Improving an ADMM-like Splitting Method via Positive-Indefinite Proximal Regularization for Three-Block Separable Convex Minimization

Bingsheng He[1] and **Xiaoming Yuan**[2]

August 14, 2016

Abstract. The augmented Lagrangian method (ALM) is fundamental for solving convex minimization models with linear constraints. When the objective function is separable such that it can be represented as the sum of more than one function without coupled variables, various splitting versions of the ALM have been well studied in the literature such as the alternating direction method of multipliers (ADMM) for the two-block separable case. It is known that the multi-block separable case where the objective function is the sum of more than two functions is more complicated than the two-block case; and it requires particular discussions on its theoretical convergence and algorithmic design. The ADMM-like splitting method in [12] allows some of its subproblems to be solved in parallel; while these subproblems should be regularized by some positive-definite proximal terms to ensure the convergence. In this paper, we further study this partially-parallel ADMM-like splitting method for a three-block separable convex minimization model and discuss how to relax the positive-definite proximal regularization terms in its subproblems. Inspired by our recent work in [10] for the ALM and ADMM, we show that it is not necessary to regularize the splitted subproblems by positive-definite proximal regularization for the method in [12]. Indeed, the proximal terms can be induced by a positive-indefiniteness proximal matrix. Thus, larger step sizes for solving these subproblems are allowed. This is an important feature to embark efficient algorithms with faster convergence. The convergence and worst-case convergence rate of the improved version of the ADMM-like splitting method in [12] with a less restricted proximal parameter are proved.

Keywords: Convex programming, augmented Lagrangian method, alternating direction method of multipliers, three-block, positive-indefinite proximal regularization, convergence analysis.

1 Introduction

Many applications arising in various areas can be modeled as convex minimization problems with linear constraints and separable objective functions that can be represented as the sum of more than one function without coupled variables. In this paper, we concentrate on the multi-block case where the objective unction is the sum of three functions without coupled variables:

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}, \tag{1.1}$$

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $C \in \Re^{m\times n_3}$; $b \in \Re^m$; $\mathcal{X} \subset \Re^{n_1}$, $\mathcal{Y} \subset \Re^{n_2}$ and $\mathcal{Z} \subset \Re^{n_3}$ are closed convex sets; and $\theta_i : \Re^{n_i} \to \Re$ $(i = 1, 2, 3)$ are closed convex but not necessarily smooth functions. For a concrete application of the abstract model (1.1), one of the functions may represent a data-fidelity term while the other two may account for various regularization terms for different applications. We refer to, e.g., [19, 23, 24, 25, 26], for some applications of (1.1). The solution set of (1.1) is assumed to be nonempty throughout.

[1]School of Management Science and Engineering, and Department of Mathematics, Nanjing University, China. This author was supported by the NSFC Grant 11471156 and 91530115. Email: hebma@nju.edu.cn

[2]Department of Mathematics, Hong Kong Baptist University, Hong Kong. This author was supported by a General Research Fund from Hong Kong Research Grants Council. Email: xmyuan@hkbu.edu.hk

The augmented Lagrangian method (ALM) originally proposed in [18, 21] is the starting point of our discussion on the model (1.1). Let the Lagrangian and augmented Lagrangian functions of (1.1) be given, respectively, by

$$L(x,y,z,\lambda) = \theta_1(x) + \theta_2(y) + \theta_3(y) - \lambda^T(Ax + By + Cz - b), \tag{1.2}$$

with $\lambda \in \Re^m$ the Lagrange multiplier; and

$$\mathcal{L}_\beta(x,y,z,\lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b) + \frac{\beta}{2}\|Ax + By + Cz - b\|^2, \tag{1.3}$$

with $\beta > 0$ the penalty parameter. When the three-block model (1.1) is purposively regarded as a generic convex minimization model and its objective function is treated as a whole, the ALM in [18, 21] can be applied directly with the resulting iterative scheme

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

Also, if two functions in the objective are treated together and two variables in the constraint are grouped accordingly, the alternating direction method of multipliers (ADMM) in [5] can also be directly applied to (1.1). The resulting iterative scheme reads as

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

Unless the functions and/or coefficient matrices in (1.1) are special enough, direct applications of ALM and ADMM in (1.4) and (1.5) usually are not preferred because the (x,y,z)-subproblem in (1.4a) and (y,z)-subproblem in (1.5b) may still be too hard (even when the functions θ_i per se are relatively easy). We are thus interested in the generic case where the three-block model (1.1) should not be ordinarily treated as a one-block or two-block case so that direct applications of the ALM and ADMM in (1.4) and (1.5) can be applied directly.

On the other hand, for specific applications of the model (1.1), functions in its objective usually have their own explanations and properties. Thus, it is usually necessary to treat them individually to design more efficient algorithms. More accurately, we are interested in such an algorithm that just needs to handle these functions θ_i separately in its iterative scheme. A natural idea is to split the subproblem in the original ALM (1.4) in the Jacobian or Gauss-Seidel manner; and obtain the schemes

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta(x^k, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases} \tag{1.6}$$

and

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y^{k+1}, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases} \tag{1.7}$$

All the subproblems in (1.6) and (1.7) are relatively easier than the original problem (1.1); only one function in its objective and a quadratic term are involved in the x-, y-, z-subproblems. However, as shown in [1, 9], neither of the schemes (1.6) and (1.7) is necessarily convergent. Therefore, although the schemes (1.6) and (1.7) can be obtained via splitting the ALM naturally, they are in good forms and they can be easily implemented, the lack of convergence requires more meticulous theoretical study and algorithmic design techniques to yield augmented-Lagrangian-based splitting algorithms for the three-block case (1.1). The results in [1, 9] also justify that designing augmented-Lagrangian-based splitting algorithms for the three-block case (1.1) is significantly different from that for the one- and two-block cases; and they need to be discussed separately despite that there is a rich literature on the ALM and ADMM. Indeed, the study of augmented-Lagrangian-based splitting methods for multi-block separable convex programming problems has recently received much attention from the community and there is a rich set of literatures, see, e.g., [9, 11, 12, 13, 14] for our previous work.

Despite their lack of convergence, the schemes (1.6) and (1.7) may empirically work well, see, e.g., [23, 26]. It is thus interesting to design an augmented-Lagrangian-based splitting method whose iterative scheme is analogous to (1.6), (1.7), or a fused one of both, while its theoretical convergence and empirical efficiency can be both ensured. To see such an method, let us recall the algorithm in [12]. Its iterative scheme for (1.1) reads as

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.8a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b), & (1.8b) \\ \left\{\begin{array}{l} y^{k+1} = \text{argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \text{argmin}\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \end{array}\right. & (1.8c) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), & (1.8d) \end{array}\right.$$

where the parameter μ is required to be $\mu \geq 2$ in [12]. The scheme (1.8) has the simplicity in sense of that each of the x-, y-, and z-subproblems involves just one function from (1.1) in its objective. Its efficiency has been verified in [12] by some sparse and low-rank models and image inpainting problems. Also, it was used in [2] to solve a dimensionality reduction problem on physical space.

As elucidated in Lemma 2.1, the scheme (1.8) can be rewritten as

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, z^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & (1.9a) \\ \left\{\begin{array}{l} y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \\ z^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau\beta}{2}\|C(z - z^k)\|^2 \;\big|\; z \in \mathcal{Z}\big\}, \end{array}\right. & (1.9b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), & (1.9c) \end{array}\right.$$

with $\tau = \mu - 1$ and thus $\tau \geq 1$ as required in [12]. The scheme (1.9) shows more clearly that it is a mixture of the augmented-Lagrangian-based splitting schemes (1.6) and (1.7), in which the x- and (y, z)-subproblems are updated in the alternating order while the (y, z)-subproblem is further splitted in parallel so that parallel computation can be implemented to the resulting y- and z-subproblems. Recall the lack of convergence of (1.6) and (1.7). Thus, it is necessary to regularize the splitted y- and z-subproblems appropriately in (1.9) to ensure the convergence. Indeed, in (1.9), the terms $\frac{\tau\beta}{2}\|B(y - y^k)\|^2$ and $\frac{\tau\beta}{2}\|C(z - z^k)\|^2$ can be regarded as proximal regularization terms and τ is the proximal parameter; see the seminal work [20, 22] of the proximal point algorithm (PPA) to better understand their roles. On the other hand, with fixed β, the proximal parameter τ determines the weights of the proximal terms in the objective functions of the subproblems in (1.9b) and it is well known (see, e.g. [20, 22]) that smaller values of τ are preferred as long as the convergence can be ensured.

As mentioned, in [12], we have shown that the condition $\tau \geq 1$ in (1.9) is sufficient to ensure the convergence. While, numerically, as shown in [12] and also in [2] (see Section V, Part B, Page 3247-3248), it has been observed that values very close to the lower bound $\tau = 1$ are much more preferred. For example, $\mu = 2.01$, i.e., $\tau = 1.01$, was recommended in [12] and then used in [2] to result in faster convergence. This raises the necessity of relaxing the theoretically sufficient condition $\tau \geq 1$ in (1.9) and seeking a lower bound smaller than 1 for τ. The main purpose of this paper is to formally show that the condition $\tau \geq 1$ in [12] can be relaxed to $\tau \geq 0.6$ and thus the numerical efficiency of the scheme (1.9) can be further improved by choosing smaller values of τ. This work is mainly inspired by our recent work [10] of relaxing the convergence conditions for the linearized versions of the ALM and ADMM via positive-indefinite proximal regularization.

The rest of this paper is organized as follows. We recall some preliminaries and prove some simple results for further analysis in Section 2. In Section 3, we show that the positive-indefiniteness proximal regularization occurs for the scheme (1.9) with $\tau \geq 0.6$. Then we provide an explanation in the prediction-correction framework for (1.9) in Section 4 and focus on an important crossing term in Section 5 that is the key for conducting convergence analysis for (1.9) with $\tau \geq 0.6$. In Section 6, we show by an example that the lower bound of τ can't be smaller than 0.5 to ensure the convergence of (1.9). The convergence and worst-case convergence rate of the scheme (1.9) with $\tau \geq 0.6$ are established in Sections 7 and 8, respectively. Finally, we make some conclusions in Section 9.

2 Preliminaries

In this section, we summarize some preliminary results, present a simple lemma and a reformulation of the scheme (1.9) for further analysis.

2.1 Variational inequality characterization

First of all, a pair of $\big((x^*, y^*, z^*), \lambda^*\big)$ is called a saddle point of the Lagrangian function defined in (1.2) if it satisfies the inequalities

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

We can rewrite these inequalities as

$$\begin{cases} x^* = \arg\min\{L(x, y^*, z^*, \lambda^*) \,|\, x \in \mathcal{X}\}, \\ y^* = \arg\min\{L(x^*, y, z^*, \lambda^*) \,|\, y \in \mathcal{Y}\}, \\ z^* = \arg\min\{L(x^*, y^*, z, \lambda^*) \,|\, z \in \mathcal{Z}\}, \\ \lambda^* = \arg\max\{L(x^*, y^*, z^*, \lambda) \,|\, \lambda \in \Re^m\}, \end{cases} \tag{2.1}$$

or the following variational inequalities:

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, \quad \forall\, y \in \mathcal{Y}, \\ z^* \in \mathcal{Z}, & \theta_3(z) - \theta_3(z^*) + (z - z^*)^T(-C^T\lambda^*) \geq 0, \quad \forall\, z \in \mathcal{Z}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* + Cz^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \end{cases} \tag{2.2}$$

We call (x, y, z) and λ the primal and dual variables, respectively.

The optimality condition of the model (1.1) can be characterized by the monotone variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.3a}$$

where

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \ \theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \ w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \ F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix} \tag{2.3b}$$

and

$$\Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m.$$

We denote by Ω^* the solution set of (2.3). Note that the operator F in (2.3b) is affine with a skew-symmetric matrix and thus we have

$$(w - \bar{w})^T(F(w) - F(\bar{w})) = 0, \quad \forall\, w, \bar{w}. \tag{2.4}$$

2.2 The equivalence of (1.9) with $\tau = \mu - 1$ and (1.8)

We show the equivalence of the schemes (1.9) with $\tau = \mu - 1$ and (1.8) in the following lemma.

Lemma 2.1 *The scheme (1.9) with $\tau = \mu - 1$ is equivalent to* (1.8).

Proof. To see the equivalence, we need only to show that their y-subproblems have the same optimality conditions. Skipping a constant, the objective function of the y-subproblem in (1.9) can be written as

$$\mathcal{L}_\beta(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 = \theta_2(y) - (\lambda^k)^T By + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 + \frac{\tau\beta}{2}\|B(y - y^k)\|^2.$$

Thus, the optimality condition of the y-subproblem in (1.9) is $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \begin{pmatrix} [-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} + Cz^k - b)] \\ +\tau\beta B^T B(y^{k+1} - y^k) \end{pmatrix} \geq 0, \ \forall y \in \mathcal{Y}. \tag{2.5}$$

Using (1.8b) to treat the term $[-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} + Cz^k - b)]$ in the above inequality, we obtain

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

Thus (2.5) can be written as $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+\frac{1}{2}} + (\tau+1)\beta B^T B(y^{k+1} - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{2.6}$$

On the other hand, the optimality condition of the y-subproblem in (1.8) is $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+\frac{1}{2}} + \mu\beta B^T B(y^{k+1} - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{2.7}$$

Setting $\tau = \mu - 1$, we see that the optimality condition of the y-subproblem (2.6) is just (2.7). The proof is complete. $\square$

2.3 A reformulation of (1.9)

Throughout, we fix β in (1.9) for the convenience of theoretical analysis in lighter notation. Indeed, because of our analysis in [1], without loss of the generality, we can just assume $\beta \equiv 1$. That is, the augmented Lagrangian function defined in (1.3) is reduced to

$$\mathcal{L}(x, y, z, \lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b) + \frac{1}{2}\|Ax + By + Cz - b\|^2; \tag{2.8}$$

and the iterative scheme of (1.9) is now simplified as

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(2.9a)} \\ \left\{\begin{array}{l} y^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \end{array}\right. & \text{(2.9b)} \\ \lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(2.9c)} \end{array}\right.$$

As mentioned, our discussion will be focused on investigating the convergence of the scheme (2.9) with $\tau \geq 0.6$, instead of $\tau \geq 1$ in [12].

3 The positive-indefiniteness in (2.9)

In this section, we revisit the proximally regularized subproblems in the scheme (2.9) from the variational inequality perspective; and show that the positive-indefiniteness proximal regularization occurs if $\tau \geq 0.6$ in (2.9) which is indeed the main difficulty of ensuring the convergence of (2.9) with this less restricted τ.

Note that the subproblem (2.9b) is specified as

$$y^{k+1} = \arg\min\{\theta_2(y) - y^T B\lambda^k + \frac{1}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 + \frac{\tau}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \tag{3.1a}$$

and

$$z^{k+1} = \arg\min\{\theta_3(z) - z^T C\lambda^k + \frac{1}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 + \frac{\tau}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}. \tag{3.1b}$$

The optimality condition of the y-subproblem in (2.9b) can be written as $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \begin{pmatrix} -B^T\lambda^k + B^T(Ax^{k+1} + By^{k+1} + Cz^k - b) \\ +\tau B^T B(y^{k+1} - y^k) \end{pmatrix} \geq 0, \ \forall y \in \mathcal{Y};$$

or equivalently: $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \begin{pmatrix} -B^T\lambda^k + B^T(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) \\ \tau B^T B(y^{k+1} - y^k) - B^T C(z^{k+1} - z^k) \end{pmatrix} \ge 0, \ \forall y \in \mathcal{Y}. \quad (3.2)$$

Furthermore, recall the equation (2.9c). The optimality condition of the y-subproblem in (2.9b) can be expressed as $y^{k+1} \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\big(-B^T\lambda^{k+1} + \tau B^T B(y^{k+1} - y^k) - B^T C(z^{k+1} - z^k)\big) \ge 0, \ \forall y \in \mathcal{Y}. \quad (3.3a)$$

Analogously, the optimality condition of the z-subproblem (3.1b) is $z^{k+1} \in \mathcal{Z}$ and

$$\theta_3(z) - \theta_3(z^{k+1}) + (z - z^{k+1})^T\big(-C^T\lambda^{k+1} - C^T B(y^{k+1} - y^k) + \tau C^T C(z^{k+1} - z^k)\big) \ge 0, \ \forall z \in \mathcal{Z}. \quad (3.3b)$$

Combining (3.3a) and (3.3b) together, we have $(y^{k+1}, z^{k+1}) \in \mathcal{Y} \times \mathcal{Z}$ and

$$\begin{pmatrix} \theta_2(y) - \theta_2(y^{k+1}) \\ \theta_3(z) - \theta_3(z^{k+1}) \end{pmatrix} + \begin{pmatrix} y - y^{k+1} \\ z - z^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -B^T\lambda^{k+1} \\ -C^T\lambda^{k+1} \end{pmatrix} + D_0 \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \end{pmatrix} \right\} \ge 0, \forall\, (y, z) \in \mathcal{Y} \times \mathcal{Z}, \quad (3.4)$$

where

$$D_0 = \begin{pmatrix} \tau B^T B & -B^T C \\ -C^T B & \tau C^T C \end{pmatrix}. \quad (3.5)$$

From the variational inequality characterization (3.4), we see that the subproblem (2.9b) is in form of the PPA in [20, 22]), with the proximal term

$$D_0 \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \end{pmatrix}$$

and the proximal matrix D_0 given in (3.5).

Obviously, the proximal matrix D_0 in (3.5) can be rewritten as

$$D_0 \;=\; \begin{pmatrix} (\tau - 1)B^T B & 0 \\ 0 & (\tau - 1)C^T C \end{pmatrix} + \begin{pmatrix} B^T \\ -C^T \end{pmatrix} (B, -C).$$

Thus, the matrix D_0 is positive- semidefinite and indefinite when $\tau \ge 1$ and $\tau \in (0, 1)$, respectively. Note that our analysis in the previous work [12] requires $\tau \ge 1$; thus the positive-semidefiniteness of D_0 is ensured and the convergence analysis can follow some PPA work in the literature. Since we are considering the case where $\tau \ge 0.6$, the matrix D_0 in (3.5) is not necessarily positive-definite and as shown later, this is the main difficulty in convergence analysis and it requires more sophisticated techniques to prove the convergence of (1.9) with $\tau \ge 0.6$.

4 A prediction-correction explanation of (2.9)

In this section, we show that the scheme (2.9) can be expressed by a prediction-correction framework. This prediction-correction explanation is only for the convenience of theoretical analysis and there is no need to follow this prediction-correction framework to implement the scheme (2.9) practically.

In the scheme (2.9), we see that x^k is not needed to generate the next $(k+1)$-th iterate; only (y^k, z^k, λ^k) are needed. Thus, we call x the intermediate variable; and (y, z, λ) essential variables. To distinguish their roles, accompanied with the notation in (2.3b), we additionally define the notation

$$v = \begin{pmatrix} y \\ z \\ \lambda \end{pmatrix}, \quad \mathcal{V} = \mathcal{Y} \times \mathcal{Z} \times R^m, \quad \text{and} \quad \mathcal{V}^* = \{(y^*, z^*, \lambda^*) \,|\, (x^*, y^*, z^*, \lambda^*) \in \Omega^*\}. \tag{4.1}$$

Moreover, we introduce the auxiliary variables $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ defined by

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \quad \tilde{z}^k = z^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - (Ax^{k+1} + By^k + Cz^k - b), \tag{4.2}$$

where $(x^{k+1}, y^{k+1}, z^{k+1})$ is the iterate generated by the scheme (2.9) from the given one (y^k, z^k, λ^k).

4.1 Predictor

First, ignoring some constant terms, the subproblem (2.9a) is equivalent to

$$x^{k+1} = \arg\min\{\theta_1(x) - x^T A\lambda^k + \frac{1}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}; \tag{4.3}$$

and its optimality condition can be rewritten as

$$\tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \; \forall\, x \in \mathcal{X}. \tag{4.4a}$$

On the other hand, since it holds

$$\begin{aligned} & -B^T\lambda^{k+1} + \tau B^T B(y^{k+1} - y^k) - B^T C(z^{k+1} - z^k) \\ &= \;\; -B^T[\lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b)] + \tau B^T B(y^{k+1} - y^k) - B^T C(z^{k+1} - z^k) \\ &= \;\; -B^T[\lambda^k - (Ax^{k+1} + By^k + Cz^k - b)] + (1+\tau)B^T B(y^{k+1} - y^k) \\ &= \;\; -B^T\tilde{\lambda}^k + (1+\tau)B^T B(\tilde{y}^k - y^k), \end{aligned}$$

the inequality (3.3a) can be written as

$$\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + (1+\tau)B^T B(\tilde{y}^k - y^k)\} \geq 0, \;\; \forall\, y \in \mathcal{Y}.$$

Accordingly, the inequality (3.4) becomes

$$\begin{aligned} &\begin{pmatrix} \theta_2(y) - \theta_2(\tilde{y}^k) \\ \theta_3(z) - \theta_3(\tilde{z}^k) \end{pmatrix} + \begin{pmatrix} y - \tilde{y}^k \\ z - \tilde{z}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -B^T \\ -C^T \end{pmatrix} \tilde{\lambda}^k + \right. \\ &\qquad \left. + \begin{pmatrix} (1+\tau)B^T B & 0 \\ 0 & (1+\tau)C^T C \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \end{pmatrix} \right\} \geq 0, \; \forall (y,z) \in \mathcal{Y} \times \mathcal{Z} \end{aligned} \tag{4.4b}$$

Note that the equality $\tilde{\lambda}^k = \lambda^k - (Ax^{k+1} + By^k + Cz^k - b)$ in (4.2) can be written as the variational inequality form

$$\tilde{\lambda}^k \in \Re^m, \;\; (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (\tilde{\lambda}^k - \lambda^k)\} \geq 0, \;\; \forall \lambda \in \Re^m. \tag{4.4c}$$

Therefore, it follows from the inequalities (4.4a), (4.4b) and (4.4c) that the auxiliary variable $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ defined in (4.2) satisfies the following variational inequality:

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

where

$$Q = \begin{pmatrix} (1+\tau)B^T B & 0 & 0 \\ 0 & (1+\tau)C^T C & 0 \\ -B & -C & I_m \end{pmatrix}. \tag{4.5b}$$

We also see that because of the introduction of the auxiliary variable $\tilde{\lambda}^k := \lambda^k - (Ax^{k+1} + By^k + Cz^k - b)$ in (4.2), the right-hand side of (4.5a) only involves the essential valuables v, without the intermediate variable x. This enables us to focus only on the essential variables for conducting the convergence analysis for the scheme (2.9).

4.2 Corrector

Recall we define by v in (4.1) the essential variables for the scheme (2.9). Using the notation in (4.2), we have

$$\begin{aligned} & (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) \\ = \;& -B(y^k - y^{k+1}) - C(z^k - z^{k+1}) + (Ax^{k+1} + By^k + Cz^k - b) \\ = \;& -B(y^k - \tilde{y}^k) - C(z^k - \tilde{z}^k) + (\lambda^k - \tilde{\lambda}^k). \end{aligned}$$

Thus, it follows from (2.9c) that

$$\lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) = \lambda^k - [-B(y^k - \tilde{y}^k) - C(z^k - \tilde{z}^k) + (\lambda^k - \tilde{\lambda}^k)];$$

and we obtain that the essential variables of (2.9) are updated by the following scheme:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{4.6a}$$

where

$$M = \begin{pmatrix} I_{n_2} & 0 & 0 \\ 0 & I_{n_3} & 0 \\ -B & -C & I_m \end{pmatrix}. \tag{4.6b}$$

Overall, the scheme (2.9) can be explained by a prediction-correction framework which generates a predictor characterized by the step (4.5) and then corrects it by the step (4.6).

4.3 Further investigation on the predictor

As we shall show, the inequality (4.5) indicates the discrepancy between $\tilde{w}^k$ and a solution point of the variational inequality (2.3) and it plays an important role in the convergence analysis for the scheme (2.9). Indeed, we can further investigate the predictor $\tilde{w}^k$ and derive a new right-hand side for the inequality (4.5) that is more preferred for establishing the convergence. For this purpose, let us define a matrix as

$$H = \begin{pmatrix} (1+\tau)B^T B & 0 & 0 \\ 0 & (1+\tau)C^T C & 0 \\ 0 & 0 & I_m \end{pmatrix}, \tag{4.7}$$

which is positive definite for any $\tau > 0$ when B and C are both full column rank. Then, for the matrices Q and M defined in (4.5b) and (4.6b), respectively, it obviously holds that

$$Q = HM. \tag{4.8}$$

In the following theorem, we polish the right-hand side of (4.5) and it helps us better discern the difference of convergence proof for the scheme (2.9) with $\tau \geq 0.6$ from that with $\tau \geq 1$ in [12]. More precisely, this theorem shows that the main difficulty for proving the convergence of (2.9) with $\tau \geq 0.6$ is due to the possible negativeness of a crossing term and it inspires us to focus on this term later for the convergence analysis.

Theorem 4.1 *Let $\{w^k\}$ be the sequence generated by (2.9) for the problem (1.1) and $\tilde{w}^k$ be defined by (4.2). Then we have $\tilde{w}^k \in \Omega$ and*

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(w) \geq \frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}(v^k-\tilde{v}^k)^T G_0(v^k-\tilde{v}^k), \quad \forall w \in \Omega, \tag{4.9}$$

where the matrix G_0 is given by

$$G_0 = Q^T + Q - M^T HM. \tag{4.10}$$

Proof. Using $Q = HM$ (see (4.8)) and the relation (4.6a), the right-hand side of (4.5a) can be written as

$$(v-\tilde{v}^k)^T H(v^k - v^{k+1}),$$

and hence we have

$$\theta(u) - \theta(\tilde{u}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \tag{4.11}$$

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 of (4.11) with

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

we obtain

$$(v-\tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \frac{1}{2}(\|v^k-\tilde{v}^k\|_H^2 - \|v^{k+1}-\tilde{v}^k\|_H^2). \tag{4.12}$$

For the last term of (4.12), we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(4.6a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&\quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\quad \overset{(4.10)}{=} \quad (v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k).
\end{aligned} \tag{4.13}$$

Substituting (4.12) into (4.13), we get

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k). \tag{4.14}$$

Recall that $(w - \tilde{w}^k)^T F(\tilde{w}^k) = (w - \tilde{w}^k)^T F(w)$ (see (2.4)). Using this fact, the assertion of this lemma follows from (4.11) and (4.14) directly. □

When G_0 in (4.10) is positive definite, as shown in [12], it is relatively easier to use the assertion (4.9) to prove the global convergence and worst-case $O(1/t)$ convergence rate of the scheme (2.9), see similar techniques in, e.g., [8, 17] and [6] (Sections 4 and 5 therein) for a tutorial proof. For the matrix G_0 given in (4.10), since $HM = Q$ and $M^T HM = M^T Q$, we have

$$\begin{aligned} M^T HM &= \begin{pmatrix} I_{n_2} & 0 & -B^T \\ 0 & I_{n_3} & -C^T \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} (1+\tau)B^T B & 0 & 0 \\ 0 & (1+\tau)C^T C & 0 \\ -B & -C & I_m \end{pmatrix} \\ &= \begin{pmatrix} (2+\tau)B^T B & B^T C & -B^T \\ C^T B & (2+\tau)C^T C & -C^T \\ -B & -C & I_m \end{pmatrix}. \end{aligned}$$

Then, using (4.5b) and the above equation, we have

$$\begin{aligned} G_0 &= (Q^T + Q) - M^T HM \\ &= \begin{pmatrix} (2+2\tau)B^T B & 0 & -B^T \\ 0 & (2+2\tau)C^T C & -C^T \\ -B & -C & 2I_m \end{pmatrix} - \begin{pmatrix} (2+\tau)B^T B & B^T C & -B^T \\ C^T B & (2+\tau)C^T C & -C^T \\ -B & -C & I_m \end{pmatrix} \\ &= \begin{pmatrix} \tau B^T B & -B^T C & 0 \\ -C^T B & \tau C^T C & 0 \\ 0 & 0 & I_m \end{pmatrix}. \end{aligned} \tag{4.15}$$

According to the definition of D_0 in (3.5), (4.15) can be rewritten as

$$G_0 = \begin{pmatrix} \multicolumn{2}{c}{\multirow{2}{*}{D_0}} & 0 \\ & & 0 \\ 0 & 0 & I_m \end{pmatrix}.$$

Therefore, for $\tau \in (\frac{3}{5}, 1)$, G_0 is not positive-definite because the matrix D_0 is not so. The positive-indefiniteness of G_0 is indeed the main difficulty of proving the convergence of the scheme (2.9) with $\tau \geq 0.6$; and we need to look into the crossing term $(v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k)$ more intensively.

5 Investigation of the crossing term $(v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k)$

As mentioned, the key point for proving the convergence of the scheme (2.9) with $\tau \geq 0.6$ is to analyze the crossing term $(v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k)$ which is not guaranteed to be positive. In this section, we focus on investigating this term and show that

$$(v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k) \geq \psi(v^k, v^{k+1}) - \psi(v^{k-1}, v^k) + \varphi(v^k, v^{k+1}), \tag{5.1}$$

where $\psi(\cdot,\cdot)$ and $\varphi(\cdot,\cdot)$ are both non-negative functions. The first two terms $\psi(v^k,v^{k+1})-\psi(v^{k-1},v^k)$ in the right-hand side of (5.1) can be manipulated consecutively between iterates and the last term $\varphi(v^k,v^{k+1})$ should be such an error bound that can measure how much w^{k+1} fails to be a solution point of (2.3). If we find such functions that guarantee the assertion (5.1), then we can substitute it into (4.9) and get the inequality

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(w)\\ \geq\;& \frac{1}{2}\big(\|v-v^{k+1}\|_H^2+\psi(v^k,v^{k+1})\big)-\frac{1}{2}\big(\|v-v^k\|_H^2+\psi(v^{k-1},v^k)\big)+\frac{1}{2}\varphi(v^k,v^{k+1}),\ \forall w\in\Omega. \qquad (5.2)\end{aligned}$$

As we shall show, all the components of the right-hand side of (5.2) in parentheses should be non-negative to establish the convergence and convergence rate of (2.9). It is indeed this requirement that implies our restriction of $\tau\geq 0.6$; we show details in Theorem 5.1. The following lemmas and theorem are for this purpose; and similar techniques can be referred to [4, 10, 15] for the convergence analysis of the ADMM.

Lemma 5.1 *Let $\{w^k\}$ be the sequence generated by* (2.9) *for the problem* (1.1) *and $\tilde{w}^k$ be defined by* (4.2)*. Then we have*

$$\begin{aligned}(v^k-\tilde{v}^k)^T G_0(v^k-\tilde{v}^k) \;=\;& (1+\tau)\|B(y^k-y^{k+1})\|^2+(1+\tau)\|C(z^k-z^{k+1})\|^2+\|\lambda^k-\lambda^{k+1}\|^2\\ &+2(\lambda^k-\lambda^{k+1})^T\big(B(y^k-y^{k+1})+C(z^k-z^{k+1})\big). \qquad (5.3)\end{aligned}$$

Proof. First, according to (4.15), we have

$$G_0=\begin{pmatrix}\tau B^TB & -B^TC & 0\\ -C^TB & \tau C^TC & 0\\ 0 & 0 & I_m\end{pmatrix}=\begin{pmatrix}(1+\tau)B^TB & 0 & 0\\ 0 & (1+\tau)C^TC & 0\\ 0 & 0 & I_m\end{pmatrix}-\begin{pmatrix}B^TB & B^TC & 0\\ C^TB & C^TC & 0\\ 0 & 0 & 0\end{pmatrix}$$

and thus

$$\begin{aligned}(v^k-\tilde{v}^k)^T G_0(v^k-\tilde{v}^k) \;=\;& (1+\tau)\|B(y^k-\tilde{y}^k)\|^2+(1+\tau)\|C(z^k-\tilde{z}^k)\|^2+\|\lambda^k-\tilde{\lambda}^k\|^2\\ &-\|B(y^k-\tilde{y}^k)+C(z^k-\tilde{z}^k)\|^2.\end{aligned}$$

For the term $\|\lambda^k-\tilde{\lambda}^k\|^2$ in the right hand-side of the above equation, because $\tilde{x}^k=x^{k+1}$,

$$\lambda^k-\tilde{\lambda}^k=Ax^{k+1}+By^k+Cz^k-b \quad\text{and}\quad Ax^{k+1}+By^{k+1}+Cz^{k+1}-b=\lambda^k-\lambda^{k+1},$$

we have

$$\lambda^k-\tilde{\lambda}^k=B(y^k-y^{k+1})+C(z^k-z^{k+1})+(\lambda^k-\lambda^{k+1}).$$

Finally, we get

$$\begin{aligned}&(v^k-\tilde{v}^k)^T G_0(v^k-\tilde{v}^k)\\ =\;& (1+\tau)\|B(y^k-y^{k+1})\|^2+(1+\tau)\|C(z^k-z^{k+1})\|^2-\|B(y^k-y^{k+1})+C(z^k-z^{k+1})\|^2\\ &+\|B(y^k-y^{k+1})+C(z^k-z^{k+1})+(\lambda^k-\lambda^{k+1})\|^2\\ =\;& (1+\tau)\|B(y^k-y^{k+1})\|^2+(1+\tau)\|C(z^k-z^{k+1})\|^2+\|\lambda^k-\lambda^{k+1}\|^2\\ &+2(\lambda^k-\lambda^{k+1})^T\big(B(y^k-y^{k+1})+C(z^k-z^{k+1})\big).\end{aligned}$$

The lemma is proved. $\square$

For the right-hand side of (5.3), there are some quadratic terms involving two consecutive iterates and they are easy to be manipulated in the convergence proof. Now, we treat the crossing term in the right-hand side of (5.3) and seek its lower bound that can also be expressed by some quadratic terms involving two consecutive iterates. We additionally define a new matrix

$$D = D_0 + \frac{2}{5}\begin{pmatrix} B^TB & 0 \\ 0 & C^TC \end{pmatrix}, \tag{5.4}$$

where the matrix D_0 is defined in (3.5). Notice that D is positive semi-definite when $\tau \geq 0.6$.

Lemma 5.2 *Let $\{w^k\}$ be the sequence generated by* (2.9) *for the problem* (1.1) *and $\tilde{w}^k$ be defined by* (4.2). *Then we have*

$$\begin{aligned} &2(\lambda^k - \lambda^{k+1})^T\big(B(y^k - y^{k+1}) + C(z^k - z^{k+1})\big) \\ \geq\ & \left\|\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1}\end{pmatrix}\right\|_D^2 - \left\|\begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k\end{pmatrix}\right\|_D^2 - \frac{6}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\Big) \\ & -\frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big), \end{aligned} \tag{5.5}$$

where the matrix D is defined in (5.4).

Proof. Recall (3.4). It holds that

$$\begin{aligned} (y^{k+1}, z^{k+1}) \in \mathcal{Y} \times \mathcal{Z},\ & \begin{pmatrix} \theta_2(y) - \theta_2(y^{k+1}) \\ \theta_3(z) - \theta_3(z^{k+1}) \end{pmatrix} + \begin{pmatrix} y - y^{k+1} \\ z - z^{k+1} \end{pmatrix}^T \\ & \left\{\begin{pmatrix} -B^T \\ -C^T \end{pmatrix}\lambda^{k+1} + D_0\begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \end{pmatrix}\right\} \geq 0, \quad \forall (y,z) \in \mathcal{Y} \times \mathcal{Z}. \end{aligned} \tag{5.6}$$

Analogously, for the previous iteration, we have

$$\begin{aligned} (y^k, z^k) \in \mathcal{Y} \times \mathcal{Z},\ & \begin{pmatrix} \theta_2(y) - \theta_2(y^k) \\ \theta_3(z) - \theta_3(z^k) \end{pmatrix} + \begin{pmatrix} y - y^k \\ z - z^k \end{pmatrix}^T \\ & + \left\{\begin{pmatrix} -B^T \\ -C^T \end{pmatrix}\lambda^k + D_0\begin{pmatrix} y^k - y^{k-1} \\ z^k - z^{k-1} \end{pmatrix}\right\} \geq 0, \quad \forall (y,z) \in \mathcal{Y} \times \mathcal{Z}. \end{aligned} \tag{5.7}$$

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

$$\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T \left\{\begin{pmatrix} B^T \\ C^T \end{pmatrix}(\lambda^k - \lambda^{k+1}) + D_0\left[\begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \end{pmatrix} - \begin{pmatrix} y^k - y^{k-1} \\ z^k - z^{k-1} \end{pmatrix}\right]\right\} \geq 0.$$

Consequently, we have

$$\begin{aligned} & 2(\lambda^k - \lambda^{k+1})^T\big(B(y^k - y^{k+1}) + C(z^k - z^{k+1})\big) \\ \geq\ & 2\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T D_0 \left[\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} - \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix}\right]. \end{aligned} \tag{5.8}$$

It follows from (3.5) and (5.4) that

$$D_0 = D - \frac{2}{5}\begin{pmatrix} B^TB & 0 \\ 0 & C^TC \end{pmatrix}.$$

Thus, using Cauchy-Schwaez inequality, from (5.8) we obtain

$$\begin{aligned}
&2(\lambda^k - \lambda^{k+1})^T\big(B(y^k - y^{k+1}) + C(z^k - z^{k+1})\big) \\
&\geq 2\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T \left\{ D - \frac{2}{5}\begin{pmatrix} B^TB & 0 \\ 0 & C^TC \end{pmatrix} \right\} \left[\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} - \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right] \\
&= 2\left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 - 2\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T D \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \\
&\quad -\frac{4}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\Big) \\
&\quad +\frac{4}{5}\Big((y^k - y^{k+1})^T B^TB(y^{k-1} - y^k) + (z^k - z^{k+1})^T C^TC(z^{k-1} - z^k)\Big) \\
&\geq \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 - \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 - \frac{6}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\Big) \\
&\quad -\frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big),
\end{aligned}$$

where the last inequality is because of the Cauchy-Schwarz inequality. We thus prove the assertion (5.5). □

Recall that we want to bound the crossing term $(v^k - \tilde{v}^k)^T G_0(v^k - \tilde{v}^k)$ in form of (5.2). Our previous analysis enables to achieve it; and this is the basis for the convergence results to be established later. We present the result in the following theorem.

Theorem 5.1 *Let* $\{w^k\}$ *be the sequence generated by* (2.9) *for the problem* (1.1) *and* $\tilde{w}^k$ *be defined by* (4.2). *Then we have*

$$\begin{aligned}
&2\big\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w)\big\} \\
&\geq \big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) \\
&\quad + \left\{ \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\Big) \right\} \\
&\quad - \left\{ \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big) \right\} \\
&\quad + \Big((\tau - \frac{3}{5})\big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\big) + \|\lambda^k - \lambda^{k+1}\|^2\Big).
\end{aligned} \tag{5.9}$$

Proof. Substituting (5.5) into (5.3), we obtain

$$
\begin{aligned}
&(v^k - \tilde{v}^k)^T G_0 (v^k - \tilde{v}^k) \\
&\geq \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 - \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 \\
&\quad + (\tau - \frac{1}{5})\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \Big) \\
&\quad - \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2 \Big) + \|\lambda^k - \lambda^{k+1}\|^2,
\end{aligned} \tag{5.10}
$$

which can be rewrite as

$$
\begin{aligned}
&(v^k - \tilde{v}^k)^T G_0 (v^k - \tilde{v}^k) \\
&\geq \left\{ \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \Big) \right\} \\
&\quad - \left\{ \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2 \Big) \right\} \\
&\quad + \Big((\tau - \frac{3}{5}) \big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \big) + \|\lambda^k - \lambda^{k+1}\|^2 \Big).
\end{aligned} \tag{5.11}
$$

The assertion (5.11) is exactly a desirable form of (5.1). Then, the assertion (5.9) follows from Theorem 4.1 and (5.11) immediately. □

6 An example

As mentioned, our restriction of $\tau \geq 0.6$ is indeed for sufficiently ensuring the non-negativeness of the coefficients in the right-hand side of (5.9). Although it is still unknown whether or not 0.6 is the smallest lower bound for τ, below we show by an example that the lower bound of τ to ensure the convergence of the scheme (2.9) can't be smaller than 0.5. Therefore, 0.6 is nearly the optimal lower bound of τ if it is not the truly optimal one.

For any given $\tau < 0.5$, we take $\epsilon = 0.5 - \tau > 0$ and consider the problem

$$
\min\{x + \frac{\epsilon}{2} y^2 + \frac{\epsilon}{2} z^2 \mid x + y + z = 0,\ x \in \{0\}, y \in \Re, z \in \Re\}, \tag{6.1}
$$

which is a special case of the model (1.1) under discussion. Obviously, the solution of this problem is $x = y = z = 0$.

The augmented Lagrangian function of (6.1) with a penalty parameter of 1 is

$$
\mathcal{L}(x, y, z, \lambda) = x + \frac{\epsilon}{2} y^2 + \frac{\epsilon}{2} z^2 - \lambda^T (x + y + z) + \frac{1}{2} \|x + y + z\|^2;
$$

and the iterative scheme (2.9) for (6.1) is

$$
\left\{
\begin{array}{lcl}
x^{k+1} & = & \arg\min\big\{ \mathcal{L}(x, y^k, z^k, \lambda^k) \bigm| x \in \{0\} \big\}, \\
y^{k+1} & = & \arg\min\big\{ \mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|y - y^k\|^2 \bigm| y \in \Re \big\}, \\
z^{k+1} & = & \arg\min\big\{ \mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|z - z^k\|^2 \bigm| z \in \Re \big\}, \\
\lambda^{k+1} & = & \lambda^k - (x^{k+1} + y^{k+1} + z^{k+1}).
\end{array}
\right. \tag{6.2}
$$

Since $\mathcal{X} = \{0\}$, we have $x^{k+1} \equiv 0$. Ignoring constant terms in the objective function of the subproblems, the recursion (6.2) can be written as

$$\left\{\begin{array}{lcl} x^{k+1} & \equiv & 0, \\ y^{k+1} & = & \arg\min\{\frac{\epsilon}{2}y^2 - y^T\lambda^k + \frac{1}{2}\|y + z^k\|^2 + \frac{\tau}{2}\|y - y^k\|^2 \mid y \in \Re\}, \\ z^{k+1} & = & \arg\min\{\frac{\epsilon}{2}z^2 - z^T\lambda^k + \frac{1}{2}\|y^k + z\|^2 + \frac{\tau}{2}\|z - z^k\|^2 \mid z \in \Re\}, \\ \lambda^{k+1} & = & \lambda^k - (y^{k+1} + z^{k+1}). \end{array}\right. \tag{6.3}$$

Further, it follows from (6.3) that

$$\left\{\begin{array}{l} \epsilon y^{k+1} - \lambda^k + (y^{k+1} + z^k) + \tau(y^{k+1} - y^k) = 0, \\ \epsilon z^{k+1} - \lambda^k + (z^{k+1} + y^k) + \tau(z^{k+1} - z^k) = 0, \\ \lambda^{k+1} = \lambda^k - (y^{k+1} + z^{k+1}). \end{array}\right. \tag{6.4}$$

Thus, the iterative scheme for $v = (y, z, \lambda)$ can be written as

$$\left\{\begin{array}{l} (\tau + 1 + \epsilon)y^{k+1} = \tau y^k - z^k + \lambda^k, \\ (\tau + 1 + \epsilon)z^{k+1} = -y^k + \tau z^k + \lambda^k, \\ \lambda^{k+1} = \lambda^k - (y^{k+1} + z^{k+1}). \end{array}\right. \tag{6.5}$$

Without loss of generality, we can take $y^0 = z^0$ and thus $y^k \equiv z^k$, for all $k > 0$. Using this fact and $\tau + \epsilon = 0.5$, we get

$$\left\{\begin{array}{l} \frac{3}{2}y^{k+1} = (\tau - 1)y^k + \lambda^k, \\ \lambda^{k+1} = \lambda^k - 2y^{k+1}. \end{array}\right. \tag{6.6}$$

With elementary manipulations, we obtain

$$\left\{\begin{array}{l} y^{k+1} = \dfrac{-2(1-\tau)}{3}y^k + \dfrac{2}{3}\lambda^k, \\ \lambda^{k+1} = \dfrac{4(1-\tau)}{3}y^k + \dfrac{-1}{3}\lambda^k, \end{array}\right. \tag{6.7}$$

which can be written as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = P(\tau)\begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} \quad \text{with} \quad P(\tau) = \frac{1}{3}\begin{pmatrix} -2(1-\tau) & 2 \\ 4(1-\tau) & -1 \end{pmatrix}. \tag{6.8}$$

Let $f_1(\tau)$ and $f_2(\tau)$ be the two eigenvalues of the matrix $P(\tau)$. Then we have

$$f_1(\tau) = \frac{1}{6}\Big((2\tau - 3) + \sqrt{(3 - 2\tau)^2 + 24(1 - \tau)}\Big),$$

and

$$f_2(\tau) = \frac{1}{6}\Big((2\tau - 3) - \sqrt{(3 - 2\tau)^2 + 24(1 - \tau)}\Big).$$

Certainly, the scheme (6.7) is divergent if the absolute value of one of the eigenvalues of the matrix $P(\tau)$ is greater than 1. Indeed, it holds that $f_2(\tau) < -1$ for any $\tau \in (0, 0.5)$. To see this assertion, we notice that

$$\begin{array}{rcrcl} f_2(\tau) < -1 & \Leftrightarrow & (2\tau-3) - \sqrt{(3-2\tau)^2 + 24(1-\tau)} & < & -6 \\ & \Leftrightarrow & 2\tau+3 & < & \sqrt{4\tau^2 - 36\tau + 33} \\ & \Leftrightarrow & 4\tau^2 + 12\tau + 9 & < & 4\tau^2 - 36\tau + 33 \\ & \Leftrightarrow & \tau & < & 0.5. \end{array}$$

Hence, the scheme (2.9) is not necessarily convergent for any $\tau < 0.5$; and the extent of improvement over the derived bound of $\tau \geq 0.6$ for (2.9), if any, should be at most 0.1.

7 Convergence

Theorem 5.1 is important for establishing the convergence of the scheme (2.9). With it, the remaining part of deriving the convergence results is subroutine. In this section, we prove the global convergence of the scheme (2.9). We first prove a lemma to show the contraction property of the sequence generated by (2.9).

Lemma 7.1 *Let $\{w^k\}$ be the sequence generated by* (2.9) *with $\tau \geq 0.6$ for the problem* (1.1)*. Then we have*

$$\begin{aligned} &\|v^{k+1} - v^*\|_H^2 + \left[\left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \Big) \right] \\ &\leq \; \|v^k - v^*\|_H^2 + \left[\left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2 \Big) \right] \\ &\quad - \Big((\tau - \frac{3}{5})\big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \big) + \|\lambda^k - \lambda^{k+1}\|^2 \Big). \end{aligned} \tag{7.1}$$

Proof. Setting $w = w^*$ in (5.9) and using

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

we obtain the assertion (7.1) immediately. □

Theorem 7.1 *Let $\{w^k\}$ be the sequence generated by* (2.9) *with $\tau \geq 0.6$ for the problem* (1.1)*. Then the sequence $\{v^k\}$ converges to v^∞ which belongs to $\mathcal{V}^*$ when B and C are both full column rank.*

Proof. First, it follows from (7.1) that

$$\begin{aligned} &(\tau - \frac{3}{5})\big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \big) + \|\lambda^k - \lambda^{k+1}\|^2 \\ &\leq \; \left\{ \|v^k - v^*\|_H^2 + \left[\left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2 \Big) \right] \right\} \\ &\quad - \left\{ \|v^{k+1} - v^*\|_H^2 + \left[\left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 \Big) \right] \right\}. \end{aligned}$$

Summarizing the inequality last inequality over $k = 1, 2, \ldots$, we obtain

$$\sum_{k=1}^{\infty}\Big\{\big(\tau - \frac{3}{5}\big)\big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\big) + \|\lambda^k - \lambda^{k+1}\|^2\Big\}$$
$$\leq \|v^1 - v^*\|_H^2 + \left[\left\|\begin{pmatrix} y^0 - y^1 \\ z^0 - z^1 \end{pmatrix}\right\|_D^2 + \frac{2}{5}\Big(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2\Big)\right]$$

and thus

$$\lim_{k\to\infty} \|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 = 0. \tag{7.2}$$

For an arbitrarily fixed $v^* \in \mathcal{V}^*$, it follows from (7.1) that for any $k > 1$, we have

$$\begin{aligned}\|v^{k+1} - v^*\|_H^2 &\leq \|v^k - v^*\|_H^2 + \left[\left\|\begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix}\right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big)\right] \\ &\leq \|v^1 - v^*\|_H^2 + \left[\left\|\begin{pmatrix} y^0 - y^1 \\ z^0 - z^1 \end{pmatrix}\right\|_D^2 + \frac{2}{5}\Big(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2\Big)\right]. \end{aligned} \tag{7.3}$$

Thus the sequence $\{v^k\}$ is bounded. Because M is non-singular, according to (4.6), $\{\tilde{v}^k\}$ is also bounded. Let v^∞ be a cluster point $\{\tilde{v}^k\}$ and $\{\tilde{v}^{k_j}\}$ be the subsequence of $\{\tilde{v}^k\}$ converging to v^∞. Let x^∞ be the vector induced by given $(y^\infty, z^\infty, \lambda^\infty) \in \mathcal{V}$. Then, it follows from (4.11) that

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

which means w^∞ is a solution point of (2.3) and its essential part $v^\infty \in \mathcal{V}^*$. Since $v^\infty \in \mathcal{V}^*$, it follows from (7.3) that

$$\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2 + \left[\left\|\begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix}\right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big)\right]. \tag{7.4}$$

Together with (7.2), it is impossible that the sequence $\{v^k\}$ has more than one cluster point. Thus $\{v^k\}$ converges to v^∞ and the proof is complete. □

8 Convergence rate

In this section, we derive a worst-case $O(1/t)$ convergence rate measured by the iteration complexity for the scheme (2.9), where t is the iteration counter. Hence, although we relax the condition $\tau \geq 1$ in [12] to $\tau \geq 0.6$, the same convergence rate result in [12] remains valid for the scheme (2.9). Similar analysis is refereed to [10, 12, 16].

First of all, recall (2.3). If we find $\tilde{w}$ satisfying the inequality

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

then $\tilde{w}$ is a solution point of (2.3). As mentioned in (2.4), we have $(w - \tilde{w})^T F(\tilde{w}) = (w - \tilde{w})^T F(\tilde{w})$. Thus, a solution point $\tilde{w}$ of (2.3) can be also characterized by

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

Therefore, as [3], for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ if it satisfies

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

where

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

In the following, we show that based on the first t iterates generated by the scheme (2.9), we can find an approximate solution of (2.3), denoted by $\tilde{w} \in \Omega$, such that

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \leq \epsilon, \tag{8.5}$$

where $\epsilon = O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate is established for the scheme (2.9). Theorem 5.1 is still the basis for the analysis in this section.

Theorem 8.1 *Let* $\{w^k\}$ *be the sequence generated by* (2.9) *with* $\tau \geq 0.6$ *for the problem* (1.1) *and* $\tilde{w}^k$ *be defined by* (4.2). *Then for any integer* t*, we have*

$$\begin{aligned} &\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \\ &\leq \frac{1}{2t}\left\{ \|v - v^1\|_H^2 + \left\| \begin{pmatrix} y^0 - y^1 \\ z^0 - z^1 \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2\Big) \right\}, \end{aligned} \tag{8.6}$$

where

$$\tilde{w}_t = \frac{1}{t}\Big(\sum_{k=1}^{t} \tilde{w}^k\Big). \tag{8.7}$$

Proof. First, it follows from (5.9) that

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ &\geq \frac{1}{2}\left\{ \|v - v^{k+1}\|_H^2 + \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\Big) \right\} \\ &\quad -\frac{1}{2}\left\{ \|v - v^k\|_H^2 + \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big) \right\}. \end{aligned}$$

Thus, we have

$$\begin{aligned} &\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) \\ &\quad +\frac{1}{2}\left\{ \|v - v^{k+1}\|_H^2 + \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2\Big) \right\} \\ &\leq \frac{1}{2}\left\{ \|v - v^k\|_H^2 + \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2\Big) \right\}. \end{aligned} \tag{8.8}$$

Summarizing the inequality(8.8) over $k = 1, 2, \ldots, t$, we obtain

$$\sum_{k=1}^{t} \theta(\tilde{u}^k) - t\theta(u) + (\sum_{k=1}^{t} \tilde{w}^k - tw)^T F(w)$$
$$\leq \frac{1}{2} \left\{ \|v - v^1\|_H^2 + \left\| \begin{pmatrix} y^0 - y^1 \\ z^0 - z^1 \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2\Big) \right\},$$

and thus

$$\frac{1}{t}(\sum_{k=1}^{t} \theta(\tilde{u}^k)) - \theta(u) + (\tilde{w}_t - w)^T F(w)$$
$$\leq \frac{1}{2t} \left\{ \|v - v^1\|_H^2 + \left\| \begin{pmatrix} y^0 - y^1 \\ z^0 - z^1 \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2\Big) \right\}, \quad (8.9)$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t}(\sum_{k=1}^{t} \tilde{u}^k),$$

we have that

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

Substituting it into (8.9), the assertion of this theorem follows directly. □

For a given compact set $\mathcal{D}_{(\tilde{w})} \subset \Omega$, let

$$d := \sup \left\{ \|v - v^1\|_H^2 + \left\| \begin{pmatrix} y^0 - y^1 \\ z^0 - z^1 \end{pmatrix} \right\|_D^2 + \frac{2}{5}\Big(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2\Big), \;\Big|\; w \in \mathcal{D}_{(\tilde{w})} \right\}$$

where $v^0 = (y^0, z^0, \lambda^0)$ and $v^1 = (y^1, z^1, \lambda^1)$ are the initial and the first generated iterates, respectively. Then, after t iterations of the scheme (2.9), the point $\tilde{w}_t \in \Omega$ defined in (8.7) satisfies

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{ \theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \big\} \leq \frac{d}{2t} = O(\frac{1}{t}),$$

which means $\tilde{w}_t$ is an approximate solution of the variational inequality (2.3) with an accuracy of $O(1/t)$ (recall (8.5)). That is, a worst-case $O(1/t)$ convergence rate is established for the scheme (2.9) with $\tau \geq 0.6$. Since $\tilde{w}_t$ defined in (8.7) is the average of all iterates of (2.9), this convergence rate is in the ergodic sense.

9 Conclusions

Inspired by our recent work [10], we revisit the ADMM-like splitting method proposed in [12] for solving a separable convex minimization model with linear constraints and its objective function is the sum of three functions without coupled variables. We show that the proximal parameter used to regularize those splitted subproblems that are eligible for parallel computation can be any number larger than or equal 0.6, rather than 1 as required in [12]. This less restricted condition may accelerate the numerical performance for a variety of applications such as those tested in [2, 12]. Theoretically, the convergence analysis for the case with this less restricted condition of the proximal parameter is much more demanding than that in [12] because positive-indefiniteness occurs in the proximal regularization terms. We prove the convergence and worst-case convergence rate for the improved version of the method in [12] with the less restricted proximal parameter. It should be mentioned that the lower bound 0.6 is still a sufficient condition to ensure the convergence of the method in [12] and it is unknown whether 0.6 is the smallest value or not. But we show by an example that the bound can't be smaller than 0.5.

References

[1] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessary convergent, Math. Program., 155 (2016), pp. 57-79.

[2] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7) (2012), pp. 3239-3252.

[3] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I*, Springer Series in Operations Research, Springer-Verlag, 2003.

[4] R. Glowinski, Numerical methods for Nonlinear Variational Problems, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

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

[6] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China, 3(2015), pp. 391-420.

[7] B. S. He, L. Z. Liao, D. Han and H. Yang, A new inexact alternating directions method for monontone variational inequalities, Math. Program., 92 (2002), pp. 103–118.

[8] 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.

[9] B. S. He, L. S. Hou and X. M. Yuan, On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming, SIAM J. Optim., 25(4)(2015), pp. 2274-2312.

[10] B. S. He, F. Ma and X. M. Yuan, Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming, manuscript, July, 2016.

[11] B. S. He, M. Tao and X. M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, SIAM J. Optim., 22 (2012), pp. 313-340.

[12] B. S. He, M. Tao and X. M. Yuan, A splitting method for separable convex programming, IMA J. Numer. Analy., 31 (2015), pp. 394-426.

[13] B. S. He, M. Tao and X. M. Yuan, Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming, Math. Oper. Res., to appear.

[14] B. S. He, H. K. Xu and X. M. Yuan, On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM, J. Sci. Comput., 66 (2016), pp. 1204-1217.

[15] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Oper. Res. Let., 23 (1998), pp. 151-161.

[16] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, SIAM J. Numer. Analy., 50 (2012), pp. 700-709.

[17] B. S. He and X. M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Num. Math., 130 (2015), pp. 567-577.

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

[19] K. C. Kiwiel, C. H. Rosa and A. Ruszczyński, Proximal decomposition via alternating linearization, SIAM J. Optim., 9 (1999), pp. 668 – 689.

[20] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, Rev. Francaise d'Inform. Recherche Oper., 4 (1970), pp. 154-159.

[21] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

[22] R. T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Cont. Optim., 14 (1976), pp. 877-898.

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

[24] R. Tibshirani, M. Saunders, M., S. Rosset, J. Zhu and K. Knight, Sparsity and smoothness via the fused lasso, J. R. Stat. Soc., 67 (2005), pp. 91-108.

[25] X. Zhou, C. Yang and W. Yu, Moving object detection by detecting contiguous outliers in the low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2013), pp. 597-610.

[26] Z. Zhou, X. Li, J. Wright, E. J. Candes and Y. Ma, Stable principal component pursuit, Proceedings of International Symposium on Information Theory, (2010), Austin, Texas, United States.