

On the Step Size of Symmetric Alternating Directions Method of Multipliers

Bingsheng He¹ Feng Ma² Xiaoming Yuan³

May 26, 2015

Abstract. The alternating direction method of multipliers (ADMM) is an application of the Douglas-Rachford splitting method; and the symmetric version of ADMM which updates the Lagrange multiplier twice at each iteration is an application of the Peaceman-Rachford splitting method. Sometimes the symmetric ADMM works empirically; but theoretically its convergence is not guaranteed. 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 shrank conservatively. In this paper, we focus on the convex programming context and specify a step size domain that can ensure the convergence of symmetric ADMM. In particular, it is shown for the first time that, we can choose Glowinski's larger step size for one of the Lagrange-multiplier-updating steps at each iteration; so shrinking both the step sizes is not necessary for the symmetric ADMM. Some known results in the ADMM literature turn out to be special cases of our discussion.

Keywords. Convex programming, Alternating directions method of multipliers, Operator splitting methods, Peaceman-Rachford splitting method, Step size.

1 Introduction

We consider the model

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

where $A \in \mathbb{R}^{m \times n_1}$, $B \in \mathbb{R}^{m \times n_2}$, $b \in \mathbb{R}^m$, $\mathcal{X} \subset \mathbb{R}^{n_1}$ and $\mathcal{Y} \subset \mathbb{R}^{n_2}$ are closed convex sets, $\theta_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R}$ and $\theta_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R}$ are convex (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

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

¹Department of Mathematics, Nanjing University, Nanjing, 210093, China. This author was supported by the NSFC Grant 11471156. Email: hebma@nju.edu.cn

²College of Communications Engineering, PLA University of Science and Technology, Nanjing, 210007, Email: mafengnju@gmail.com

³Department of Mathematics, Hong Kong Baptist University, Hong Kong. This author was supported by the General Research Fund from Hong Kong Research Grants Council: 12302514. Email: xmyuan@hkbu.edu.hk

where λ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter for the linear constraints. Thus, applying directly the augmented Lagrangian method (ALM) in [21, 24] to (1.1), the iterative scheme is

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

We are only interested in the case where the functions θ_1 and θ_2 may have their own properties/structures and it is worth to take 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 inspired the alternating directions method of multipliers (ADMM), which was originally proposed in [2, 16] and now has received much attention from a wide range of areas such as image processing, statistical learning, computer vision, wireless communication network and cloud computing. Starting with an initial iterate $(y^0, \lambda^0) \in \mathcal{Y} \times \Re^m$, the ADMM generates the sequence via the scheme

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

We refer to, e.g., [1, 10, 14], for some reviews on ADMM.

In [13], Glowinski suggested attaching a relaxation factor to the Lagrange-multiplier-updating step in (1.4). This results in the scheme

$$\begin{cases} x^{k+1} = \arg \min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.5a) \\ y^{k+1} = \arg \min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & (1.5b) \\ \lambda^{k+1} = \lambda^k - s\beta(Ax^{k+1} + By^{k+1} - b), & (1.5c) \end{cases}$$

where the parameter s can be chosen in the interval $(0, \frac{1+\sqrt{5}}{2})$ and thus it becomes possible to enlarge the step size for updating the Lagrange multiplier. The ADMM scheme (1.5) with Glowinski's larger step size differs from the original ADMM scheme (1.4) only in the fact that the step size for updating the Lagrange multiplier can be even larger than 1. Numerically, this larger step size is crucial for leading to better numerical performance, see some numerical results in [13] and later in [3, 5, 19, 25] for some new applications and numerical verification.

As analyzed in [8, 12], the ADMM scheme (1.4) is an application of the Douglas-Rachford splitting method (DRSM) in [6, 22] to the dual of (1.1). If the Peaceman-Rachford splitting method (PRSM), which was originally proposed in [23, 22] and is equally important as the DRSM in PDE literature, to the dual of (1.1), it was shown in [15] that we can obtain the following scheme:

$$\begin{cases} x^{k+1} = \arg \min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.6a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & (1.6b) \\ y^{k+1} = \arg \min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathcal{Y}\}, & (1.6c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b). & (1.6d) \end{cases}$$

In the context of (1.1), the applications of PRSM and DRSM (a.k.a. (1.6) and (1.4), respectively) are very similar, with the only difference of the Lagrange multiplier being updated twice at each iteration in (1.6). Alternatively, the scheme (1.6) can be regarded as an 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. One explanation can be found in [18], showing that that the sequence generated by the symmetric ADMM (1.6) 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). Counter examples showing the divergence of (1.6) can also be found in [4, 7]. To overcome the difficulty of the lack of strict contraction, in [18] we suggested updating the Lagrange multiplier in (1.6) more conservatively and obtained the symmetric ADMM scheme with shorter step sizes:

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

where the parameter $\alpha \in (0, 1)$ is for shrinking the step sizes in (1.6). It was shown in [18] that the sequence generated by (1.7) is strictly contractive with respect to the solution set of (1.1). Thus, in [18] we also called the symmetric ADMM (1.7) the strictly contractive PRSM. It worths to mention that the restriction $\alpha \in (0, 1)$ makes the update of Lagrange multipliers more conservative with shorter step sizes; but it plays a crucial theoretical role in ensuring the strict contraction for the sequence generated by (1.7) and hence sufficiently ensuring the convergence of the symmetric ADMM (1.7). Despite of the significant theoretical role of the shorter step sizes in (1.7), it was recommended in [18] to take larger values close to 1 for α to lead to better numerical performance. Naturally, one may ask if it is necessary to shrink both the step sizes for the symmetric ADMM (1.7) in order to ensure its convergence.

We will provide a negative answer to this question and show that we can take the Glowinski's larger step size in [13] to update one of the Lagrange-multiplier-updating steps in (1.7) while the convergence can still be ensured. Furthermore, we will specify a step size domain that can ensure the convergence of the symmetric ADMM (1.7). To emphasize that we can take different step sizes for updating the Lagrange multiplier, we revise the symmetric ADMM (1.7) slightly as following

$$\begin{cases} x^{k+1} = \arg \min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.8a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & (1.8b) \\ y^{k+1} = \arg \min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathcal{Y}\}, & (1.8c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(Ax^{k+1} + By^{k+1} - b), & (1.8d) \end{cases}$$

in which r and s play the same role of α as in (1.7) but they are independent. We discuss the case where the parameters r and s are restricted into the domain

$$\mathcal{D} = \{(s, r) \mid s \in (0, \frac{1+\sqrt{5}}{2}), r \in (-1, 1) \ \& \ r + s > 0, |r| < 1 + s - s^2\}. \quad (1.9)$$

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 examples, the original ADMM (1.4) corresponds to the point $(s = 1, r = 0)$ in Figure 1, the ADMM with Glowinski's larger step size (1.5) is the case where $s \in (0, \frac{1+\sqrt{5}}{2})$ and $r = 0$, the generalized ADMM scheme proposed in [8] is the case where $s = 1$ and $r \in (-1, 1)$ (see Remark 5.6 for elaboration); and the symmetric ADMM in [18] 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 (1.8). 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 (1.8) 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. Finally some conclusions are made in Section 7.

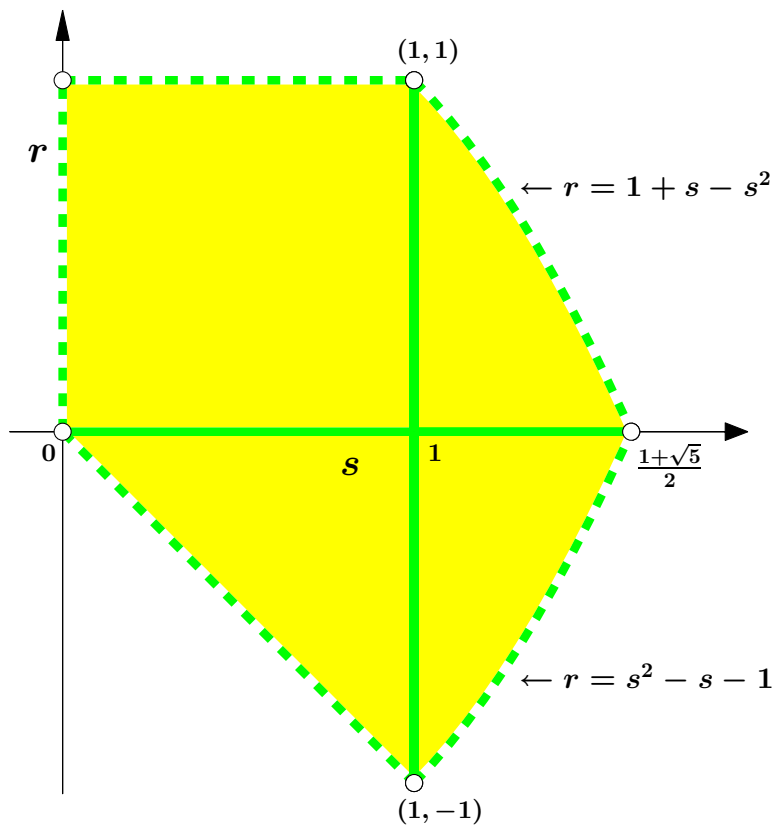


Fig. 1. Step size domain \mathcal{D} for the symmetric ADMM (1.8)

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.

Let the Lagrangian function of the problem (1.1) be

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

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

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

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

$$\begin{cases} x^* &= \arg \min\{L(x, y^*, \lambda^*) | x \in \mathcal{X}\}, \\ y^* &= \arg \min\{L(x^*, y, \lambda^*) | y \in \mathcal{Y}\}, \\ \lambda^* &= \arg \max\{L(x^*, y^*, \lambda) | \lambda \in \mathfrak{R}^m\}, \end{cases} \quad (2.2)$$

which can be rewritten as

$$\begin{cases} x^* \in \mathcal{X}, & L(x, y^*, \lambda^*) - L(x^*, y^*, \lambda^*) \geq 0, \quad \forall x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & L(x^*, y, \lambda^*) - L(x^*, y^*, \lambda^*) \geq 0, \quad \forall y \in \mathcal{Y}, \\ \lambda^* \in \mathfrak{R}^m, & L(x^*, y^*, \lambda^*) - L(x^*, y^*, \lambda) \geq 0, \quad \forall \lambda \in \mathfrak{R}^m. \end{cases}$$

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

Proposition 2.1. Let $\mathcal{X} \subset \mathfrak{R}^n$ be a closed convex set, $\theta(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}$ and $f(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}$ 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,

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

if and only if

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

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

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

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

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

where

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

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

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

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 (1.8) 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 (1.8) from a given iterate (y^k, λ^k) . If*

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

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 to show that

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

First, it follows from (2.6) that

$$Ax^{k+1} + By^k - b = 0 \quad \text{and} \quad Ax^{k+1} + By^{k+1} - b = 0. \quad (2.8)$$

Consequently, together with (1.8b) and (1.8d), we get

$$\lambda^{k+\frac{1}{2}} = \lambda^k \quad \text{and} \quad \lambda^{k+1} = \lambda^k. \quad (2.9)$$

On the other hand, using (2.3), the optimality conditions of the subproblems (1.8a) and (1.8c) are

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{-A^T \lambda^k + \beta A^T (Ax^{k+1} + By^k - b)\} \geq 0, \quad \forall x \in \mathcal{X} \quad (2.10a)$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^{k+1} + By^{k+1} - b)\} \geq 0, \quad \forall y \in \mathcal{Y}, \quad (2.10b)$$

respectively. Substituting (2.8) in (2.10), it 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. □

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 (1.8), in which $\epsilon > 0$ is the tolerance specified by the user.

2.3 Some notation

Like the original ADMM scheme (1.4), in (1.8) the variable x also plays an intermediate role in the sense that x^k is not involved in the iteration. Thus, as [1], 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 (1.8), 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 (1.8). Note that this is only for the convenience of algebraically presenting the convergence proof with compact notations.

First, for the iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ generated by the symmetric ADMM (1.8), we define an auxiliary vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as

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

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 (1.8) and \tilde{w}^k be defined by (3.1). Then, we have*

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

where

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

Proof. Using the notation defined in (3.1), we can rewrite $\lambda^{k+\frac{1}{2}}$ in (1.8b) as

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

Notice that the objective functions of the x - and y -subproblem in (1.8) are

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

and

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

respectively. According to (2.3), the optimality condition of the x -subproblem of (1.8a) is

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

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

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

Ignoring some constant terms in the objective function of the y -subproblem (1.8c) and using (3.3), we have

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

Consequently, using (2.3), we obtain

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

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

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

Thus, using the notation of w^k and \tilde{w}^k , the optimality condition of the y -subproblem can be written as

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

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

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

Combining (3.5a), (3.5b) and (3.5c), and using the notations of (2.5), we prove the assertion of this lemma. \square

Lemma 3.2. For given v^k , let w^{k+1} be generated by (1.8) and \tilde{w}^k be defined by (3.1). Then, we have

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

where

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

Proof. It follows from (1.8) that

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

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. \square

Therefore, based on the previous two lemmas, the scheme (1.8) 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 Preparation for convergence proof

As mentioned, our main purpose is proving the convergence for the symmetric ADMM (1.8) with the parameters r and s restricted into the domain \mathcal{D} defined in (1.9). Since the proof is not trivial, in this section we prove some theorems first so as to prepare for the convergence proof for (1.8).

First, 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 (3.2a) as

$$(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^T H(v^k - v^{k+1}). \quad (4.1)$$

For this purpose, we define

$$H = QM^{-1} \quad (4.2)$$

and prove a simple fact for the matrix H .

Lemma 4.1. The matrix H defined in (4.2) 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 (4.2) and (3.2b) that

$$\begin{aligned} H &= QM^{-1} = \begin{pmatrix} \beta B^T B & -rB^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{s}{r+s} \beta B & \frac{1}{r+s} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1 - \frac{rs}{r+s}) \beta B^T B & -\frac{r}{r+s} B^T \\ -\frac{r}{r+s} B & \frac{1}{(r+s)\beta} I_m \end{pmatrix}. \end{aligned} \quad (4.3)$$

For any $(s, r) \in \mathcal{D}$, we have $r + s > 0$. From (4.3), we know that

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

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

$$r + s - rs = \begin{cases} \text{for } r \in [0, 1) \text{ and } s > 0 \\ \text{for } r \in (-1, 0), s > 0 \text{ and } s + r > 0 \end{cases} = \begin{cases} r + s(1 - r) \\ (r + s) - rs \end{cases} > 0,$$

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. \square

For the case where H is positive definite, we have $HM = Q$. Thus, substituting (4.1) into (3.2a), we obtain

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega. \quad (4.5)$$

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 (4.5) with

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

we obtain

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2} (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \quad (4.6)$$

The following theorem is the basis for proving the strict contraction property for the sequence generated by the scheme (1.8). It is also useful for estimating the convergence rate for the sequence generated by (1.8).

Theorem 4.2. *For the sequence $\{w^k\}$ generated by the symmetric ADMM (1.8), we have*

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ & \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2} \|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega, \end{aligned} \quad (4.7)$$

where H is defined in (4.2) and

$$G = Q^T + Q - M^T H M. \quad (4.8)$$

Proof. Substituting (4.6) into the right-hand side of (4.5), we obtain

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ & \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2} (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2), \quad \forall w \in \Omega, \end{aligned} \quad (4.9)$$

For the last term of the right-hand side of (4.9), we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & = \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \stackrel{(3.6a)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ & = 2(v^k - \tilde{v}^k)^T H M (v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T H M (v^k - \tilde{v}^k) \\ & \stackrel{(4.2)}{=} (v^k - \tilde{v}^k)^T (Q^T + Q - M^T H M) (v^k - \tilde{v}^k). \end{aligned} \quad (4.10)$$

Substituting this equation into (4.9) and using the definition of the matrix G in (4.8), the assertion of this theorem is proved. \square

The next theorem clearly shows the difficulty of proving the convergence for (1.8). The significant difference in convergence proof between the general symmetric ADMM scheme (1.8) and the special symmetric ADMM scheme (1.7) with shrank step sizes is also reflected in this theorem.

Theorem 4.3. *For the sequence $\{w^k\}$ generated by the symmetric ADMM (1.8), we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \quad (4.11)$$

Proof. Setting $w = w^*$ in (4.7), we get

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \quad (4.12)$$

Then, using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0$$

and thus

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

The assertion (4.11) follows directly. \square

Obviously, if the matrix G defined in (4.8) is guaranteed to be positive definite, then the inequality (4.11) 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 (1.8) can be easily established. Now, let us investigate the positive definiteness for the matrix G . Since $HM = Q$ (see (4.2)), we have $M^T HM = M^T Q$. Note that

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

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

$$\begin{aligned} G &= (Q^T + Q) - M^T HM \\ &= \begin{pmatrix} 2\beta B^T B & -(1+r)B^T \\ -(1+r)B & \frac{2}{\beta}I_m \end{pmatrix} - \begin{pmatrix} (1+s)\beta B^T B & -(r+s)B^T \\ -(r+s)B & \frac{1}{\beta}(r+s)I_m \end{pmatrix} \\ &= \begin{pmatrix} (1-s)\beta B^T B & -(1-s)B^T \\ -(1-s)B & \frac{1}{\beta}(2-(r+s))I_m \end{pmatrix}. \end{aligned} \quad (4.13)$$

Note that

$$G = \begin{pmatrix} B^T & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \beta(1-s)I & -(1-s)I \\ -(1-s)I & \frac{1}{\beta}(2-(r+s))I \end{pmatrix} \begin{pmatrix} B & 0 \\ 0 & I \end{pmatrix}. \quad (4.14)$$

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

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

is positive definite. Indeed, under the assumption $r \in [0, 1)$, to ensure

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

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 (1.7) with shrink step sizes is convergent, as proved in [18]. However, for $(s, r) \in \mathcal{D}$ defined in (1.9) 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 (1.8). In the following, we split the discussion for the convergence of (1.8) into five scenarios:

$$\left\{ \begin{array}{l} \mathcal{D}_1 = \{(s, r) \mid s \in (0, 1), r \in (-1, 1), r + s > 0\}, \\ \mathcal{D}_2 = \{(s, r) \mid s = 1, r \in (-1, 1)\}, \\ \mathcal{D}_3 = \{(s, r) \mid s \in (1, \frac{1+\sqrt{5}}{2}), r = 0\}, \\ \mathcal{D}_4 = \{(s, r) \mid s \in (1, \frac{1+\sqrt{5}}{2}), r \in (0, 1) \ \& \ r < 1 + s - s^2\}, \\ \mathcal{D}_5 = \{(s, r) \mid s \in (1, \frac{1+\sqrt{5}}{2}), r \in (-1, 0) \ \& \ -r < 1 + s - s^2\}. \end{array} \right. \quad (4.15)$$

These five sub-domains are displayed separately in Figure 2. Figure 1 is re-displayed in Figure 3, in which the composition of these sub-domains defined in (4.15) is clearly shown. It is clear (see Figure 1) that

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

and

$$\mathcal{D}_i \cap \mathcal{D}_j = \emptyset, \quad \forall i, j \in \{1, 2, 3, 4, 5\}, \quad i \neq j.$$

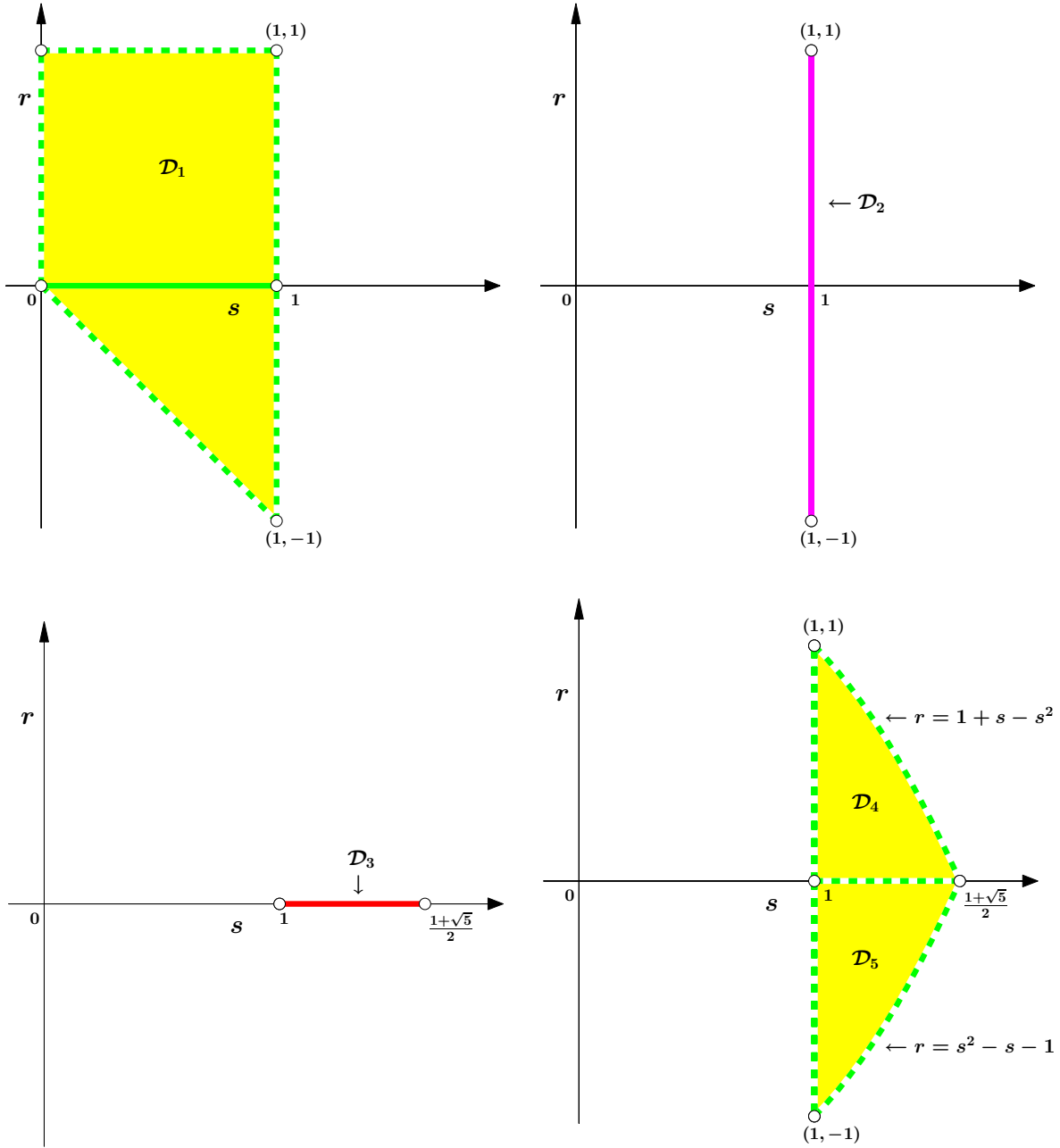


Figure 2. Sub-domains \mathcal{D}_1 , \mathcal{D}_2 , \mathcal{D}_3 , \mathcal{D}_4 and \mathcal{D}_5

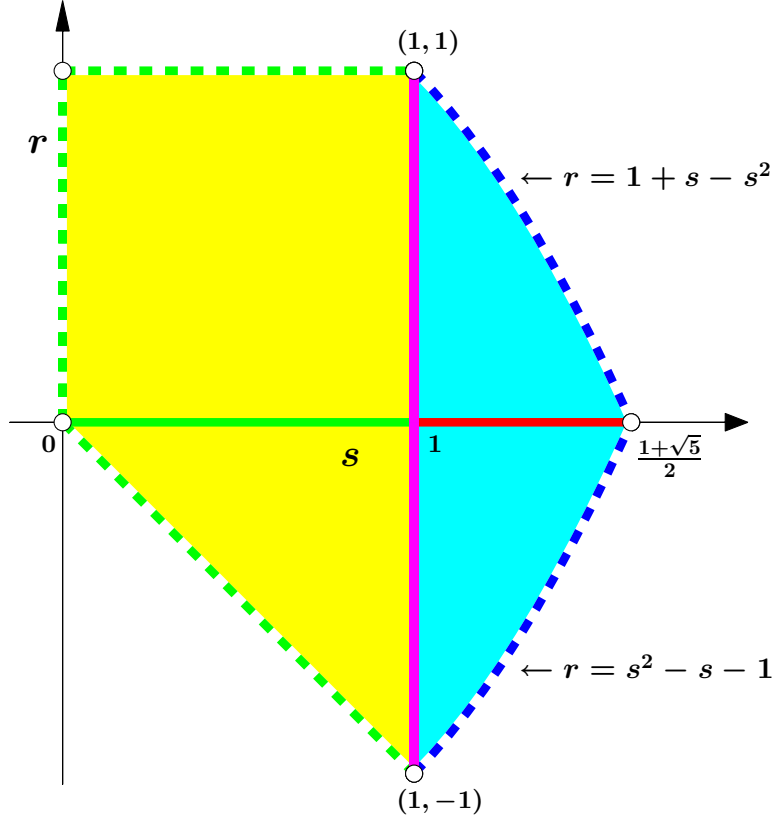


Figure 3. $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$

For showing the main theorem in the next section, we need to deal with the term $\|v^k - \tilde{v}^k\|_G^2$ beforehand.

Lemma 4.4. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.8) and \tilde{w}^k be defined by (3.1). Then we have*

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 &= (1-r)\beta\|B(y^k - y^{k+1})\|^2 + (2-r-s)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &\quad + 2(1-r)\beta(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}). \end{aligned} \quad (4.16)$$

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 (4.13)), $v = \begin{pmatrix} y \\ \lambda \end{pmatrix}$ and $\tilde{y}^k = y^{k+1}$, we have

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 &= (1-s)\beta\|B(y^k - y^{k+1})\|^2 - 2(1-s)(\lambda^k - \tilde{\lambda}^k)^T B(y^k - y^{k+1}) \\ &\quad + \frac{2-(r+s)}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2. \end{aligned} \quad (4.17)$$

Notice that (see (3.1))

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

Thus, we have

$$\begin{aligned}
& (\lambda^k - \tilde{\lambda}^k)^T B(y^k - y^{k+1}) \\
&= \beta((Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1}))^T B(y^k - y^{k+1}) \\
&= \beta(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) + \beta\|B(y^k - y^{k+1})\|^2,
\end{aligned} \tag{4.18}$$

and

$$\begin{aligned}
\|\lambda^k - \tilde{\lambda}^k\|^2 &= \beta^2\|(Ax^{k+1} + By^{k+1} - b) + B(y^k - y^{k+1})\|^2 \\
&= \beta^2\|Ax^{k+1} + By^{k+1} - b\|^2 + 2\beta^2(Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \\
&\quad + \beta^2\|B(y^k - y^{k+1})\|^2.
\end{aligned} \tag{4.19}$$

Substituting (4.18) and (4.19) into the right-hand side of (4.17), we obtain (4.16) immediately. \square

In the following, we treat the crossing term in the right-hand side of (4.16), namely,

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

and find a lower bound for it.

Lemma 4.5. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.8). Then we have*

$$\begin{aligned}
& (Ax^{k+1} + By^{k+1} - b)^T B(y^k - y^{k+1}) \\
&\geq \frac{1-s}{1+r}(Ax^k + By^k - b)^T B(y^k - y^{k+1}) - \frac{r}{1+r}\|B(y^k - y^{k+1})\|^2.
\end{aligned} \tag{4.20}$$

Proof. By ignoring the constant terms, the y -subproblem in (1.8) can be written as

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

Thus, according to (3.4b), the optimality condition of the y -subproblem (1.8c) is

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+\frac{1}{2}} + \beta B^T (Ax^{k+1} + By^{k+1} - b)\} \geq 0, \quad \forall y \in \mathcal{Y}. \tag{4.21}$$

Analogously, for the previous iteration, we have

$$y^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^{k-\frac{1}{2}} + \beta B^T (Ax^k + By^k - b)\} \geq 0, \quad \forall y \in \mathcal{Y}. \tag{4.22}$$

Setting $y = y^k$ and $y = y^{k+1}$ in (4.21) and (4.22), respectively, and then adding them, we get

$$(y^k - y^{k+1})^T B^T \{(\lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}}) - \beta(Ax^k + By^k - b) + \beta(Ax^{k+1} + By^{k+1} - b)\} \geq 0. \tag{4.23}$$

Note that in the k -th iteration (see (1.8b)), we have

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \tag{4.24}$$

and in the previous iteration (see (1.8d)),

$$\lambda^k = \lambda^{k-\frac{1}{2}} - s\beta(Ax^k + By^k - b). \quad (4.25)$$

It follows from (4.25) and (4.24) that

$$\begin{aligned} \lambda^{k-\frac{1}{2}} - \lambda^{k+\frac{1}{2}} &= r\beta(Ax^{k+1} + By^k - b) + s\beta(Ax^k + By^k - b) \\ &= r\beta(Ax^{k+1} + By^{k+1} - b) + r\beta B(y^k - y^{k+1}) + s\beta(Ax^k + By^k - b). \end{aligned} \quad (4.26)$$

Substituting (4.26) into (4.23) and with a simple manipulation, we have

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

and obtain (4.20) directly from the last inequality. This lemma is proved. \square

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

Theorem 4.6. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.8) and \tilde{w}^k be defined by (3.1). Then we have*

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 &\geq \frac{(1-r)^2}{1+r}\beta\|B(y^k - y^{k+1})\|^2 + (2-r-s)\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ &\quad + \frac{2(1-r)(1-s)}{1+r}\beta(Ax^k + By^k - b)^T B(y^k - y^{k+1}). \end{aligned} \quad (4.27)$$

Proof. Substituting (4.20) into (4.16), we obtain the assertion (4.27) immediately. \square

5 Main Theorem

In this section, we prove the main theorem for proving the convergence and estimating the convergence rate for the symmetric ADMM (1.8).

Theorem 5.1. *For the symmetric ADMM (1.8) with $(s, r) \in \mathcal{D}$, we have*

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

$$\|v^k - \tilde{v}^k\|_G^2 \geq C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \quad (5.1a)$$

1. 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

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 &\geq C_0\beta(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2) \\ &\quad + C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned} \quad (5.1b)$$

Below, we prove Theorem 5.1 for the sub-domains defined in (4.15) one by one.

5.1 (s, r) in the sub-domain \mathcal{D}_1

Lemma 5.2. *For arbitrarily fixed $(s, r) \in \mathcal{D}_1$, there are constants $C_1, C_2 > 0$ such that the inequality (5.1a) holds and Theorem 5.1 is true for $(s, r) \in \mathcal{D}_1$.*

Proof. Recall that $\mathcal{D}_1 = \{(s, r) \mid s \in (0, 1), r \in (-1, 1), r + s > 0\}$ (see (4.15)). According to (4.14), we have

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

where

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

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

$$\|v^k - \tilde{v}^k\|_G^2 = \left\| \begin{pmatrix} B(y^k - \tilde{y}^k) \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \right\|_{G_0}^2 = \left\| \begin{pmatrix} B(y^k - y^{k+1}) \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix} \right\|_{\tilde{G}_0}^2, \quad (5.2)$$

where

$$\tilde{G}_0 = L^T G_0 L \quad \text{and} \quad L = \begin{pmatrix} I & 0 \\ \beta I & \beta I \end{pmatrix}.$$

Due to the positive definiteness of G_0 and non-singularity of L , \tilde{G}_0 is positive definite. Finally, the assertion follows from (5.2) and the positive definiteness of \tilde{G}_0 . \square

Remark 5.3. The symmetric ADMM (1.7) with shrunk step sizes is a special case of (1.8) with $r = s \in (0, 1)$. In this case (see (4.4)),

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

and (see (4.14))

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

5.2 (s, r) in the sub-domain \mathcal{D}_2

Theorem 4.6 is crucial for proving the convergence of (1.8) with $s \geq 1$. Now, we use it to show that Theorem 5.1 is true for $(s, r) \in \mathcal{D}_2$.

Lemma 5.4. *For any $(s, r) \in \mathcal{D}_2$, the inequality (5.1a) holds with*

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

Proof. Notice that $\mathcal{D}_2 = \{(s, r) \mid s = 1, r \in (-1, 1)\}$ (see (4.15)). Setting $s = 1$ in (4.27), we obtain

$$\|v^k - \tilde{v}^k\|_G^2 \geq \frac{(1-r)^2}{1+r} \beta \|B(y^k - y^{k+1})\|^2 + (1-r) \beta \|Ax^{k+1} + By^{k+1} - b\|^2. \quad (5.3)$$

The lemma is proved. \square

Remark 5.5. Recall that the original ADMM (1.4) is a special case of (1.8) with $r = 0$ and $s = 1$. In this special case, the matrix H has the form (see (4.3))

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \quad (5.4)$$

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 (5.3), 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 (5.4), we get

$$\|v^k - \tilde{v}^k\|_G^2 \geq \|v^k - v^{k+1}\|_H^2. \quad (5.5)$$

Remark 5.6. Indeed, the generalized ADMM in [8] is a special case of (1.8) with $(s, r) \in \mathcal{D}_2$. Let us elaborate on the detail, which is not straightforward. In [8], only the special case of (1.1) with $B = -I$ and $b = 0$ was discussed. But it is trivial (see, e.g., [9, 10, 11]) to extend the generalized ADMM in [8] to the model (1.1). The resulting iterative scheme can be written as

$$\begin{cases} x^{k+1} = \arg \min \{ \theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X} \}, & (5.6a) \\ y^{k+1} = \arg \min \{ \theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2} \|[\alpha Ax^{k+1} - (1-\alpha)(By^k - b)] + By - b\|^2 \mid y \in \mathcal{Y} \}, & (5.6b) \\ \lambda^{k+1} = \lambda^k - \beta \{ [\alpha Ax^{k+1} - (1-\alpha)(By^k - b)] + By^{k+1} - b \}, & (5.6c) \end{cases}$$

where the parameter $\alpha \in (0, 2)$ is a relaxation factor. To see why (5.6) is a special case of (1.8), let us choose $r = \alpha - 1$. Then, the y -subproblem in (5.6b) can be written as

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

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

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

Recall (1.8b). We have

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), \quad (5.7)$$

which enables us to rewrite the y -subproblem as

$$y^{k+1} = \arg \min \{ \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y} \}. \quad (5.8)$$

Analogously, since $\alpha = 1 + r$, the step (5.6c) 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 (5.7), yields

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

Combining (5.6a), (5.7), (5.8) and (5.9), we now can rewrite the generalized ADMM (5.6) as

$$\begin{cases} x^{k+1} = \arg \min \{ \theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X} \}, & (5.10a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(Ax^{k+1} + By^k - b), & (5.10b) \\ y^{k+1} = \arg \min \{ \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y} \}, & (5.10c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), & (5.10d) \end{cases}$$

where $r \in (-1, 1)$. Thus, the generalized ADMM (5.6) is a special case of the symmetric ADMM scheme (1.8) with the restriction $(s, r) \in \mathcal{D}_2$.

5.3 (s, r) in the sub-domain \mathcal{D}_3

Now, we turn to verify Theorem 5.1 for the sub-domain \mathcal{D}_3 . First, let us define

$$S = 1 + s - s^2. \quad (5.11a)$$

Because

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

we have

$$S > 0, \quad \forall s \in \left(1, \frac{1 + \sqrt{5}}{2}\right). \quad (5.11b)$$

The result in Theorem 4.6 is the basis for the proof in this section.

Recall that $\mathcal{D}_3 = \{(s, r) \mid s \in (1, \frac{1+\sqrt{5}}{2}), r = 0\}$ (see (4.15)). Since $r = 0$, it follows from Theorem 4.6 that

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

Now, we define a constant T_1 as

$$T_1 := \frac{1}{3}(s^2 - s + 5), \quad (5.13)$$

and thus we have

$$T_1 - s = \frac{1}{3}(s^2 - 4s + 5) = \frac{1}{3}[(s-2)^2 + 1] > \frac{1}{3}. \quad (5.14)$$

Using the Cauchy-Schwarz inequality to the crossing term in the right-hand side of (5.12) and because of $T_1 - s > 0$, we have

$$\begin{aligned} & 2(1-s)\beta(Ax^k + By^k - b)^T B(y^k - y^{k+1}) \\ & \geq -(T_1 - s)\beta\|Ax^k + By^k - b\|^2 - \frac{(1-s)^2}{T_1 - s}\beta\|B(y^k - y^{k+1})\|^2. \end{aligned} \quad (5.15)$$

Lemma 5.7. *For any $(s, r) \in \mathcal{D}_3$, the inequality (5.1b) holds with*

$$C_0 = T_1 - s > \frac{1}{3}, \quad C_1 \geq S > 0 \quad \text{and} \quad C_2 = \frac{1}{3}S > 0.$$

Proof. Substituting (5.15) into (5.12), we obtain

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 & \geq (T_1 - s)\beta(\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2) \\ & \quad + \left(1 - \frac{(1-s)^2}{T_1 - s}\right)\beta\|B(y^k - y^{k+1})\|^2 + (2 - T_1)\beta\|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned} \quad (5.16)$$

According to (5.1b), we set

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

It follows from (5.14) that $C_0 \geq \frac{1}{3}$. Then, using (5.13), 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 $s \in (1, \frac{1+\sqrt{5}}{2})$, it follows that

$$C_1 = \frac{2(1+s-s^2)}{1+(s-2)^2} \geq 1 + s - s^2 = S. \quad (5.17)$$

Finally,

$$C_2 = 2 - T_1 = 2 - \frac{1}{3}(s^2 - s + 5) = \frac{1}{3}(1 + s - s^2) = \frac{1}{3}S. \quad (5.18)$$

The assertion of this lemma is proved \square

5.4 (s, r) in the sub-domain \mathcal{D}_4

Recall that $\mathcal{D}_4 = \{(s, r) \mid s \in (1, \frac{1+\sqrt{5}}{2}), r \in (0, 1) \ \& \ r < 1 + s - s^2\}$ (see (4.15)). We say $r < 1 + s - s^2$ is an additional restriction in \mathcal{D}_4 and define

$$T_2 = r + s + (1-s)^2. \quad (5.19)$$

Notice that $T_2 - (r + s) = (1 - s)^2 > 0$ for all $s \in (1, \frac{1+\sqrt{5}}{2})$. For the crossing term in the right-hand side of (4.27), using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \frac{2(1-r)(1-s)}{1+r} \beta (Ax^k + By^k - b)^T B(y^k - y^{k+1}) \\ & \geq -[T_2 - (r + s)] \beta \|Ax^k + By^k - b\|^2 - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_2 - (r + s)]} \beta \|B(y^k - y^{k+1})\|^2. \end{aligned} \quad (5.20)$$

Lemma 5.8. *For any $(s, r) \in \mathcal{D}_4$, the inequality (5.1b) holds with*

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

where T_2 is defined in (5.19).

Proof. Substituting (5.23) into (4.27) yields

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 & \geq [T_2 - (r + s)] \beta (\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2) \\ & \quad + \left(\frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_2 - (r + s)]} \right) \beta \|B(y^k - y^{k+1})\|^2 \\ & \quad + (2 - T_2) \beta \|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned} \quad (5.21)$$

According to (5.1b), we set

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

In the following, we show that $C_0, C_1, C_2 > 0$. First, it follows from (5.19) that $C_0 = (1 - s)^2 > 0$ for all $s \in (1, \frac{1+\sqrt{5}}{2})$. Indeed, using (5.19), 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 (5.19), the left-hand side of the last inequality is T_2 and thus $C_2 = 2 - T_2 > 0$. The lemma is proved. \square

5.5 (s, r) in the sub-domain \mathcal{D}_5

Recall that $\mathcal{D}_5 = \{(s, r) \mid s \in (1, \frac{1+\sqrt{5}}{2}), r \in (-1, 0) \text{ \& } -r < 1 + s - s^2\}$ (see (4.15)). We say $s^2 - s - 1 < r$ is an additional restriction in \mathcal{D}_5 and define

$$T_3 = r + s + \frac{(s^2 - s)(2 - s)}{1 + r}. \quad (5.22)$$

Notice that $T_3 - (r + s) = \frac{(s^2 - s)(2 - s)}{1 + r} > 0$ for all $s \in (1, \frac{1 + \sqrt{5}}{2})$ and $r \in (-1, 0)$. For the crossing term in the right-hand side of (4.27), using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \frac{2(1-r)(1-s)}{1+r} \beta (Ax^k + By^k - b)^T B(y^k - y^{k+1}) \\ & \geq -[T_3 - (r + s)] \beta \|Ax^k + By^k - b\|^2 - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_3 - (r + s)]} \beta \|B(y^k - y^{k+1})\|^2. \end{aligned} \quad (5.23)$$

Lemma 5.9. *For any $(s, r) \in \mathcal{D}_5$, the inequality (5.1b) holds with*

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

where T_3 is defined in (5.22).

Proof. Substituting (5.23) into (4.27) yields

$$\begin{aligned} \|v^k - \tilde{v}^k\|_G^2 & \geq [T_3 - (r + s)] \beta (\|Ax^{k+1} + By^{k+1} - b\|^2 - \|Ax^k + By^k - b\|^2) \\ & \quad + \left(\frac{(1-r)^2}{1+r} - \frac{(1-r)^2(1-s)^2}{(1+r)^2[T_3 - (r + s)]} \right) \beta \|B(y^k - y^{k+1})\|^2 \\ & \quad + (2 - T_3) \beta \|Ax^{k+1} + By^{k+1} - b\|^2. \end{aligned} \quad (5.24)$$

According to (5.1b), we set

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

In the following, we show that $C_0, C_1, C_2 > 0$. First, it follows from (5.22) that $C_0 = \frac{(s^2 - s)(2 - s)}{r + 1} > 0$ for all $s \in (1, \frac{1 + \sqrt{5}}{2})$ and $r \in (-1, 0)$. Indeed, using (5.22), we have

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

Since $r \in (-1, 0)$ and $s \in (1, \frac{1 + \sqrt{5}}{2})$ in \mathcal{D}_5 and $1 + s - s^2 > 0$ (see (5.11)), 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 $s \in (1, \frac{1 + \sqrt{5}}{2})$, we also have $s^2 - s > 0$ and thus

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

Substituting the last inequality into (5.22), we get

$$T_3 = r + s + \left(\frac{s^2 - s}{1 + r} \right) (2 - s) < r + s + (2 - s) = 2 + r, \quad (5.25)$$

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

Since we have proved Lemma 5.2, Lemma 5.4, Lemma 5.7, Lemma 5.8 and Lemma 5.9, the proof for Theorem 5.1 is complete. In the next section, we will use Theorem 4.2 and Theorem 5.1 to prove the convergence and convergence rate.

6 Convergence analysis

In this section, we show the global convergence and estimate the convergence rate in terms of the iteration complexity for the symmetric ADMM (1.8). Theorems 4.2 and 5.1 will be used.

6.1 Global convergence

We summarize the global convergence of (1.8) in the following theorem

Theorem 6.1. *For the sequence $\{w^k\}$ generated by the symmetric ADMM (1.8), we have*

$$\lim_{k \rightarrow \infty} (\|B(y^k - y^{k+1})\|^2 + \|Ax^{k+1} + By^{k+1} - b\|^2) = 0. \quad (6.1)$$

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. First, for $(s, r) \in \mathcal{D}_1 \cup \mathcal{D}_2$, according to Theorems 4.3 and 5.1, there are constants $C_1, C_2 > 0$, such that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - (C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2), \quad \forall v^* \in \mathcal{V}^*. \quad (6.2)$$

It follows from the above inequality that

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

and thus we obtain the assertion (6.1). Analogously, for $(s, r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$, it follows from Theorems 4.3 and 5.1 that

$$\begin{aligned} & \|v^{k+1} - v^*\|_H^2 + C_0\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \\ & \leq (\|v^k - v^*\|_H^2 + C_0\beta\|Ax^k + By^k - b\|^2) \\ & \quad - (C_1\beta\|B(y^k - y^{k+1})\|^2 + C_2\beta\|Ax^{k+1} + By^{k+1} - b\|^2), \quad \forall v^* \in \mathcal{V}^*. \end{aligned} \quad (6.3)$$

Consequently (note that x^1 is the computational result of (y^0, λ^0)), we have

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

and the assertion (6.1) is true for $(s, r) \in \mathcal{D}_3 \cup \mathcal{D}_4 \cup \mathcal{D}_5$. The convergence of (1.8) is thus proved in sense of (6.1).

Furthermore, from (6.2) and (6.3) we know 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}^∞ be induced by (1.8a) with given $(y^\infty, \lambda^\infty)$. Recall the matrix B is assumed to be full column rank. Due to (6.1), we have

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

Then, it follows from (3.2a) that

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

and thus $\tilde{w}^\infty = w^\infty$ is a solution point of (2.5). On the other hand, because (6.3) (Resp. (6.2)) is true for any solution point of (2.5) and $w^\infty \in \Omega^*$, we have

$$\|v^{k+1} - v^\infty\|_H^2 + C_0\beta\|Ax^{k+1} + By^{k+1} - b\|^2 \leq \|v^k - v^\infty\|_H^2 + C_0\beta\|Ax^k + By^k - b\|^2. \quad (6.5)$$

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

6.2 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 [17] (Theorem 2.3.5) or [20] (Theorem 2.1).

Theorem 6.2. *The solution set of VI(Ω, F, θ) is convex and it can be characterized as*

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

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

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

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

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

Before we present the analysis for estimating the convergence rate for (1.8), we notice that it follows from the monotonicity of F that

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

Thus, substituting this inequality into (4.7), we obtain

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ & \geq \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega, \end{aligned} \quad (6.8)$$

which, together with (5.1), enables us to derive the assertions in the following lemma.

Lemma 6.3. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.8) 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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2} \|v - v^k\|_H^2 \geq \frac{1}{2} \|v - v^{k+1}\|_H^2, \quad \forall w \in \Omega. \quad (6.9a)$$

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

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \left(\frac{1}{2} \|v - v^k\|_H^2 + C_0 \beta \|Ax^k + By^k - b\|^2 \right) \\ & \geq \left(\frac{1}{2} \|v - v^{k+1}\|_H^2 + C_0 \beta \|Ax^{k+1} + By^{k+1} - b\|^2 \right), \quad \forall w \in \Omega. \end{aligned} \quad (6.9b)$$

Theorem 6.4. *Let the sequence $\{w^k\}$ be generated by the symmetric ADMM (1.8) and \tilde{w}^k be defined by (3.1). Then, for $(s, r) \in \mathcal{D}_1 \cup \mathcal{D}_2$ and any integer number $t > 0$, we have*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega, \quad (6.10a)$$

where

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^t \tilde{w}^k. \quad (6.10b)$$

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

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2t} (\|v - v^1\|_H^2 + 2C_0 \|Ax^1 + By^1 - b\|^2) \quad \forall w \in \Omega, \quad (6.11a)$$

where

$$\tilde{w}_t = \frac{1}{t} \sum_{k=1}^t \tilde{w}^k. \quad (6.11b)$$

Proof. First, summarizing the inequalities (6.9a) over $k = 0, 1, \dots, t$, we obtain

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

Then, using the notation of \tilde{w}_t in (6.10b), the last inequality can be written as

$$\frac{1}{t+1} \sum_{k=0}^t \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \quad (6.12)$$

It follows from the definition of \tilde{w}_t in (6.10b) that

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

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

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

Substituting it into (6.12), the assertion (6.10) follows directly. The proof for the assertion (6.11) is similar and omitted. \square

It follows from (6.7), (6.10) and (6.11) that the symmetric ADMM (1.8) is able to generate an approximate solution point (i.e., \tilde{w}_t defined in (6.10b) or (6.11b)) 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 (1.8).

7 Conclusions

In this paper, we comprehensively investigate how to choose step sizes to update the Lagrange multiplier for the symmetric version of the alternating direction method of multipliers (ADMM) in the convex programming context, which can also be called the strictly contractive Peaceman-Rachford splitting method. Theoretically, smaller step sizes usually offer more analytical convenience such as the possibility of deriving convergence more easily; while numerically it is a common target to pursue larger step sizes even though they may cause more difficulties in theoretical analysis. Thus, it is interesting to find out if it is possible to use larger step sizes for the symmetric ADMM while simultaneously ensuring its convergence. We identify a step size domain that can theoretically ensure the convergence but allows to use Glowinski's larger step size for one Lagrange-multiplier-updating step at each iteration. The step size domain includes some ADMM-like schemes in the literature as special cases; thus our discussion extends some known convergence results in the ADMM literature and provides a comprehensive framework to study the convergence of ADMM-like schemes.

Because of the equal role of the parameters r and s in the symmetric ADMM scheme (1.8), it is clear that the scheme (1.8) can be rewritten as

$$\begin{cases} y^{k+1} = \arg \min\{\mathcal{L}_\beta(x^k, y, \lambda^k) \mid y \in \mathcal{Y}\}, & (7.13a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - s\beta(Ax^k + By^{k+1} - b), & (7.13b) \\ x^{k+1} = \arg \min\{\mathcal{L}_\beta(x, y^{k+1}, \lambda^{k+\frac{1}{2}}) \mid x \in \mathcal{X}\}, & (7.13c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - r\beta(Ax^{k+1} + By^{k+1} - b). & (7.13d) \end{cases}$$

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

$$\mathcal{D} = \{(s, r) \mid r + s > 0, \quad |r| < 1 + s - s^2, \quad |s| < 1 + r - r^2\}, \quad (7.14)$$

which is displayed in Figure 4. Essentially, discussing the convergence for the symmetric ADMM (1.8) over the domain \mathcal{D} defined in (1.9) is not easy; while its extension to the symmetric domain in Figure 4 is trivial. Thus, for succinctness, we omit the detail of this extension to the domain in Figure 4.

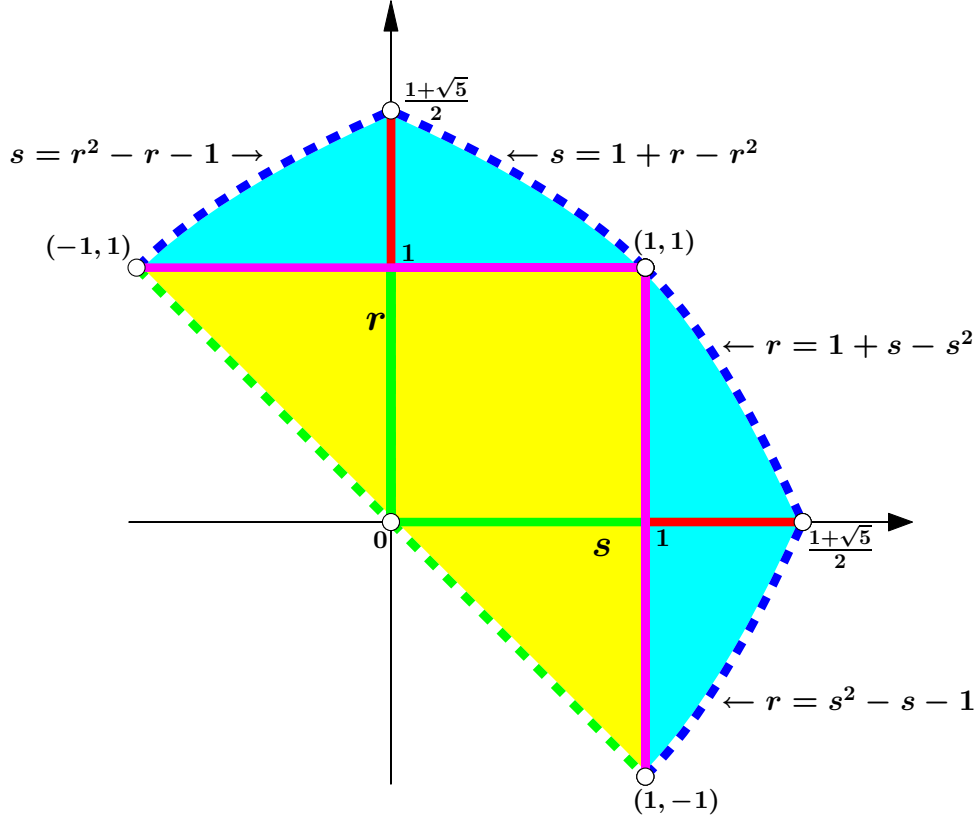


Figure 4. A symmetric step size domain with convergence for the symmetric ADMM (7.14)

References

- [1] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1-122.
- [2] 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, 1978.
- [3] 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.
- [4] E. Corman and X. M. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optim., 24 (2014), pp. 1614-1638.
- [5] W. Deng and W. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, manuscript, 2012.
- [6] 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.

- [7] J. Eckstein, Splitting methods for monotone operators with applications to parallel optimization, Ph.D thesis, MIT, 1989.
- [8] 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.
- [9] J. Eckstein and M. Fukushima, *Reformulations and applications of the alternating direction method of multipliers*, Large Scale Optimization, W.W. Hager, D.W. Hearn and P.M. Pardalos, eds., Springer US, (1994), pp. 115–134.
- [10] J. Eckstein and W. Yao, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, RUTCOR Research Report RRR 32-2012, December 2012.
- [11] 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(2)(2015), pp. 149-187.
- [12] D. Gabay, *Applications of the method of multipliers to variational inequalities*, Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, (1983), pp. 299–331.
- [13] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.
- [14] 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. Neittaanmki & O. Pironneau, eds., Computational Methods in Applied Sciences, Vol. 34, Springer, Dordrecht, 2014, 59-82.
- [15] 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 & O.Pironneau, eds., CIMNE, Barcelona, 2003, 67-79.
- [16] R. Glowinski, and A. Marrocco, *Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Revue Fr. Autom. Inform. Rech. Opér., Anal. Numér. 2 (1975), pp. 41–76.
- [17] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I*, Springer Series in Operations Research, Springer-Verlag, 2003.
- [18] 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.
- [19] 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. Appli., 32(2011), pp. 136-152.

- [20] B. S. He and X. M. Yuan, *On the $O(1/t)$ convergence rate of the alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700-709.
- [21] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4(1969), pp. 303-320.
- [22] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16(1979), pp. 964-979.
- [23] D. H. Peaceman and H. H. Rachford, *The numerical solution of parabolic elliptic differential equations*, SIAM J. Appl. Math. 3 (1955), pp. 28-41.
- [24] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.
- [25] Z. Wen, D. Goldfarb and W. Yin, *Alternating direction augmented Lagrangian methods for semidefinite programming*, Math. Program. Comput. 2 (3-4), 203-230, 2010.