative scheme

$$\begin{cases} (x^{k+1}, y^{k+1}) = \arg\min\{\mathcal{L}_\beta(x, y, \lambda^k) \,|\, x \in \mathcal{X},\ y \in \mathcal{Y}\}, & \text{(1.3a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.3b)} \end{cases}$$

We are interested in only the case where the functions θ_1 and θ_2 may have their own properties/structures and it is worth taking advantage of them individually in algorithmic implementation. For this case, solving the (x, y)-minimization problem in (1.3) may not be efficient and we prefer splitting the (x, y)-subproblem into two easier ones, each involving only one function θ_i in its objective. This idea can be realized by the classical alternating direction method of multipliers (ADMM) (see [7, 25]). Starting with an initial iterate $(y^0, \lambda^0) \in \mathcal{Y}\times\Re^m$, the ADMM generates its sequence via the scheme

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(1.4a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, & \text{(1.4b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.4c)} \end{cases}$$

We refer to, e.g., [5, 15, 22] for some reviews on ADMM.

The split Bregman algorithm (SBA) was proposed in [27] and it has received much attention in various image processing domains; see, e.g., [1, 2, 6, 16, 28, 29, 44], to mention just a few. Now, we elaborate on the relationship between the SBA and ADMM schemes. The iterative scheme of SBA for (1.1) can be written as

$$\begin{cases} \begin{cases} \text{Perform the following two minimization problems for } N \text{ rounds:} \\ \quad x^{k+1} \leftarrow \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \quad y^{k+1} \leftarrow \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \end{cases} & \text{(1.5a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & \text{(1.5b)} \end{cases}$$

where $N > 0$ is an integer. Clearly, if N is chosen as 1 in (1.5), i.e., the (x, y)-subproblem in (1.3a) is approximated by one Gauss–Seidel pass of the block-minimization approach, then the SBA (1.5) reduces to the ADMM scheme (1.4). It is interesting to notice the comment in [27]: for many applications optimal efficiency is obtained when only one iteration of the inner loop is performed (i.e., $N = 1$ in the above algorithm). We refer to, e.g., [16, 41, 42], for more elaborations on the relationship between the SBA and ADMM schemes.

As reported in [21], Fortin and Glowinski suggested, in [18, 19], attaching a relaxation factor to the Lagrange-multiplier-updating step in (1.4). This results in the scheme

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(1.6a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, & \text{(1.6b)} \\ \lambda^{k+1} = \lambda^k - s\beta(Ax^{k+1} + By^{k+1} - b), & \text{(1.6c)} \end{cases}$$

where the parameter s can be chosen in the interval $\left(0, \frac{1+\sqrt{5}}{2}\right)$ and thus it becomes possible to enlarge the step size for updating the Lagrange multiplier. An advantage of this larger step size in (1.6) is that it can numerically lead to faster convergence empirically; see some numerical results in [21] and more in, e.g., [8, 32]. Moreover, the parameter s for convergence acceleration is just a constant and it can be chosen without additional computation. We see that the scheme (1.6) differs from the original ADMM scheme (1.4) only in the fact that the step size for updating the Lagrange multiplier can be larger than 1. But, as commented in [15], technically they are "actually two distinct families of ADMM algorithms, one derived from the operator-splitting framework and the other derived from Lagrangian splitting." Thus, despite the similarity in notation, the ADMM scheme (1.6) with Fortin and Glowinski's larger step size and the original ADMM scheme (1.4) are completely different in nature.

As analyzed in [13, 20], the ADMM scheme (1.4) is an application of the Douglas–Rachford splitting method (DRSM) in [11, 37] to the dual of (1.1). If the Peaceman–Rachford splitting method (PRSM), which was originally proposed in [37, 38] and is equally important as the DRSM in PDE literature, is applied to the dual of (1.1), it was shown in [24] that we can obtain the following scheme:

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(1.7a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(1.7b)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, & \text{(1.7c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.7d)} \end{cases}$$

In the context of (1.1), the applications of PRSM and DRSM (i.e., (1.7) and (1.4), respectively) have the only difference that the Lagrange multiplier is updated twice at each iteration in (1.7). In [24], it was commented that the scheme (1.7) "is always faster than DRSM whenever it is convergent"; its numerical efficiency can be found in [3, 23]. Alternatively, the scheme (1.7) can be regarded as a symmetric version of the ADMM scheme (1.4) in the sense that the variables x and y are treated equally, each of which is followed consequently by an update of the Lagrange multiplier. Analytically, however, these two schemes are of significant difference because of their different roots respectively in DRSM and PRSM. One more explanation can be found in [31], showing that the sequence generated by the symmetric ADMM (1.7) is not necessarily strictly contractive with respect to the solution set of (1.1) while this property can be ensured by the sequence generated by the ADMM (1.4). Counterexamples showing the divergence of (1.7) can also be found in [10, 12]. Once again, despite the similarity in notation (only when the specific model (1.1) is considered), the symmetric ADMM (1.7) and the original ADMM scheme (1.4) are completely different in nature.

To summarize, the ADMM scheme (1.6) with larger step sizes and the symmetric ADMM (1.7) with an equal treatment on both variables are of different natures from the original ADMM scheme (ADMM); but they are two commonly used techniques to accelerate the original ADMM scheme (1.4). These two numerically favorable techniques, however, are usually used separately; thus it is natural to ask if we can combine them together and thus propose an symmetric version of ADMM but with larger step sizes. Before answering this question, let us briefly recall the work [31], which seems to provide a puzzling clue to this question. In [31], to overcome the difficulty of the lack of strict contraction, we suggested updating the Lagrange multiplier in (1.7) more conservatively and obtained the symmetric ADMM scheme with smaller step sizes:

$$\left\{\begin{aligned}
&x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, && \text{(1.8a)}\\
&\lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), && \text{(1.8b)}\\
&y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, && \text{(1.8c)}\\
&\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), && \text{(1.8d)}
\end{aligned}\right.$$

where the parameter $\alpha \in (0, 1)$ is for shrinking the step sizes in (1.7). It was shown in [31] that the sequence generated by (1.8) is strictly contractive with respect to the solution set of (1.1). Thus, in [31] we also called the symmetric ADMM (1.8) the strictly contractive PRSM. It is worthwhile to mention that the restriction $\alpha \in (0, 1)$ makes the update of the Lagrange multiplier more conservative with smaller step sizes, but it plays a crucial theoretical role in ensuring the strict contraction for the sequence generated by (1.8) and hence sufficiently ensuring the convergence of the symmetric ADMM (1.8). We also refer the reader to [35, 36] for proximal versions of (1.8) which treat the subproblems more sophisticatedly, and some applications to image processing are tested therein. Despite their significant theoretical role in (1.8), smaller step sizes should be strongly avoided in practice, and on the contrary, we have the desire to seek larger step sizes whenever possible in order to accelerate the numerical performance. (Indeed it was recommended in [31] to take larger values close to 1 for α to lead to better numerical performance.)

Hence, the theory in [31] for ensuring convergence of the symmetric ADMM (1.7) with smaller step sizes seems to show that these two techniques (1.6) and (1.7) are incompatible in

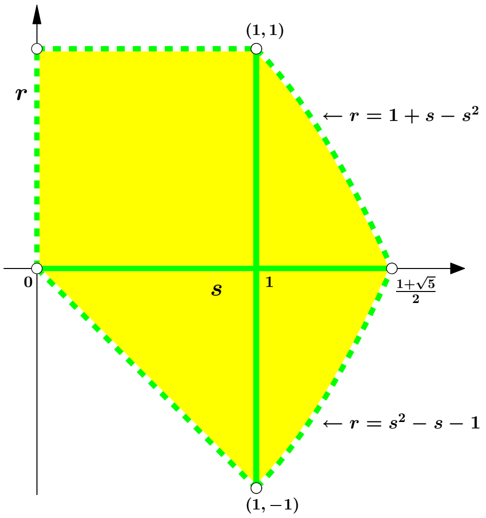

Figure 1. *Step size domain $\mathcal{D}$ for the symmetric ADMM* (1.9).

combination with the original ADMM scheme (1.4) due to the conflict between the theoretical role in ensuring the convergence with smaller step sizes (e.g., (1.8)) and the empirical necessity in accelerating numerical performance with larger step sizes. Indeed, given the already known divergence of (1.7) and the fact that ensuring the convergence of an algorithm with larger step sizes is usually more demanding to be analyzed, it is easy to have the intuition that the combination of (1.7) with larger step sizes as in (1.6) is more likely to be divergent.

Our main purpose in this paper is, counterintuitively, to show the rigorous convergence of the symmetric version of ADMM (1.7) with step sizes that could be enlarged by Fortin and Glowinski's constant in [18, 19, 21]. We conduct our convergence analysis by specifying a domain for the constants to be attached to the step sizes in the symmetric ADMM (1.7). To emphasize that we can take different step sizes for updating the Lagrange multiplier, we assign two different constants to different step sizes in (1.7) and propose the following version of the symmetric ADMM (1.8):

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(1.9a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & \text{(1.9b)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, & \text{(1.9c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + By^{k+1} - b), & \text{(1.9d)} \end{array}\right.$$

in which r and s are independent constants that are restricted into the domain

$$\mathcal{D} = \left\{(s, r) \,|\, s \in \left(0, \tfrac{1+\sqrt{5}}{2}\right),\ r \in (-1, 1)\ \&\ r + s > 0,\ |r| < 1 + s - s^2\right\}. \tag{1.10}$$

In Figure 1, we show the domain $\mathcal{D}$ for the convenience of discussion, and various ADMM-like algorithms can be retrieved by choosing different values for r and s in the domain. Indeed, it is easy to see that our discussion includes some known ADMM-based schemes as special cases. For example, the original ADMM (1.4) corresponds to the point $(s = 1, r = 0)$ in Figure 1, the ADMM with Fortin and Glowinski's larger step size (1.6) is the case where $s \in (0, \frac{1+\sqrt{5}}{2})$ and $r = 0$, the generalized ADMM scheme proposed in [13] is the case where $s = 1$ and $r \in (-1, 1)$ (see Remark 5.8 for elaboration), and the symmetric ADMM in [31] corresponds to the region where $s \in (0, 1)$ and $r \in (0, 1)$. Therefore, with the restriction of r and s in $\mathcal{D}$, we provide a comprehensive study on choosing the step sizes for the symmetric ADMM (1.9). Indeed, it turns out that the proof is more demanding than those for the mentioned special cases, mainly because one step size can be enlarged to the interval $(0, \frac{1+\sqrt{5}}{2})$.

The rest of this paper is organized as follows. In section 2, we summarize some facts that are useful for further analysis. In particular, the variational inequality characterization of the model (1.1) is presented. Since the proof of convergence for (1.9) is highly nontrivial, we prepare for the main convergence analysis step by step in sections 3–5. Then, the convergence analysis is conducted in section 6. We report some preliminary numerical results in section 7 to support our theoretical assertions. Some conclusions are made in section 8.

Finally, we refer to [26] for further recent developments on the ADMM and SBA, especially their applications to image processing and others in science and engineering.

2. Preliminaries.

2.1. Optimality condition in terms of variational inequality.

In this section, we summarize some preliminaries that will be used in later analysis. First, we show how to represent the optimality condition of the model (1.1) in the variational inequality context, which is the basis of the convergence analysis to be presented. This technique has been used in, e.g., [31, 33].

Let the Lagrangian function of the problem (1.1) be

$$L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b), \tag{2.1}$$

defined on $\mathcal{X} \times \mathcal{Y} \times \Re^m$. In (2.1), (x, y) and λ are primal and dual variables, respectively. We call $\big((x^*, y^*), \lambda^*\big) \in \mathcal{X} \times \mathcal{Y} \times \Re^m$ a saddle point of $L(x, y, \lambda)$ if the following inequalities are satisfied:

$$L_{\lambda\in\Re^m}(x^*, y^*, \lambda) \le L(x^*, y^*, \lambda^*) \le L_{x\in\mathcal{X}, y\in\mathcal{Y}}(x, y, \lambda^*).$$

Obviously, a saddle point $\big((x^*, y^*), \lambda^*\big)$ can be characterized by the system

$$\left\{\begin{array}{lcl} x^* & = & \arg\min\{L(x, y^*, \lambda^*) | x \in \mathcal{X}\}, \\ y^* & = & \arg\min\{L(x^*, y, \lambda^*) | y \in \mathcal{Y}\}, \\ \lambda^* & = & \arg\max\{L(x^*, y^*, \lambda) | \lambda \in \Re^m\}, \end{array}\right. \tag{2.2}$$

which can be rewritten as

$$\left\{\begin{array}{ll} x^* \in \mathcal{X}, & L(x, y^*, \lambda^*) - L(x^*, y^*, \lambda^*) \ge 0 \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & L(x^*, y, \lambda^*) - L(x^*, y^*, \lambda^*) \ge 0 \quad \forall\, y \in \mathcal{Y}, \\ \lambda^* \in \Re^m, & L(x^*, y^*, \lambda^*) - L(x^*, y^*, \lambda) \ge 0 \quad \forall\, \lambda \in \Re^m. \end{array}\right.$$

Below we summarize how to characterize the optimality condition of an optimization model by a variational inequality. The proof is obvious and is thus omitted.

Proposition 2.1. *Let $\mathcal{X} \subset \Re^n$ be a closed convex set and let $\theta(x) : \Re^n \to \Re$ and $f(x) : \Re^n \to \Re$ be convex functions. In addition, $f(x)$ is differentiable. We assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then,*

$$x^* = \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{2.3a}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0 \quad \forall\, x \in \mathcal{X}. \tag{2.3b}$$

Hence, using (2.1) and (2.3), we can rewrite the system (2.2) as

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0 \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0 \quad \forall\, y \in \mathcal{Y}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0 \quad \forall\, \lambda \in \Re^m. \end{cases} \tag{2.4}$$

In other words, a saddle point $\big((x^*, y^*), \lambda^*\big)$ of the Lagrangian function (2.1) can be characterized by a solution point of the following variational inequality:

$$\mathrm{VI}(\Omega, F, \theta) \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0 \quad \forall\, w \in \Omega, \tag{2.5a}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{2.5b}$$

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \text{and} \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m.$$

We denote by Ω^* the solution set of $\mathrm{VI}(\Omega, F, \theta)$.

2.2. Characterization of a solution point of (1.1).

The following theorem provides us a criterion for judging if an iterate generated by the scheme (1.9) is an approximate solution point of (1.1) with sufficient accuracy.

Theorem 2.2. *For $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ generated by the symmetric ADMM (1.9) from a given iterate (y^k, λ^k), if*

$$B(y^k - y^{k+1}) = 0 \qquad \text{and} \qquad Ax^{k+1} + By^{k+1} - b = 0, \tag{2.6}$$

then $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution point of the variational inequality (2.5).

Proof. Because of (2.4), we need only show that

$$\begin{cases} x^{k+1} \in \mathcal{X}, & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1}\} \geq 0 \quad \forall\, x \in \mathcal{X}, & (2.7a) \\ y^{k+1} \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0 \quad \forall\, y \in \mathcal{Y}, & (2.7b) \\ \lambda^{k+1} \in \Re^m, & (\lambda - \lambda^{k+1})^T(Ax^{k+1} + By^{k+1} - b) \geq 0 \quad \forall\, \lambda \in \Re^m. & (2.7c) \end{cases}$$

First, it follows from (2.6) that

$$Ax^{k+1} + By^k - b = 0 \qquad \text{and} \qquad Ax^{k+1} + By^{k+1} - b = 0. \tag{2.8}$$

Consequently, together with (1.9b) and (1.9d), we get

$$\lambda^{k+\frac{1}{2}} = \lambda^k \qquad \text{and} \qquad \lambda^{k+1} = \lambda^k. \tag{2.9}$$

On the other hand, using (2.3), the optimality conditions of the subproblems (1.9a) and (1.9c) are

(2.10a)

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

and

(2.10b)

$$y^{k+1} \in \mathcal{Y},\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} \geq 0\ \forall\, y \in \mathcal{Y},$$

respectively. Substituting (2.8) in (2.10) yields (2.7a) and (2.7b). Finally, notice that (2.7c) can be specified as

$$Ax^{k+1} + By^{k+1} - b = 0.$$

The proof is complete. ■

Remark 2.3. According to Theorem 2.2, it is obvious that we can use

$$\max\{\|B(y^k - y^{k+1})\|^2, \|Ax^{k+1} + By^{k+1} - b\|^2\} \leq \epsilon$$

as a stopping criterion to implement the symmetric ADMM (1.9), in which $\epsilon > 0$ is the tolerance specified by the user.

2.3. Some notation. Like the original ADMM scheme (1.4), in (1.9) the variable x also plays an intermediate role in the sense that x^k is not involved in the iteration. Thus, as [5], we still call x an *intermediate variable* and (y, λ) *essential variables* because they are essentially needed in the iteration. Correspondingly, for $w = (x, y, \lambda)$ and $w^k = (x^k, y^k, \lambda^k)$ generated by (1.9), we use the notation

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix} \quad \text{and} \quad v^k = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix}$$

to represent the essential variables in w and w^k, respectively. Moreover, we use $\mathcal{V}^*$ to denote the set of v^* for all subvectors of w^* in Ω^*.

3. A prediction-correction interpretation.

In this section, we show a prediction-correction interpretation to the symmetric ADMM (1.9). Note that this is only for the convenience of algebraically presenting the convergence proof with compact notation.

First, for the iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ generated by the symmetric ADMM (1.9), we define an auxiliary vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as

$$\tilde{x}^k = x^{k+1}, \qquad \tilde{y}^k = y^{k+1}, \tag{3.1a}$$

and

$$\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b). \tag{3.1b}$$

Below we show the discrepancy between the auxiliary vector $\tilde{w}^k$ and a solution point of VI(Ω, F, θ).

Lemma 3.1. *For given $v^k = (y^k, \lambda^k)$, let w^{k+1} be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). Then, we have*

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) \;\; \forall w \in \Omega, \tag{3.2a}$$

where

$$Q = \begin{pmatrix} \beta B^T B & -rB^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.2b}$$

Proof. Using the notation defined in (3.1), we can rewrite $\lambda^{k+\frac{1}{2}}$ in (1.9b) as

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k). \tag{3.3}$$

Notice that the objective functions of the x- and y-subproblems in (1.9) are

$$\mathcal{L}_\beta(x, y^k, \lambda^k) = \theta_1(x) + \theta_2(y^k) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2, \tag{3.4a}$$

and

$$\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) = \theta_1(x^{k+1}) + \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2, \tag{3.4b}$$

respectively. According to (2.3), the optimality condition of the x-subproblem of (1.9a) is

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

and it can be written as (by using the auxiliary vector $\tilde{w}^k$ defined in (3.1))

$$\tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0 \;\; \forall\, x \in \mathcal{X}. \tag{3.5a}$$

Ignoring some constant terms in the objective function of the y-subproblem (1.9c) and using (3.3), we have

$$\tilde{y}^k = \arg\min\left\{\theta_2(y) - \big(\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)\big)^T By + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \,|\, y \in \mathcal{Y}\right\}.$$

Consequently, using (2.3), we obtain

$$\tilde{y}^k \in \mathcal{Y}, \; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\big(\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)\big) + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0 \; \forall y \in \mathcal{Y}.$$

Now, we treat the $\{\cdot\}$ term in the last inequality. Using $\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k)$ (see (3.1b)), we obtain

$$\begin{aligned}
&-B^T\big(\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)\big) + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\\
&= -B^T(\tilde{\lambda}^k + (r-1)(\tilde{\lambda}^k - \lambda^k)) + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T(A\tilde{x}^k + By^k - b)\\
&= -B^T\tilde{\lambda}^k - (r-1)B^T(\tilde{\lambda}^k - \lambda^k) + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k)\\
&= -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - rB^T(\tilde{\lambda}^k - \lambda^k).
\end{aligned}$$

Thus, the optimality condition of the y-subproblem can be written as

(3.5b)
$$\tilde{y}^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - rB^T(\tilde{\lambda}^k - \lambda^k)\big\} \geq 0 \ \ \forall y \in \mathcal{Y}.$$

According to the definition of $\tilde{w}^k$ in (3.1), we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0,$$

and it can be written as

(3.5c) $$\tilde{\lambda}^k \in \Re^m, \ \ (\lambda - \tilde{\lambda}^k)^T\big\{(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\big\} \geq 0 \ \ \forall \lambda \in \Re^m.$$

Combining (3.5a), (3.5b), and (3.5c), and using the notation of (2.5), we prove the assertion of this lemma. ∎

Lemma 3.2. *For given v^k, let w^{k+1} be generated by (1.9) and $\tilde{w}^k$ be defined by (3.1). Then we have*

(3.6a) $$v^{k+1} = v^k - M(v^k - \tilde{v}^k),$$

where

(3.6b) $$M = \begin{pmatrix} I & 0 \\ -s\beta B & (r+s)I_m \end{pmatrix}.$$

Proof. It follows from (1.9) that

$$\begin{aligned}
\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \big[-s\beta B(y^k - \tilde{y}^k) + s\beta(Ax^{k+1} + By^k - b)\big]\\
&= \lambda^k - r(\lambda^k - \tilde{\lambda}^k) - \big[-s\beta B(y^k - \tilde{y}^k) + s(\lambda^k - \tilde{\lambda}^k)\big]\\
&= \lambda^k - \big[-s\beta B(y^k - \tilde{y}^k) + (r+s)(\lambda^k - \tilde{\lambda}^k)\big].
\end{aligned}$$

Thus, together with $y^{k+1} = \tilde{y}^k$, we have the following useful relationship:

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -s\beta B & (r+s)I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

The proof is complete. ∎

Therefore, based on the previous two lemmas, the scheme (1.9) can be interpreted as a prediction-correction procedure which consists of the prediction step (3.2) and the correction step (3.6). We thus also frequently call $\tilde{w}^k$ and w^{k+1} the predictor and corrector, respectively, in our proof.

4. Main result.

As mentioned, our main purpose is to prove the convergence for the symmetric ADMM (1.9) with the parameters r and s restricted into the domain $\mathcal{D}$ defined in (1.10). As we shall show, the technique of our analysis basically follows the convergence analysis framework of contraction methods; see, e.g., in [4]. We thus need to investigate the strict contraction property of the sequence $\{\|v^k - v^*\|_H^2\}$, which is not obvious for some parts of the domain $\mathcal{D}$ under discussion. In this section, we present a main theorem that is the basis for the convergence proof; then the detailed proof of this main theorem will be provided in the next section.

4.1. Contraction of the sequence $\{\|v^k - v^*\|_H^2\}$.

First, we investigate the contraction of the sequence $\{\|v^k - v^*\|_H^2\}$ that plays a pivotal role in the convergence analysis. We know from the monotonicity of F that

$$(w - \tilde{w}^k)^T F(w) \geq (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting this inequality into (3.2a), we get

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) \quad \forall w \in \Omega, \tag{4.1}$$

We notice that for the matrices Q defined in (3.2b) and M defined in (3.6b), if there is a positive definite matrix H such that $Q = HM$, then using (3.6a), we can rewrite the right-hand side of (4.1) as

$$(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^T H(v^k - v^{k+1}). \tag{4.2}$$

For this purpose, we define

$$H = QM^{-1} \tag{4.3}$$

and prove a simple fact for the matrix H in the following lemma.

Lemma 4.1. *The matrix H defined in (4.3) is positive definite for any $(s, r) \in \mathcal{D}$ when the matrix B in (1.1) is full column rank.*

Proof. For the matrix M given in (3.6b), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \frac{s}{r+s}\beta B & \frac{1}{r+s} I_m \end{pmatrix}.$$

Thus, it follows from (4.3) and (3.2b) that

$$\begin{aligned} H = QM^{-1} &= \begin{pmatrix} \beta B^T B & -rB^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{s}{r+s}\beta B & \frac{1}{r+s} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1 - \frac{rs}{r+s})\beta B^T B & -\frac{r}{r+s} B^T \\ -\frac{r}{r+s} B & \frac{1}{(r+s)\beta} I_m \end{pmatrix}. \end{aligned} \tag{4.4}$$

For any $(s, r) \in \mathcal{D}$, we have $r + s > 0$. From (4.4), we know that

$$
\begin{aligned}
H &= \begin{pmatrix} (1 - \frac{rs}{r+s})\beta B^T B & -\frac{r}{r+s} B^T \\ -\frac{r}{r+s} B & \frac{1}{(r+s)\beta} I_m \end{pmatrix} \\
&= \frac{1}{r+s} \begin{pmatrix} B^T & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta(r+s-rs)I & -rI \\ -rI & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix}.
\end{aligned} \tag{4.5}
$$

Because the matrix B in (1.1) is assumed to be full column rank, to check the positive definiteness of H, we need only to show that the matrix

$$
\begin{pmatrix} \beta(r+s-rs) & -r \\ -r & \frac{1}{\beta} \end{pmatrix}
$$

is positive definite. Indeed, for any $r \in (-1, 1)$, $s > 0$, and $s + r > 0$, we have

$$
r + s - rs = \left\{ \begin{array}{c} \text{for} \quad r \in [0,1) \;\text{ and }\; s > 0 \\ \text{for} \quad r \in (-1,0), s > 0 \;\text{ and }\; s + r > 0 \end{array} \right\} = \left\{ \begin{array}{c} r + s(1-r) \\ (r+s) - rs \end{array} \right\} > 0
$$

and

$$
\det \begin{pmatrix} \beta(r+s-rs) & -r \\ -r & \frac{1}{\beta} \end{pmatrix} = (1-r)(r+s) > 0.
$$

The positive definiteness of H follows immediately. ∎

For the case where H is positive definite, we have $HM = Q$. Thus, substituting (4.2) into (4.1), we obtain

$$
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}) \quad \forall w \in \Omega. \tag{4.6}
$$

Applying the identity

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

to the right-hand side in (4.6) with

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

we obtain

$$
(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \tag{4.7}
$$

The following theorem is useful for estimating the convergence rate for the sequence generated by (1.9).

Theorem 4.2. *For the sequence $\{w^k\}$ generated by the symmetric ADMM (1.9), we have*

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(w)\\ &\geq\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2\quad\forall w\in\Omega,\end{aligned}\tag{4.8}$$

where H is defined in (4.3) and

$$G=Q^T+Q-M^THM.\tag{4.9}$$

Proof. Substituting (4.7) into the right-hand side of (4.6), we obtain

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(w)\\ &\geq\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big)\quad\forall w\in\Omega.\end{aligned}\tag{4.10}$$

For the last term of the right-hand side of (4.10), we have

$$\begin{aligned}&\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\\ &\quad=\quad\|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-(v^k-v^{k+1})\|_H^2\\ &\quad\overset{(3.6a)}{=}\quad\|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\\ &\quad=\quad2(v^k-\tilde{v}^k)^THM(v^k-\tilde{v}^k)-(v^k-\tilde{v}^k)^TM^THM(v^k-\tilde{v}^k)\\ &\quad\overset{(4.3)}{=}\quad(v^k-\tilde{v}^k)^T(Q^T+Q-M^THM)(v^k-\tilde{v}^k).\end{aligned}\tag{4.11}$$

Substituting this equation into (4.10) and using the definition of the matrix G in (4.9), the assertion of this theorem is proved. ∎

The next theorem clearly shows the difficulty of proving the convergence for (1.9). The significant difference in convergence proof between the general symmetric ADMM scheme (1.9) and the special symmetric ADMM scheme (1.8) with shrunken step sizes is also reflected in this theorem.

Theorem 4.3. *For the sequence $\{w^k\}$ generated by the symmetric ADMM (1.9), we have*

$$\|v^{k+1}-v^*\|_H^2\leq\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2\quad\forall v^*\in\mathcal{V}^*.\tag{4.12}$$

Proof. Setting $w=w^*$ in (4.8), we get

$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\geq\|v^k-\tilde{v}^k\|_G^2+2\{\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(w^*)\}.\tag{4.13}$$

Since $\tilde{w}^k\in\Omega$, using the optimality of w^* (see (2.5a)), we have

$$\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(w^*)\geq0$$

and thus

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

The assertion (4.12) follows directly. ∎

Obviously, if the matrix G defined in (4.9) is guaranteed to be positive definite, then the inequality (4.12) indicates that the sequence $\{v^k\}$ is strictly contractive with respect to the solution set $\mathcal{V}^*$ and the convergence of the sequence generated by (1.9) can be easily established. Now, let us investigate the positive definiteness for the matrix G. Since $HM = Q$ (see (4.3)), we have $M^T HM = M^T Q$. Note that

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

Using (3.2b) and the above equation, we have

$$\begin{aligned} G &= (Q^T + Q) - M^T HM \\ &= \begin{pmatrix} 2\beta B^T B & -(1+r)B^T \\ -(1+r)B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} (1+s)\beta B^T B & -(r+s)B^T \\ -(r+s)B & \frac{1}{\beta}(r+s)I_m \end{pmatrix} \\ &= \begin{pmatrix} (1-s)\beta B^T B & -(1-s)B^T \\ -(1-s)B & \frac{1}{\beta}(2-(r+s))I_m \end{pmatrix}. \end{aligned} \tag{4.14}$$

Note that

$$G = \begin{pmatrix} B^T & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta(1-s)I & -(1-s)I \\ -(1-s)I & \frac{1}{\beta}\big(2-(r+s)\big)I \end{pmatrix} \begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix}. \tag{4.15}$$

Thus, when the matrix B in (1.1) is assumed to be full column rank, the matrix G is positive definite if and only if

$$G_0 = \begin{pmatrix} \beta(1-s)I & -(1-s)I \\ -(1-s)I & \frac{1}{\beta}\big(2-(r+s)\big)I \end{pmatrix}$$

is positive definite. Indeed, under the assumption $r \in (-1, 1)$, to ensure

$$\beta(1-s) > 0 \qquad \text{and} \qquad \det \begin{pmatrix} \beta(1-s) & -(1-s) \\ -(1-s) & \frac{1}{\beta}(2-r-s) \end{pmatrix} = (1-s)(1-r) > 0,$$

the parameter s must be in $(0, 1)$. Obviously, when $r = s \in (0, 1)$, the positive definiteness of G is guaranteed; thus the symmetric ADMM (1.8) with shrunken step sizes is convergent, as proved in [31]. However, for $(s, r) \in \mathcal{D}$ defined in (1.10) where s may be greater than 1, G may not be positive definite. This is the main difficulty of proving the convergence for the symmetric ADMM (1.9).

4.2. Main theorem. Now we present the main theorem that enables us to establish the convergence for the symmetric ADMM (1.9) with the parameters r and s restricted into the domain $\mathcal{D}$ defined in (1.10). As just mentioned, depending on the domain of (s, r), the matrix G in (4.9) may not be positive definite, and thus the inequality (4.12) in Theorem 4.3 does not necessarily imply the strict contraction for the sequence $\{\|v^k - v^*\|_H^2\}$. This is also why we need to further discuss the term $\|v^k-\tilde{v}^k\|_G^2$ in different subdomains of (s,r). More specifically, we should split this discussion into five scenarios:

$$(4.16)\qquad \begin{cases} \mathcal{D}_1 &= \big\{(s,r)\,|\, s\in(0,1),\ r\in(-1,1),\ r+s>0\big\},\\ \mathcal{D}_2 &= \big\{(s,r)\,|\, s=1,\ r\in(-1,1)\big\},\\ \mathcal{D}_3 &= \big\{(s,r)\,|\, s\in\big(1,\frac{1+\sqrt{5}}{2}\big),\ r=0\big\},\\ \mathcal{D}_4 &= \big\{(s,r)\,|\, s\in\big(1,\frac{1+\sqrt{5}}{2}\big),\ r\in(0,1)\ \&\ r<1+s-s^2\big\},\\ \mathcal{D}_5 &= \big\{(s,r)\,|\, s\in\big(1,\frac{1+\sqrt{5}}{2}\big),\ r\in(-1,0)\ \&\ -r<1+s-s^2\big\}. \end{cases}$$

These five subdomains are displayed separately in Figure 2. Figure 1 is redisplayed in Figure 3, in which the composition of these subdomains defined in (4.16) is clearly shown. It is clear (see Figure 1) that

$$\mathcal{D}=\bigcup_{k=1}^{5}\mathcal{D}_k \qquad \text{and} \qquad \mathcal{D}_i\bigcap\mathcal{D}_j=\emptyset\ \ \forall\, i,j\in\{1,2,3,4,5\},\ i\neq j.$$

In the following, we summarize the main theorem, which gives us a unified presentation of how to bound the term $\|v^k-\tilde{v}^k\|_G^2$ in different subdomains of (s,r). Since the proof is highly nontrivial, the detailed proof will be given in the next section.

Theorem 4.4. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). We have the following:*

1. *For arbitrarily fixed $(s,r)\in\mathcal{D}_1\cup\mathcal{D}_2$, there exist constants C_1, $C_2>0$, such that*

$$(4.17a)\qquad \|v^k-\tilde{v}^k\|_G^2\geq C_1\beta\|B(y^k-y^{k+1})\|^2+C_2\beta\|Ax^{k+1}+By^{k+1}-b\|^2.$$

2. *For arbitrarily fixed $(s,r)\in\mathcal{D}_3\cup\mathcal{D}_4\cup\mathcal{D}_5$, there exist constants C_0, C_1, $C_2>0$, such that*

$$\begin{aligned}\|v^k-\tilde{v}^k\|_G^2 &\geq C_0\beta\big(\|Ax^{k+1}+By^{k+1}-b\|^2-\|Ax^k+By^k-b\|^2\big)\\ &\quad +C_1\beta\|B(y^k-y^{k+1})\|^2+C_2\beta\|Ax^{k+1}+By^{k+1}-b\|^2.\end{aligned}\qquad (4.17b)$$

5. Proof of Theorem 4.4.

In this section, we prove Theorem 4.4.

5.1. Further investigation of $\|v^k-\tilde{v}^k\|_G^2$.

Let us further investigate the term $\|v^k-\tilde{v}^k\|_G^2$ and then show the assertions in Theorem 4.4.

Lemma 5.1. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). Then we have*

$$\begin{aligned}\|v^k-\tilde{v}^k\|_G^2 &=(1-r)\beta\|B(y^k-y^{k+1})\|^2+(2-r-s)\beta\|Ax^{k+1}+By^{k+1}-b\|^2\\ &\quad +2(1-r)\beta(Ax^{k+1}+By^{k+1}-b)^TB(y^k-y^{k+1}).\end{aligned}\qquad (5.1)$$

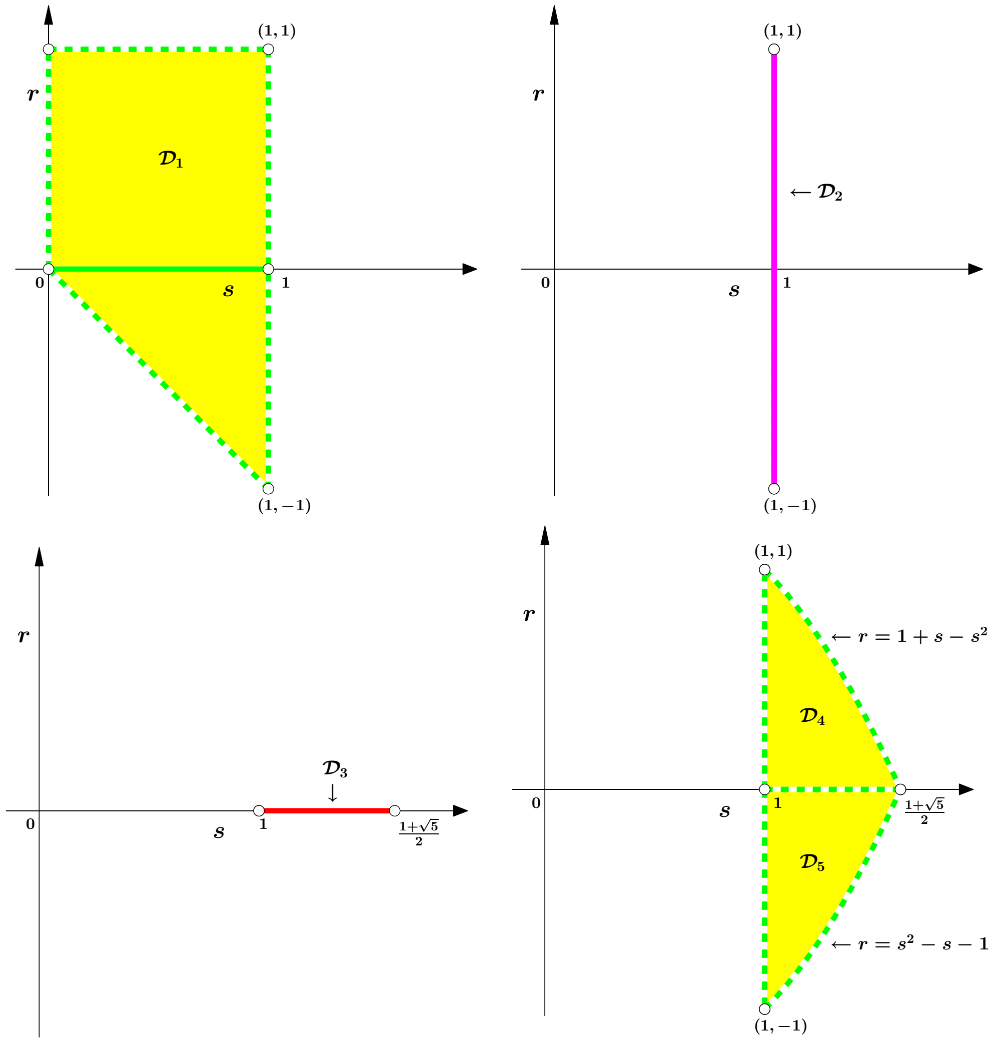

Figure 2. *Subdomains* $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$, $\mathcal{D}_4$, *and* $\mathcal{D}_5$.

Proof. Because $G = \begin{pmatrix} (1-s)\beta B^T B & -(1-s)B^T \\ -(1-s)B & \frac{2-(r+s)}{\beta} I \end{pmatrix}$ (see (4.14)), $v = \begin{pmatrix} y \\ \lambda \end{pmatrix}$ and $\tilde{y}^k = y^{k+1}$, we have

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 = (1-s)\beta\|B(y^k - y^{k+1})\|^2 - 2(1-s)(\lambda^k - \tilde{\lambda}^k)^T B(y^k - y^{k+1}) \\ + \frac{2-(r+s)}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2. \end{aligned} \tag{5.2}$$

Notice that (see (3.1))

$$\lambda^k - \tilde{\lambda}^k = \beta(Ax^{k+1} + By^k - b) = \beta\big\{(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\big\}.$$

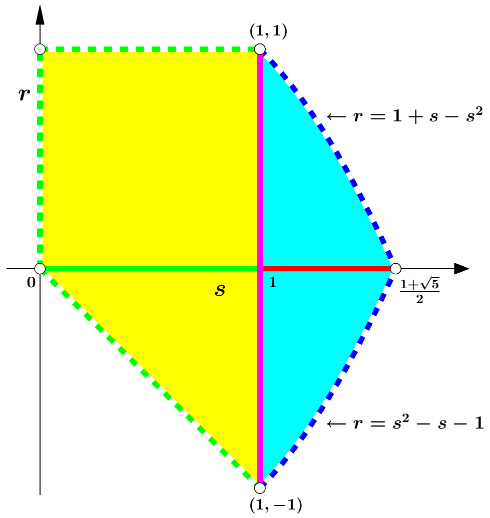

Figure 3. $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$.

Thus, we have

$$
\begin{aligned}
&(\lambda^k - \tilde{\lambda}^k)^T B(y^k - y^{k+1}) \\
&= \beta\big((Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\big)^T B(y^k - y^{k+1}) \\
&= \beta(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) + \beta\|B(y^k - y^{k+1})\|^2
\end{aligned} \tag{5.3}
$$

and

$$
\begin{aligned}
\|\lambda^k - \tilde{\lambda}^k\|^2 &= \beta^2\|(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\|^2 \\
&= \beta^2\|Ax^{k+1} + By^{k+1} - b\|^2 + 2\beta^2(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \\
&\quad + \beta^2\|B(y^k - y^{k+1})\|^2.
\end{aligned} \tag{5.4}
$$

Substituting (5.3) and (5.4) into the right-hand side of (5.2), we obtain (5.1) immediately. ■

In the following, we treat the crossing term in the right-hand side of (5.1), namely,

$$(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}),$$

and find a lower bound for it.

Lemma 5.2. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9). Then we have*

$$
\begin{aligned}
&(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \\
&\geq \frac{1-s}{1+r}(Ax^k + By^k - b)^T B(y^k - y^{k+1}) - \frac{r}{1+r}\|B(y^k - y^{k+1})\|^2.
\end{aligned} \tag{5.5}
$$

Proof. By ignoring the constant terms, the y-subproblem in (1.9) can be written as

$$\begin{aligned} y^{k+1} &= \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\} \\ &= \arg\min\left\{\theta_1(x^{k+1}) + \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\right\} \\ &= \arg\min\left\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\right\}. \end{aligned}$$

Thus, according to (3.4b), the optimality condition of the y-subproblem (1.9c) is

$$y^{k+1} \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} \geq 0 \ \ \forall y \in \mathcal{Y}. \tag{5.6}$$

Analogously, for the previous iteration, we have

$$y^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^{k-\frac{1}{2}} + \beta B^T(Ax^k + By^k - b)\} \geq 0 \ \ \forall y \in \mathcal{Y}. \tag{5.7}$$

Setting $y = y^k$ and $y = y^{k+1}$ in (5.6) and (5.7), respectively, and then adding them, we get

$$(y^k - y^{k+1})^T B^T\{(\lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}}) - \beta(Ax^k + By^k - b) + \beta(Ax^{k+1} + By^{k+1} - b)\} \geq 0. \tag{5.8}$$

Note that in the kth iteration (see (1.9b)), we have

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \tag{5.9}$$

and in the previous iteration (see (1.9d)),

$$\lambda^k = \lambda^{k-\frac{1}{2}} - s\beta(Ax^k + By^k - b). \tag{5.10}$$

It follows from (5.10) and (5.9) that

$$\begin{aligned} \lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} &= r\beta(Ax^{k+1} + By^k - b) + s\beta(Ax^k + By^k - b) \\ &= r\beta(Ax^{k+1} + By^{k+1} - b) + r\beta B(y^k - y^{k+1}) + s\beta(Ax^k + By^k - b). \end{aligned} \tag{5.11}$$

Substituting (5.11) into (5.8) and with a simple manipulation, we have

$$\begin{aligned} (y^k - y^{k+1})^T B^T \Big\{&(1 + r)\beta(Ax^{k+1} + By^{k+1} - b) \\ &-(1 - s)\beta(Ax^k + By^k - b) + r\beta B(y^k - y^{k+1})\Big\} \geq 0 \end{aligned}$$

and obtain (5.5) directly from the last inequality. This lemma is proved. ∎

Consequently, we get the following theorem, which is important for the proof of the main theorem in the next section.

Theorem 5.3. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). Then we have*

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 \geq\ & \frac{(1 - r)^2}{1 + r}\beta\|B(y^k - y^{k+1})\|^2 + (2 - r - s)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &+ \frac{2(1 - r)(1 - s)}{1 + r}\beta(Ax^k + By^k - b)^T B(y^k - y^{k+1}). \end{aligned} \tag{5.12}$$

Proof. Substituting (5.5) into (5.1), we obtain the assertion (5.12) immediately. ∎

5.2. Proofs for different subdomains. Now, we are at the stage to prove Theorem 4.4 for the subdomains defined in (4.16) one by one.

5.2.1. (s,r) in the subdomain $\mathcal{D}_1$.

Lemma 5.4. *For arbitrarily fixed $(s,r)\in\mathcal{D}_1$, there are constants $C_1, C_2>0$ such that the inequality (4.17a) holds and Theorem 4.4 is true for $(s,r)\in\mathcal{D}_1$.*

Proof. Recall that $\mathcal{D}_1=\big\{(s,r)\,|\,s\in(0,1),\ r\in(-1,1),\ r+s>0\big\}$ (see (4.16)). According to (4.15), we have

$$\|v^k-\tilde v^k\|_G^2=\left\|\begin{array}{c} B(y^k-\tilde y^k)\\ \lambda^k-\tilde\lambda^k\end{array}\right\|_{G_0}^2,$$

where

$$G_0=\begin{pmatrix}\beta(1-s)I & -(1-s)I\\ -(1-s)I & \frac{1}{\beta}\big(2-(r+s)\big)I\end{pmatrix}\succ 0\quad\forall\,(s,r)\in\mathcal{D}_1.$$

On the other hand, according to (3.1), we have

$$\begin{aligned}\begin{pmatrix}B(y^k-\tilde y^k)\\ \lambda^k-\tilde\lambda^k\end{pmatrix}&=\begin{pmatrix}B(y^k-y^{k+1})\\ \beta(Ax^{k+1}+By^k-b)\end{pmatrix}\\&=\begin{pmatrix}B(y^k-y^{k+1})\\ \beta B(y^k-y^{k+1})+\beta(Ax^{k+1}+By^{k+1}-b)\end{pmatrix}\\&=\begin{pmatrix}I & 0\\ \beta I & \beta I\end{pmatrix}\begin{pmatrix}B(y^k-y^{k+1})\\ Ax^{k+1}+By^{k+1}-b\end{pmatrix}.\end{aligned}$$

Consequently, we obtain

$$\|v^k-\tilde v^k\|_G^2=\left\|\begin{array}{c} B(y^k-\tilde y^k)\\ \lambda^k-\tilde\lambda^k\end{array}\right\|_{G_0}^2=\left\|\begin{array}{c} B(y^k-y^{k+1})\\ Ax^{k+1}+By^{k+1}-b\end{array}\right\|_{\widetilde G_0}^2, \tag{5.13}$$

where

$$\widetilde G_0=L^TG_0L\qquad\text{and}\qquad L=\begin{pmatrix}I & 0\\ \beta I & \beta I\end{pmatrix}.$$

Due to the positive definiteness of G_0 and nonsingularity of L, $\widetilde G_0$ is positive definite. Finally, the assertion follows from (5.13) and the positive definiteness of $\widetilde G_0$. ■

Remark 5.5. The symmetric ADMM (1.8) with shrunken step sizes is a special case of (1.9) with $r=s\in(0,1)$. In this case (see (4.5)),

$$H=\frac{1}{2r}\begin{pmatrix}B^T & 0\\ 0 & I\end{pmatrix}\begin{pmatrix}r(2-r)\beta I & -rI\\ -rI & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix}B & 0\\ 0 & I\end{pmatrix}$$

and (see (4.15))

$$G=(1-r)\begin{pmatrix}B^T & 0\\ 0 & I\end{pmatrix}\begin{pmatrix}\beta I & -I\\ -I & \frac{2}{\beta}I\end{pmatrix}\begin{pmatrix}B & 0\\ 0 & I\end{pmatrix}.$$

5.2.2. (s, r) in the subdomain $\mathcal{D}_2$. Theorem 5.3 is crucial for proving the convergence of (1.9) with $s \geq 1$. Now, we use it to show that Theorem 4.4 is true for $(s, r) \in \mathcal{D}_2$.

Lemma 5.6. *For any $(s, r) \in \mathcal{D}_2$, the inequality (4.17a) holds with*

$$C_1 = \frac{(1-r)^2}{1+r} > 0 \qquad \text{and} \qquad C_2 = 1 - r > 0.$$

Proof. Notice that $\mathcal{D}_2 = \{(s, r) \,|\, s = 1,\ r \in (-1, 1)\}$ (see (4.16)). Setting $s = 1$ in (5.12), we obtain

$$\|v^k - \tilde{v}^k\|_G^2 \geq \frac{(1-r)^2}{1+r}\beta\|B(y^k - y^{k+1})\|^2 + (1-r)\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \tag{5.14}$$

The lemma is proved. ∎

Remark 5.7. Recall that the original ADMM (1.4) is a special case of (1.9) with $r = 0$ and $s = 1$. In this special case, the matrix H has the form (see (4.4))

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{5.15}$$

Notice that in this special case, we have

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

Thus, setting $r = 0$ in (5.14), we get

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 &\geq \beta\|B(y^k - y^{k+1})\|^2 + \beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &= \beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2. \end{aligned}$$

Using the special matrix H in (5.15), we get

$$\|v^k - \tilde{v}^k\|_G^2 \geq \|v^k - v^{k+1}\|_H^2. \tag{5.16}$$

Remark 5.8. Indeed, the generalized ADMM in [13] is a special case of (1.9) with $(s, r) \in \mathcal{D}_2$. Let us elaborate on the detail, which is not straightforward. In [13], only the special case of (1.1) with $B = -I$ and $b = 0$ was discussed. But it is trivial (see, e.g., [14, 15, 17]) to extend the generalized ADMM in [13] to the model (1.1). The resulting iterative scheme can be written as

$$\left\{\begin{aligned} x^{k+1} &= \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, && \text{(5.17a)} \\ y^{k+1} &= \arg\min\{\theta_2(y) - y^T B^T \lambda^k && \text{(5.17b)} \\ &\qquad + \frac{\beta}{2}\|[\alpha A x^{k+1} - (1-\alpha)(By^k - b)] + By - b\|^2 | y \in \mathcal{Y}\}, \\ \lambda^{k+1} &= \lambda^k - \beta\{[\alpha A x^{k+1} - (1-\alpha)(By^k - b)] + By^{k+1} - b\}, && \text{(5.17c)} \end{aligned}\right.$$

where the parameter $\alpha \in (0,2)$ is a relaxation factor. To see why (5.17) is a special case of (1.9), let us choose $r = \alpha - 1$. Then, the y-subproblem in (5.17b) can be written as

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

Ignoring some constant terms in the objective function, we further rewrite it as

$$y^{k+1} = \arg\min\left\{\theta_2(y) - y^T B^T[\lambda^k - r\beta(Ax^{k+1} + By^k - b)] + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\right\}.$$

Recall (1.9b). We have

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \tag{5.18}$$

which enables us to rewrite the y-subproblem as

$$y^{k+1} = \arg\min\left\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\right\}. \tag{5.19}$$

Analogously, since $\alpha = 1 + r$, the step (5.17c) can be rewritten as

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

which, together with (5.18), yields

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b). \tag{5.20}$$

Combining (5.17a), (5.18), (5.19), and (5.20), we now can rewrite the generalized ADMM (5.17) as

$$\begin{cases} x^{k+1} = \arg\min\left\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\right\}, & \text{(5.21a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & \text{(5.21b)} \\ y^{k+1} = \arg\min\left\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\right\}, & \text{(5.21c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & \text{(5.21d)} \end{cases}$$

where $r \in (-1,1)$. Thus, the generalized ADMM (5.17) is a special case of the symmetric ADMM scheme (1.9) with the restriction $(s,r) \in \mathcal{D}_2$.

5.2.3. (s,r) in the subdomain $\mathcal{D}_3$. Now, we turn to verify Theorem 4.4 for the subdomain $\mathcal{D}_3$. First, let us define

$$S := 1 + s - s^2. \tag{5.22a}$$

Because

$$1 + s - s^2 = -\Big(s - \frac{1-\sqrt{5}}{2}\Big)\Big(s - \frac{1+\sqrt{5}}{2}\Big),$$

we have

$$S > 0 \quad \forall\, s \in \Big(1, \frac{1+\sqrt{5}}{2}\Big). \tag{5.22b}$$

The result in Theorem 5.3 is the basis for the proof in this section.

Recall that $\mathcal{D}_3 = \big\{(s,r)\,|\, s \in \big(1, \frac{1+\sqrt{5}}{2}\big), r = 0\big\}$ (see (4.16)). Since $r = 0$, it follows from Theorem 5.3 that

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 \ge{}& \beta\|B(y^k - y^{k+1})\|^2 + (2-s)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &+ 2(1-s)\beta(Ax^k + By^k - b)^T B(y^k - y^{k+1}). \end{aligned} \tag{5.23}$$

Now, we define a constant T_1 as

$$T_1 := \frac{1}{3}(s^2 - s + 5), \tag{5.24}$$

and thus we have

$$T_1 - s = \frac{1}{3}(s^2 - 4s + 5) = \frac{1}{3}[(s-2)^2 + 1] > \frac{1}{3}. \tag{5.25}$$

Using the Cauchy–Schwarz inequality to the crossing term in the right-hand side of (5.23) and because of $T_1 - s > 0$, we have

$$\begin{aligned} &2(1-s)\beta(Ax^k + By^k - b)^T B(y^k - y^{k+1}) \\ &\ge -\big(T_1 - s\big)\beta\|Ax^k + By^k - b\|^2 - \frac{(1-s)^2}{T_1 - s}\beta\|B(y^k - y^{k+1})\|^2. \end{aligned} \tag{5.26}$$

Lemma 5.9. *For any* $(s,r) \in \mathcal{D}_3$*, the inequality* (4.17b) *holds with*

$$C_0 = T_1 - s > \frac{1}{3}, \qquad C_1 = \frac{2(1+s-s^2)}{1 + (s-2)^2} > 0, \qquad \text{and} \qquad C_2 = \frac{1}{3}S > 0.$$

Proof. Substituting (5.26) into (5.23), we obtain

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 \ge{}& \big(T_1 - s\big)\beta\big(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\big) \\ &+ \Big(1 - \frac{(1-s)^2}{T_1 - s}\Big)\beta\|B(y^k - y^{k+1})\|^2 + (2 - T_1)\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned} \tag{5.27}$$

According to (4.17b), we set

$$C_0 = T_1 - s, \qquad C_1 = 1 - \frac{(1-s)^2}{T_1 - s}, \qquad \text{and} \qquad C_2 = 2 - T_1.$$

It follows from (5.25) that $C_0 \ge \frac{1}{3}$. Then, using (5.24), we obtain

$$C_1 = 1 - \frac{(1-s)^2}{T_1 - s} = \frac{(s^2-4s+5) - 3(1-s)^2}{s^2-4s+5} = \frac{2(1+s-s^2)}{1+(s-2)^2}.$$

Because $1+(s-2)^2 < 2$ for all $s \in \big(1, \frac{1+\sqrt{5}}{2}\big)$, it follows that

$$C_1 = \frac{2(1+s-s^2)}{1+(s-2)^2} \ge 1+s-s^2 = S. \tag{5.28}$$

Finally,

$$C_2 = 2 - T_1 = 2 - \frac{1}{3}(s^2-s+5) = \frac{1}{3}(1+s-s^2) = \frac{1}{3}S. \tag{5.29}$$

The assertion of this lemma is proved ∎

5.2.4. (s,r) in the subdomain $\mathcal{D}_4$. Recall that $\mathcal{D}_4 = \big\{(s,r)\,|\, s \in \big(1, \frac{1+\sqrt{5}}{2}\big),\ r \in (0,1)\ \&\ r < 1+s-s^2\big\}$ (see (4.16)). We say $r < 1+s-s^2$ is an additional restriction in $\mathcal{D}_4$ and define

$$T_2 = r + s + (1-s)^2. \tag{5.30}$$

Notice that $T_2 - (r+s) = (1-s)^2 > 0$ for all $s \in \big(1, \frac{1+\sqrt{5}}{2}\big)$. For the crossing term in the right-hand side of (5.12), using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} &\frac{2(1-r)(1-s)}{1+r}\beta(Ax^k+By^k-b)^T B(y^k-y^{k+1}) \\ &\ge -[T_2-(r+s)]\beta\|Ax^k+By^k-b\|^2 - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_2-(r+s)]}\beta\|B(y^k-y^{k+1})\|^2. \end{aligned} \tag{5.31}$$

Lemma 5.10. *For any $(s,r) \in \mathcal{D}_4$, the inequality (4.17b) holds with*

$$C_0 = (1-s)^2 > 0, \qquad C_1 = r\frac{(1-r)^2}{(1+r)^2} > 0, \qquad \text{and} \qquad C_2 = 2 - T_2 > 0,$$

where T_2 is defined in (5.30).

Proof. Substituting (5.31) into (5.12) yields

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 \ge\ & [T_2-(r+s)]\beta\big(\|Ax^{k+1}+By^{k+1}-b\|^2 - \|Ax^k+By^k-b\|^2\big) \\ &+\Big(\frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_2-(r+s)]}\Big)\beta\|B(y^k-y^{k+1})\|^2 \\ &+(2-T_2)\beta\|Ax^{k+1}+By^{k+1}-b\|^2. \end{aligned} \tag{5.32}$$

According to (4.17b), we set

$$C_0 = T_2 - (r+s), \qquad C_1 = \frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_2-(r+s)]}, \qquad \text{and} \qquad C_2 = 2 - T_2.$$

In the following, we show that $C_0, C_1, C_2 > 0$. First, it follows from (5.30) that $C_0 = (1-s)^2 > 0$ for all $s \in \big(1, \frac{1+\sqrt{5}}{2}\big)$. Indeed, using (5.30), we have

$$C_1 = \frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_2-(r+s)]} = \frac{(1-r)^2}{1+r} - \frac{(1-r)^2}{(1+r)^2} = r\frac{(1-r)^2}{(1+r)^2}.$$

Since $r \in (0,1)$ in $\mathcal{D}_4$, we have $C_1 > 0$. Adding the term $1 - s + s^2$ to both sides of the additional restriction

$$r < 1 + s - s^2,$$

we obtain

$$r + s + (1-s)^2 < 2.$$

According to the definition (5.30), the left-hand side of the last inequality is T_2 and thus $C_2 = 2 - T_2 > 0$. The lemma is proved. ∎

5.2.5. (s,r) in the subdomain $\mathcal{D}_5$. Recall that $\mathcal{D}_5 = \big\{(s,r)\,|\, s \in \big(1, \frac{1+\sqrt{5}}{2}\big),\ r \in (-1,0)\ \&\ -r < 1+s-s^2\big\}$ (see (4.16)). We say $s^2 - s - 1 < r$ is an additional restriction in $\mathcal{D}_5$ and define

$$T_3 = r + s + \frac{(s^2-s)(2-s)}{1+r}. \tag{5.33}$$

Notice that $T_3-(r+s) = \frac{(s^2-s)(2-s)}{1+r} > 0$ for all $s \in \big(1, \frac{1+\sqrt{5}}{2}\big)$ and $r \in (-1,0)$. For the crossing term in the right-hand side of (5.12), using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
&\frac{2(1-r)(1-s)}{1+r}\beta(Ax^k + By^k - b)^T B(y^k - y^{k+1}) \\
&\geq -[T_3 - (r+s)]\beta\|Ax^k + By^k - b\|^2 - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_3-(r+s)]}\beta\|B(y^k - y^{k+1})\|^2. \qquad (5.34)
\end{aligned}$$

Lemma 5.11. *For any $(s,r) \in \mathcal{D}_5$, the inequality (4.17b) holds with*

$$C_0 = \frac{(s^2-s)(2-s)}{1+r} > 0, \qquad C_1 = \frac{(1-r)^2(1+s-s^2)}{s(1+r)(2-s)} > 0, \qquad \textit{and} \qquad C_2 = 2 - T_3 > 0,$$

where T_3 is defined in (5.33).

Proof. Substituting (5.34) into (5.12) yields

$$\begin{aligned}
\|v^k - \tilde{v}^k\|_G^2 \geq{}& [T_3 - (r+s)]\beta\big(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2\big) \\
&+\Big(\frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_3-(r+s)]}\Big)\beta\|B(y^k - y^{k+1})\|^2 \\
&+ (2-T_3)\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \qquad (5.35)
\end{aligned}$$

According to (4.17b), we set

$$C_0 = T_3 - (r+s), \qquad C_1 = \frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_3-(r+s)]}, \qquad \text{and} \qquad C_2 = 2 - T_3.$$

In the following, we show that $C_0, C_1, C_2 > 0$. First, it follows from (5.33) that $C_0 = \frac{(s^2-s)(2-s)}{r+1} > 0$ for all $s \in \left(1, \frac{1+\sqrt{5}}{2}\right)$ and $r \in (-1, 0)$. Indeed, using (5.33), we have

$$C_1 = \frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_3-(r+s)]} = \frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)(s^2-s)(2-s)}$$
$$= \frac{(1-r)^2}{1+r} + \frac{(1-r)^2(1-s)}{s(1+r)(2-s)} = \frac{(1-r)^2(1+s-s^2)}{s(1+r)(2-s)}.$$

Since $r \in (-1, 0)$ and $s \in \left(1, \frac{1+\sqrt{5}}{2}\right)$ in $\mathcal{D}_5$ and $1 + s - s^2 > 0$ (see (5.22)), we have $C_1 > 0$. From the restriction in $\mathcal{D}_5$, namely, $s^2 - s - 1 < r$, we obtain

$$s^2 - s < 1 + r.$$

Notice that $s \in \left(1, \frac{1+\sqrt{5}}{2}\right)$, we also have $s^2 - s > 0$, and thus

$$\frac{s^2-s}{1+r} < 1.$$

Substituting the last inequality into (5.33), we get

$$T_3 = r + s + \left(\frac{s^2-s}{1+r}\right)(2-s) < r + s + (2-s) = 2 + r, \tag{5.36}$$

which, together with $r < 0$, implies $C_2 = 2 - T_3 > 0$. The lemma is proved. ■

Since we have proved Lemmas 5.4, 5.6, 5.9, 5.10, and 5.11, the proof for Theorem 4.4 is complete. In the next section, we will prove the convergence and convergence rate.

6. Convergence analysis.

In this section, we establish the global convergence and estimate the convergence rate in terms of the iteration complexity for the symmetric ADMM (1.9).

6.1. Some corollaries.

First, we show two immediate corollaries of the main result in Theorem 4.4, whose proofs are omitted. The first one is based on (4.12) and Theorem 4.4.

Corollary 6.1. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). Then, we have the following assertions:*

1. *For any given $(s, r) \in \mathcal{D}_1 \cup \mathcal{D}_2$, there are constants $C_1, C_2 > 0$ such that*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \left(C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2\right) \quad \forall v^* \in \mathcal{V}^*. \tag{6.1a}$$

2. *For $(s, r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$, there are constants $C_0, C_1, C_2 > 0$ such that*

$$\begin{aligned} &\|v^{k+1} - v^*\|_H^2 + C_0\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &\le \left(\|v^k - v^*\|_H^2 + C_0\beta\|Ax^k + By^k - b\|^2\right) \\ &\quad -\left(C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2\right) \quad \forall v^* \in \mathcal{V}^*. \end{aligned} \tag{6.1b}$$

The second one is based on (4.8) and Theorem 4.4.

Corollary 6.2. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). Then, we have the following assertions:*

1. *For $(s,r) \in \mathcal{D}_1 \cup \mathcal{D}_2$, it holds that*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \geq \frac{1}{2}\|v - v^{k+1}\|_H^2 \ \ \forall w \in \Omega. \tag{6.2a}$$

2. *For $(s,r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$, we have*

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\big(\|v - v^k\|_H^2 + C_0\beta\|Ax^k + By^k - b\|^2\big) \\ &\geq \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 + C_0\beta\|Ax^{k+1} + By^{k+1} - b\|^2\big) \ \ \forall w \in \Omega. \end{aligned} \tag{6.2b}$$

These two corollaries will be used for the upcoming convergence analysis.

6.2. Global convergence. We summarize the global convergence of (1.9) in the following theorem and its proof follows straightforwardly from Corollary 6.1.

Theorem 6.3. *For the sequence $\{w^k\}$ generated by the symmetric ADMM (1.9), we have*

$$\lim_{k\to\infty}\big(\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2\big) = 0. \tag{6.3}$$

Moreover, if the matrix B is assumed to be full column rank, then the sequence $\{v^k\}$ converges to a solution point $v^\infty \in \mathcal{V}^*$.

Proof. Let (y^0, λ^0) be the initial iterate. For $(s,r) \in \mathcal{D}_1 \cup \mathcal{D}_2$, according to (6.1a) in Corollary 6.1, there are constants $C_1, C_2 > 0$ such that

$$\sum_{k=0}^{\infty}\big(C_1\|B(y^k - y^{k+1})\|^2 + C_2\|Ax^{k+1} + By^{k+1} - b\|^2\big) \leq \|v^0 - v^*\|_H^2$$

and thus we obtain the assertion (6.3). For $(s,r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$, using (6.1b) in Corollary 6.1 (note that x^1 is generated by the given (y^0, λ^0)), we have

$$\begin{aligned} &\sum_{k=1}^{\infty}\big(C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2\big) \\ &\leq \big(\|v^1 - v^*\|_H^2 + C_0\beta\|Ax^1 + By^1 - b\|^2\big), \end{aligned}$$

and the assertion (6.3) is proved. The convergence of (1.9) is thus proved in sense of (6.3).

Furthermore, it follows from (6.1a) and (6.1b) that the sequence $\{v^k\}$ is in a compact set and it has a subsequence $\{v^{k_j}\}$ converging to a cluster point, say, v^∞. Let $\tilde{x}^\infty$ be induced by (1.9a) with given $(y^\infty, \lambda^\infty)$. Recall the matrix B is assumed to be full column rank. Because of (6.3), we have

$$B(y^\infty - \tilde{y}^\infty) = 0 \qquad \text{and} \qquad \lambda^\infty - \tilde{\lambda}^\infty = 0.$$

Then, it follows from (3.2a) that

$$\tilde{w}^\infty \in \Omega, \quad \theta(u) - \theta(\tilde{u}^\infty) + (w - \tilde{w}^\infty)^T F(\tilde{w}^\infty) \geq 0 \quad \forall w \in \Omega, \tag{6.4}$$

and thus $\tilde{w}^\infty = w^\infty$ is a solution point of (2.5). On the other hand, because (6.1a) (resp., (6.1b)) holds for any solution point of (2.5) and $w^\infty \in \Omega^*$, we have

$$\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2 \tag{6.5a}$$

and

$$\|v^{k+1} - v^\infty\|_H^2 + C_0\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \leq \|v^k - v^\infty\|_H^2 + C_0\beta\|Ax^k + By^k - b\|^2, \tag{6.5b}$$

respectively. Because $\lim_{k\to\infty} \|Ax^k + By^k - b\|^2 = 0$, the sequence $\{v^k\}$ cannot have another cluster point and thus it converges to a solution point $v^* = v^\infty \in \mathcal{V}^*$. ■

6.3. Convergence rate. To estimate the convergence rate in terms of the iteration complexity, we need a characterization of the solution set of VI (2.5), which is described in the following theorem. The proof can be found in [30, Theorem 2.3.5] or [33, Theorem 2.1].

Theorem 6.4. *The solution set of VI(Ω, F, θ) is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega} \big\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\big\}. \tag{6.6}$$

Therefore, for a given accuracy $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution point of VI(Ω, F, θ) if it satisfies

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

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \leq 1\}.$$

To estimate the convergence rate in terms of the iteration complexity for a sequence $\{w^k\}$, we need to show that for given $\epsilon > 0$, after t iterations, this sequence can offer a point $\tilde{w} \in \Omega$ such that

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w\in\mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \leq \epsilon. \tag{6.7}$$

Now, using the results in Corollary 6.2, we prove the $O(1/t)$ convergence rate theorem.

Theorem 6.5. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.9) and $\tilde{w}^k$ be defined by (3.1). Then, for $(s, r) \in \mathcal{D}_1 \cup \mathcal{D}_2$ and any integer $t > 0$, we have*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|v - v^0\|_H^2 \quad \forall w \in \Omega, \tag{6.8a}$$

where

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{6.8b}$$

For $(s,r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$ *and any integer number* $t > 0$, *we have*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2t}\big(\|v - v^1\|_H^2 + C_0\|Ax^1 + By^1 - b\|^2\big) \quad \forall w \in \Omega, \tag{6.9a}$$

where

$$\tilde{w}_t = \frac{1}{t}\sum_{k=1}^{t}\tilde{w}^k. \tag{6.9b}$$

Proof. Let us rewrite the results in Corollary 6.2 as

$$\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) + \frac{1}{2}\|v - v^{k+1}\|_H^2 \le \frac{1}{2}\|v - v^k\|_H^2 \ \ \forall w \in \Omega \tag{6.10a}$$

and

$$\begin{aligned} &\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) + \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 + C_0\beta\|Ax^{k+1} + By^{k+1} - b\|^2\big) \\ &\le \frac{1}{2}\big(\|v - v^k\|_H^2 + C_0\beta\|Ax^k + By^k - b\|^2\big) \ \ \forall w \in \Omega, \end{aligned} \tag{6.10b}$$

for $(s,r) \in \mathcal{D}_1 \cup \mathcal{D}_2$ and $(s,r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$, respectively. Summarizing the inequalities (6.10a) over $k = 0, 1, \ldots, t$, we obtain

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

Then, using the notation of $\tilde{w}_t$ in (6.8b), we can rewrite the last inequality as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2 \quad \forall w \in \Omega. \tag{6.11}$$

It follows from the definition of $\tilde{w}_t$ in (6.8b) that

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

Since $\theta(u)$ is convex, it follows that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it into (6.11), the assertion (6.8) follows directly. The proof for the assertion (6.9) is similar and omitted. ∎

It follows from (6.7), (6.8), and (6.9) that the symmetric ADMM (1.9) is able to generate an approximate solution point (i.e., $\tilde{w}_t$ defined in (6.8b) or (6.9b)) with an accuracy of $O(1/t)$ after t iterations. That is, a worse-case $O(1/t)$ convergence rate in the ergodic sense is established for the symmetric ADMM (1.9).

7. Numerical results.

In the literature, some special cases of the symmetric ADMM (1.9) have been widely verified by many applications such as the original ADMM where $r = 0$ and $s = 1$ and the ADMM with Fortin and Glowinski's larger step size where $r = 0$ and $s = \frac{\sqrt{5}+1}{2}$. In this section, we supplement with some more numerical results to verify the efficiency of the scheme (1.9) with other values of r and s. Given the well-verified efficiency of the original ADMM, our emphasis is simply verifying that it is still possible to outperform the original ADMM by choosing some other values of r and s in the symmetric ADMM scheme (1.9).

We test two fundamental models: the basis pursuit problem and total-variational image deblurring problem. Using the original ADMM to solve these two models or some of their variants has been well studied in the literature; thus we use the original ADMM as the benchmark to show the efficiency of the symmetric ADMM (1.9) with some values of r and s. All the codes were written in MATLAB R2015a and all experiments were performed on a desktop with Windows 7 and an Intel Core i5-4590 CPU processor (3.30 GHz) with 8 GB memory.

7.1. Basis pursuit problem.

We first test the basis pursuit model:

$$\min_{x\in\mathbb{R}^n} \|x\|_1 + \frac{1}{2\mu}\|Ax-b\|_2^2, \tag{7.1}$$

where $\|x\|_1 := \sum_{i=1}^n |x_i|$, $A \in \mathbb{R}^{m\times n}(m \ll n)$ is an encoding matrix (full row-rank), $b \in \mathbb{R}^m$ represents a compressed signal, and $\mu > 0$ is a parameter. We refer to, e.g., [9] for details. The model (7.1) is a core problem in areas such as compressive sensing, variable selection, and so on. It is well known that the model (7.1) can be reformulated as

$$\min\left\{ \|x\|_1 + \frac{1}{2\mu}\|Ay-b\|_2^2 \mid x-y=0, x\in\mathbb{R}^n, y\in\mathbb{R}^n \right\}, \tag{7.2}$$

where $y \in \mathbb{R}^n$ is an auxiliary variable. Thus, (7.2) is a special case of (1.1) and the proposed symmetric ADMM (1.9) is applicable.

In our experiments, A and b in (7.1) are generated in the same way as [43]: $A \in \mathbb{R}^{m\times n}$ is a random Gaussian matrix whose rows are orthonormalized by the QR factorization. The true signal x^* has p nonzero elements whose positions are determined randomly, and the nonzero values are generated with the standard deviation as 1. The vector b has zero mean white noise generated by MATLAB script: `b = A*xstar + sigma*randn(m,1)`, where σ(`sigma`) is the standard deviation of the additive Gaussian noise.

For succinctness, we report only the results for the scenarios of $(m/n = 0.2, p/m = 0.2)$, and $(m/n = 0.2, p/m = 0.1)$ when $\sigma = 10^{-3}$, $n = 2^{12}$, and $\mu = 10^{-4}$ are fixed in (7.1). To implement (1.9) with different values of r and s, we fix $\beta = m/\|b\|_1$ as in [43], and the initial iterate is chosen as $(y^0, \lambda^0) = (\mathbf{1}, \mathbf{0})$. Below, we list some representative values of r and s such that the general symmetric version of ADMM (1.9) is slightly faster than its specific choice of $r = 0$ and $s = 1$, i.e., the original ADMM (1.4). Many other values close to the choices in Table 1 generate similar results; thus their results are omitted. In Table 1, It., CPU, and Obj represent the iteration numbers, computing time in seconds, and objective function value

Table 1
Numerical results of (1.9) *for* (7.1) $(n = 2^{12}, \sigma = 10^{-3}, \mu = 10^{-4})$.

	$m/n = 0.2, p/m = 0.2$			$m/n = 0.2, p/m = 0.1$		
	It.	CPU	Obj	It.	CPU	Obj
$r = -0.3, \quad s = 1.41$	69	1.9	142.1	54	1.5	67.03
$r = -0.2, \quad s = 1.48$	67	1.9	142.1	53	1.5	67.03
$r = -0.2, \quad s = 1.52$	66	1.8	142.2	53	1.5	66.93
$r = 0.1, \quad s = 1.57$	69	1.9	142.0	56	1.5	67.03
$r = 0, s = 1$ (ADMM (1.4))	74	2.0	142.1	64	1.8	67.01

of current iteration, respectively. The efficiency of the general symmetric ADMM (1.9) with these values of r and s is thus shown.

7.2. Total-variational image deblurring model. The original ADMM (1.4) has been widely used in various image processing fields. In this subsection, we test the total-variational image deblurring model whose discretized version can be written as

$$\min_{y} \|Ay\|_1 + \frac{\lambda}{2}\|By - z\|^2, \tag{7.3}$$

where $y \in \mathcal{R}^n$ represents a digital clean image, $z \in \mathcal{R}^n$ is a corrupted input image, $A := (\partial_1, \partial_2)\colon \mathcal{R}^n \to \mathcal{R}^n \times \mathcal{R}^n$ is the discrete gradient operator with $\partial_1 : \mathcal{R}^n \to \mathcal{R}^n$ and $\partial_2 : \mathcal{R}^n \to \mathcal{R}^n$ the standard finite differences with periodic boundary conditions in the horizontal and vertical directions, respectively (see details in [40]), $B\colon \mathcal{R}^n \to \mathcal{R}^n$ is the matrix representation of a spatially invariant blurring operator, $\lambda > 0$ is a constant balancing the data-fidelity and total-variational regularization terms, and $\|\cdot\|_1$ defined on $\mathcal{R}^n \times \mathcal{R}^n$ is given by

$$\|x\|_1 := \sum_{\{i,j\}=1}^{n} \sqrt{|x^1_{i,j}|^2 + |x^2_{i,j}|^2} \quad \forall x = (x^1, x^2) \in \mathcal{R}^n \times \mathcal{R}^n.$$

This is a basic model for various more advanced image processing tasks and it has been studied extensively in the literature. Introducing the auxiliary variable x, we can reformulate (7.3) as

$$\begin{array}{ll} \min & \|x\|_1 + \frac{\lambda}{2}\|By - z\|^2 \\ \text{s.t.} & x - Ay = 0, \end{array}$$

which is a special case of the generic model (1.1) under our discussion and thus the symmetric ADMM (1.9) is applicable.

We test the image of cameraman.png (256×256). The clean image is degraded by the convolution generated via the scripts fspecial and imfilter in the MATLAB Image Processing Toolbox with the "motion" and then added by the zero-mean Gaussian noise with standard deviation $\sigma = 10^{-3}$ via the script of imnoise. We set the angle parameter "theta" $= 135$ and the motion distance parameter "len" $= 91$ for the "motion" blurring. The clean and corrupted images are shown in (a) and (b) in Figure 4.

(a) (b) (c)

Figure 4. (a) *Clean image.* (b) *Corrupted image.* (c) *Restored image by* (1.9) *with* $r = 0.8$ *and* $s = 1.17$.

Table 2
Numerical results of (1.9) *for* (7.3).

The symmetric version of ADMM (1.9)	It.	CPU	SNR
$r = 0.7, \quad s = 1.24$	13	0.2	21.29
$r = 0.7, \quad s = 1.14$	13	0.19	21.29
$r = 0.8, \quad s = 1.17$	12	0.14	21.29
$r = 0.8, \quad s = 1.07$	12	0.16	21.28
$r = 0.9, \quad s = 1.09$	12	0.16	21.29
$r = 0.9, \quad s = 1$	12	0.17	21.29
$r = 0, \quad s = 1$ (i.e., original ADMM (1.4))	19	0.28	21.26
$r = 0, \quad s = 1.61$	20	0.27	21.26

For succinctness, we fix $\lambda = 250$ in (7.3) and $\beta = 0.1$ when implementing the symmetric ADMM (1.9). We report the iteration number (It.) and computing time in seconds (CPU) when comparing the performance of (1.9) with different values of r and s. The quality of restored images is measured by the value of the SNR given by

$$\text{SNR}: = 20 \log_{10} \frac{\|y^*\|}{\|y^k - y^*\|}, \tag{7.4}$$

where y^k is the restored image and y^* is the ground truth.

In Table 2, we list several interior point cases of the domain D to show that they could result in better numerical performance than the original ADMM ($s = 1, r = 0$) and the point ($s = 1.61, r = 0$) that is very close to the corner point ($s = \frac{1+\sqrt{5}}{2} \approx 1.618, r = 0$) which corresponds to the ADMM with Fortin and Glowinski's larger step size (1.6). We choose some interior points in D for the comparison. For succinctness, we show only the image restored by the case of (1.9) with $r = 0.8$ and $s = 1.17$; see (c) in Figure 4. It can be seen that interior points in D may accelerate the convergence of the original ADMM (1.4); thus the necessity of considering the general symmetric version of ADMM (1.9) is meaningful.

Note that the point ($r = 0, s = 1.61$) is already much better than some other corner points such as ($r = -1, s = 1$) and ($r = 0, s = 0$). So we skip the comparison with other corner points for succinctness. Also, a number of other scenarios (the same image but with different parameters and/or blurring sizes, or other images) and many other standard total-variational image restoration models can be tested, but they are omitted for succinctness.

8. Conclusions. In this paper, we conducted a convergence analysis for the symmetric version of the ADMM with step sizes enlarged by Fortin and Glowinski's constant and justified the rationale of combining the original ADMM with these two numerically favorable techniques. Smaller step sizes are usually required to conservatively ensure convergence in theory, while they should be avoided practically to avoid slow convergence. Thus, ensuring the convergence of an algorithm with larger step sizes is usually more demanding to analyze. On the other hand, the symmetric ADMM with nonenlarged step sizes is already known to be not necessarily convergent despite its possible empirical efficiency. Thus, it seems not promising to guarantee the convergence if the symmetric ADMM is combined with larger step sizes. We provide a counterintuitive answer to this question and prove rigorously the convergence of the symmetric ADMM with larger step size enlarged by Fortin and Glowinski's constant. Our analysis is conducted in the generic convex programming context, and it unifies the convergence analysis for some ADMM-like algorithms, including some existing results in the literature as special cases.

The rule for enlarging step sizes by Fortin and Glowinski's constant is applicable to the general iteration scheme of the symmetric ADMM when the generic setting of the model (1.1) is considered, and it requires no additional computation (even though the "optimal" value of the constants may still be problem-dependent). Thus, our theory well explains the justification of the combination of two commonly used techniques in numerical implementation: using the symmetric version of ADMM and enlarging step sizes by Fortin and Glowinski's constants, which conventionally are used separately in the ADMM literature to accelerate the numerical efficiency for various applications. The proposed symmetric ADMM with larger step sizes is ready to use, with rigorous convergence analysis, for immediately accelerating the speed of solving a variety of applications by ADMM-like algorithms, including a wide range of imaging processing applications as well demonstrated in the rich literature.

Because of the equal role of the parameters r and s in the symmetric ADMM scheme (1.9), it is clear that the scheme (1.9) can be rewritten as

$$\begin{cases} y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^k, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(8.1a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - s\beta(Ax^k + By^{k+1} - b), & \text{(8.1b)} \\ x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^{k+1}, \lambda^{k+\frac{1}{2}}) \mid x \in \mathcal{X}\}, & \text{(8.1c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r\beta(Ax^{k+1} + By^{k+1} - b). & \text{(8.1d)} \end{cases}$$

Thus, we can easily extend our previous discussion from the domain $\mathcal{D}$ in Figure 1 to the larger and symmetric domain defined as

$$\mathcal{D} = \big\{(s,r) \mid r+s>0, \;\; |r| < 1+s-s^2, \;\; |s| < 1+r-r^2\big\}, \tag{8.2}$$

which is displayed in Figure 5. Essentially, discussing the convergence for the symmetric ADMM (1.9) over the domain $\mathcal{D}$ defined in (1.10) is not easy, while its extension to the symmetric domain in Figure 5 is trivial. Thus, for succinctness, we omit the detail of this extension to the domain in Figure 5.

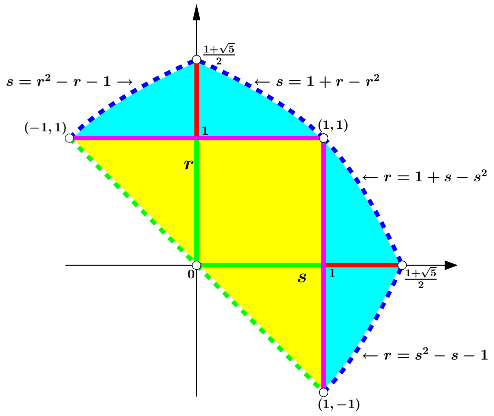

Figure 5. *A symmetric step size domain with convergence for the symmetric ADMM* (8.2).

REFERENCES

[1] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, *An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems*, IEEE Trans. Image Process., 20 (2011), pp. 681–695.

[2] M. Almeida and M. Figueiredo, *Deconvolving images with unknown boundaries using the alternating direction method of multipliers*, IEEE Trans. Image Process., 22 (2013), pp. 3084–3096.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[4] E. Blum and W. Oettli, *Mathematische Optimierung,* Econometrics Oper. Rese. 20, Springer-Verlag, Berlin, 1975.

[5] 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.

[6] J. F. Cai, S. J. Osher, and Z. Shen, *Split Bregman methods and frame based image restoration*, Multiscale Model. Simul., 8 (2009), pp. 337–369.

[7] T. F. Chan and R. Glowinski, *Finite Element Approximation and Iterative Solution of a Class of Mildly Nonlinear Elliptic Equations*, Technical Report STAN-CS-78-674, Computer Science Department, Stanford University, Stanford, CA, 1978.

[8] C. H. Chen, B. S. He, and X. M. Yuan, *Matrix completion via alternating direction method*, IMA J. Numer. Anal., 32 (2012), pp. 227–245.

[9] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), pp. 129–159.

[10] E. Corman and X. M. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optim., 24 (2014), pp. 1614–1638.

[11] 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.

[12] J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph.D. thesis, MIT, Cambridge, MA, 1989.

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

[14] J. Eckstein and M. Fukushima, *Reformulations and applications of the alternating direction method of multipliers*, in Large Scale Optimization, W. W. Hager, D. W. Hearn, and P. M. Pardalos, eds., Springer, New York, US, 1994, pp. 115–134.

[15] J. Eckstein and W. Yao, *Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives*, Pacific J. Optim., 11 (2015), pp. 619–644.

[16] E. Esser, *Applications of Lagrangian-Based Alternating Direction Methods and Connections to Split Bregman*, manuscript, 2009.

[17] E. X. Fang, B. S. He, H. Liu, and X. M. Yuan, *The generalized alternating direction method of multipliers: New theoretical insights and applications*, Math. Program. Comput., 7 (2015), pp. 149–187.

[18] M. Fortin and R. Glowinski, eds., *Méthodes de Lagrangien Augmenté: Application à la Résolution Numérique de Problèmes aux Limites*, Dunod, Paris, 1982.

[19] M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam, The Netherlands, 1983.

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

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

[22] R. Glowinski, *On alternating direction methods of multipliers: A historical perspective*, in Modeling, Simulation and Optimization for Science and Technology, W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanmäki, and O. Pironneau, eds., Comput. Methods Appli. Sci. 34, Springer, Dordrecht, The Netherlands, 2014, pp. 59–82.

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

[24] R. Glowinski, T. Karkkainen, and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scientific Computing, Variational Problems and Applications, E. Heikkola, Y. Kuznetsov, P. Neittaanmäki, and O. Pironneau, eds., CIMNE, Barcelona, 2003, pp. 67–79.

[25] R. Glowinski and A. Marrocco, *Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Rev. Fr. Autom. Inform. Rech. Opér. Anal. Numér., 2 (1975), pp. 41–76.

[26] R. Glowinski, S. J. Osher, and W. Yin, eds., *Splitting Methods in Communication and Imaging*, Science, and Engineering, Springer, to appear.

[27] T. Goldstein and S. J. Osher, *The split Bregman method for* L1 *regularized problems*, SIAM J. Imaging. Sci., 2 (2008), pp. 323–343.

[28] T. Goldstein, X. Bresson, and S. J. Osher, *Geometric applications of the split Bregman method: Segmentation and surface reconstruction*, J. Sci. Comput., 45 (2010), pp. 272–293.

[29] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, *Fast alternating direction optimization methods*, SIAM J. Imaging. Sci., 7 (2014), pp. 1588–1623.

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

[31] B. S. He, H. Liu, Z. R. Wang, and X. M. Yuan, *A strictly contractive Peaceman-Rachford splitting method for convex programming*, SIAM J. Optim., 24 (2014), pp. 1011–1040.

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

[33] 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), pp. 702–709.

[34] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl, 4 (1969), pp. 303–320.

[35] M. Li and X. M. Yuan, *A strictly contractive Peaceman-Rachford splitting method with logarithmic-quadratic proximal regularization for convex programming*, Math. Oper. Res., 40 (2015), pp. 842–858.

[36] X. X. Li and X. M. Yuan, *A proximal strictly contractive Peaceman-Rachford splitting method for convex programming with applications to imaging*, SIAM J. Imaging. Sci., 8 (2015), pp. 1332–1365.

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

[38] D. H. Peaceman and H. H. Rachford, *The numerical solution of parabolic elliptic differential equations*, SIAM J. Appl. Math., 3 (1955), pp. 28–41.

[39] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletchered, Academic Press, New York, 1969, pp. 283–298.

[40] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D., 60 (1992), pp. 259–268.

[41] S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, in Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, X.-C. Tai et al., eds., Lecture Notes in Comput. Sci. 5567, Springer, Berlin, 2009, pp. 464–476.

[42] S. Setzer, *Operator splittings, Bregman methods and frame shrinkage in image processing*, Int. J. Comput. Vis., 92 (2011), pp. 265–280.

[43] J. F. Yang and Y. Zhang, *Alternating direction algorithms for* L1-problems in compressive sensing, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.

[44] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2010), pp. 20–46.