

On the Optimal Proximal Parameter of an ADMM-like Splitting Method for Separable Convex Programming¹

Bingsheng He² and Xiaoming Yuan³

September 30, 2017

Abstract. An ADMM-based splitting method is proposed in [11] for solving convex minimization problems with linear constraints and multi-block separable objective functions; while a relatively large proximal parameter is required for theoretically ensuring the convergence. In this paper, we further study this method and find its optimal (smallest) proximal parameter. For succinctness, we focus on the case where the objective function is the sum of three functions and show that the optimal proximal parameter is 0.5. This optimal proximal parameter generates positive indefiniteness in the regularization of the subproblems and thus the convergence analysis for the improved method is more challenging than those for existing methods of the same kind in the literature, which all require positive definiteness (or positive semi-definiteness plus additional assumptions) of the regularization. We establish the convergence and estimate the convergence rate in terms of iteration complexity for the improved method with the optimal proximal parameter.

Keywords: Convex programming, splitting method, positive indefinite proximal regularization, convergence analysis.

1 Introduction

Our purpose is finding the optimal (smallest) proximal parameter for the splitting method in [11] for separable convex programming models. To expose our main idea and technique more clearly, we focus on the special convex minimization problem with linear constraints and a separable objective function that can be represented as the sum of three functions without coupled variables:

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

where $A \in \mathbb{R}^{m \times n_1}$, $B \in \mathbb{R}^{m \times n_2}$, $C \in \mathbb{R}^{m \times n_3}$; $b \in \mathbb{R}^m$; $\mathcal{X} \subset \mathbb{R}^{n_1}$, $\mathcal{Y} \subset \mathbb{R}^{n_2}$ and $\mathcal{Z} \subset \mathbb{R}^{n_3}$ are closed convex sets; and $\theta_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ ($i = 1, 2, 3$) are closed convex but not necessarily smooth functions. Such a model may arise from a concrete application in which one of the functions represents a data-fidelity term while the other two account for various regularization terms. We refer to, e.g., [16, 20, 21, 22, 23], for some applications of (1.1). The solution set of (1.1) is assumed to be nonempty throughout.

To recall the splitting method in [11] for the model (1.1), we start from the augmented Lagrangian method (ALM) that was originally proposed in [15, 18]. Let the Lagrangian and augmented Lagrangian functions of (1.1) be given, respectively, by

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

and

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

¹This is an improved version of our earlier manuscript released on Optimization Online in August 2016 (#5585) entitled “Improving an ADMM-like Splitting Method via Positive-Indefinite Proximal Regularization for Three-Block Separable Convex Minimization”, with the proved optimality of 0.5 for the parameter τ .

²Department of Mathematics, Southern University of Science and Technology, and Department of Mathematics, Nanjing University, China. This author was supported by the NSFC Grant 11471156. Email: hebma@nju.edu.cn

³Department of Mathematics, Hong Kong Baptist University, Hong Kong. This author was supported by the General Research Fund from Hong Kong Research Grants Council: 12300317. Email: xmyuan@hkbu.edu.hk

In (1.2) and (1.3), $\lambda \in \mathfrak{R}^m$ is the Lagrange multiplier; and in (1.3), $\beta > 0$ is the penalty parameter. When the three-block separable convex minimization model (1.1) is purposely regarded as a generic convex minimization model and its objective function is treated as a whole, the ALM in [15, 18] can be applied directly and the resulting iterative scheme is

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

$$(1.4b)$$

If two functions in the objective are treated together and two variables in the constraints 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

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

$$(1.5b)$$

$$(1.5c)$$

Unless the functions and/or coefficient matrices in (1.1) are special enough, direct applications of the (1.4) and the ADMM (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). Therefore, generally the three-block model (1.1) should not be treated as a one-block or two-block case and the ALM (1.4) or ADMM (1.5) should not be applied directly.

On the other hand, for specific applications of the model (1.1), functions in its objective usually have their own physical explanations and mathematical 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 Gaussian manner; and obtain the schemes

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

or

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

All the subproblems in (1.6) and (1.7) are 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. But, as shown in [1, 8], neither of the schemes (1.6) and (1.7) is necessarily convergent. Therefore, although schemes such as (1.6) and (1.7) can be easily generated, the lack of convergence may require more meticulous theoretical study and algorithmic design techniques for the three-block case (1.1). The results in [1, 8] also justify that designing augmented-Lagrangian-based splitting algorithms for the three-block case (1.1) is significantly different from that for the one- or two-block case; and they need to be discussed separately despite that there is a rich literature of the ALM and ADMM. Indeed, research on augmented-Lagrangian-based splitting methods for the three-block case (1.1) has recently received much attention from the community.

Despite of their lack of convergence, the schemes (1.6) and (1.7) may empirically work well, see, e.g., [20, 22, 23]. 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. The method in [11] is such one; its iterative scheme for (1.1) reads as

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

where the parameter μ is required to be $\mu \geq 2$ in [11]. 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 [11] by some sparse and low-rank models and image inpainting problems. Also, it was used in [2] for solving a dimensionality reduction problem on physical space.

It is easy to see that the scheme (1.8) can be rewritten as

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

with $\tau = \mu - 1$ and thus $\tau \geq 1$ as shown in [11]. 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, the terms $\frac{\tau\beta}{2}\|B(y - y^k)\|^2$ and $\frac{\tau\beta}{2}\|C(z - z^k)\|^2$ in (1.9) can be regarded as proximal regularization terms with τ as the proximal parameter.

On the other hand, with fixed β , the proximal parameter τ determines the weight of the proximal terms in the subproblems (1.9b) and its reciprocal plays the role of step size for an algorithm implemented internally to solve the subproblems (1.9b). We hence prefer smaller values of τ whenever the convergence of (1.9) can be theoretically guaranteed. As mentioned, in [11], we have shown that the condition $\tau \geq 1$ is sufficient to ensure the convergence of (1.9). While, numerically, as shown in [11] and also in [2] (see Section V, Part B, Page 3247-3248), it has been observed that values very close to 1 are preferred for τ . For example, $\mu = 2.01$, i.e., $\tau = 1.01$, was recommended in [11] and used in [2] to result in faster convergence. This raises the necessity of seeking the optimal (smallest) value of τ that can ensure the convergence of (1.9). The main purpose of this paper is to rigorously prove that the optimal value of τ is 0.5 for the method (1.9). That is, any $\tau > 0.5$ ensures the convergence of (1.9) yet any $\tau \in (0, 0.5)$ yields divergence.

Note that, 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

$$\mathcal{L}(x, y, z, \lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b) + \frac{1}{2}\|Ax + By + Cz - b\|^2; \quad (1.10)$$

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

$$\begin{cases} x^{k+1} = \arg \min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.11a) \\ \begin{cases} y^{k+1} = \arg \min\{\mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg \min\{\mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \end{cases} & (1.11b) \\ \lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & (1.11c) \end{cases}$$

The rest of this paper is organized as follows. We recall some preliminaries in Section 2. In Section 3, we show why positive indefiniteness occurs in the proximal regularization for the scheme (1.11) when $\tau > 0.5$. Then, we provide an explanation in the prediction-correction framework for (1.11) in Section 4; and focus on analyzing an important quadratic term in Section 5 that is the key for conducting convergence analysis for (1.11). The convergence of (1.11) with $\tau > 0.5$ is proved in Section 6; and the divergence of (1.11) with $\tau \in (0, 0.5)$ is shown in Section 7 by an example. We estimate the worst-case convergence rate in terms of iteration complexity for the scheme (1.11) in Section 8. Finally, we make some conclusions in Section 9.

2 Preliminaries

In this section, we recall some preliminary results for further analysis. First of all, a pair of $((x^*, y^*, z^*), \lambda^*)$ is called a saddle point of the Lagrangian function defined in (1.2) if it satisfies the inequalities

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

Or, we can rewrite these inequalities as

$$\begin{cases} x^* = \arg \min\{L(x, y^*, z^*, \lambda^*) \mid x \in \mathcal{X}\}, \\ y^* = \arg \min\{L(x^*, y, z^*, \lambda^*) \mid y \in \mathcal{Y}\}, \\ z^* = \arg \min\{L(x^*, y^*, z, \lambda^*) \mid z \in \mathcal{Z}\}, \\ \lambda^* = \arg \max\{L(x^*, y^*, z^*, \lambda) \mid \lambda \in \mathfrak{R}^m\}. \end{cases} \quad (2.1)$$

Indeed, a saddle point of the Lagrangian function defined in (1.2) can also be characterized by the following variational inequality:

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

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:

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

where

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

and

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

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

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

3 The positive indefiniteness of (1.11) with $\tau > 0.5$

In this section, we revisit the scheme (1.11) from the variational inequality perspective; and show that it can be represented as a proximal version of the direct application of ADMM (1.5) but the proximal regularization term is not positive definite for the case of $\tau > 0.5$. The positive indefiniteness of the proximal regularization excludes the application of a vast set of known convergence results in the literature of ADMM and its proximal versions, because they all require positive definiteness or semi-definiteness (plus additional assumptions on the model (1.1)) for the proximal regularization term to validate the convergence analysis.

Let us first take a look at the optimality conditions of the subproblems in (1.11). Note that the subproblem (1.11b) are specified as

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

and

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

Thus, the optimality condition of the y -subproblem in (1.11b) can be written as $y^{k+1} \in \mathcal{Y}$ and

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

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

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

Similarly, the optimality condition of the z -subproblem in (1.11b) can be written as $z^{k+1} \in \mathcal{Z}$ and

$$\theta_3(z) - \theta_3(z^{k+1}) + (z - z^{k+1})^T \begin{pmatrix} -C^T \lambda^k + C^T (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) \\ -C^T B (y^{k+1} - y^k) + \tau C^T C (z^{k+1} - z^k) \end{pmatrix} \geq 0, \quad \forall z \in \mathcal{Z}. \quad (3.2b)$$

Then, with (1.11c), we can rewrite the inequalities (3.2a) and (3.2b) as $(y^{k+1}, z^{k+1}) \in \mathcal{Y} \times \mathcal{Z}$ and

$$\begin{aligned} & \begin{pmatrix} \theta_2(y) - \theta_2(y^{k+1}) \\ \theta_3(z) - \theta_3(z^{k+1}) \end{pmatrix} + \begin{pmatrix} y - y^{k+1} \\ z - z^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -B^T \lambda^{k+1} \\ -C^T \lambda^{k+1} \end{pmatrix} + D_0 \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \end{pmatrix} \right\} \\ & \geq 0, \quad \forall (y, z) \in \mathcal{Y} \times \mathcal{Z}, \end{aligned} \quad (3.3)$$

where

$$D_0 = \begin{pmatrix} \tau B^T B & -B^T C \\ -C^T B & \tau C^T C \end{pmatrix}. \quad (3.4)$$

Obviously, D_0 is positive semidefinite and indefinite when $\tau \geq 1$ and $\tau \in (0, 1)$, respectively.

Then, it is easy to see that the scheme (1.11) can be rewritten as

$$\begin{cases} x^{k+1} = \arg \min \{ \mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X} \}, & (3.5a) \\ \begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} = \arg \min \left\{ \mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2} \left\| \begin{pmatrix} y - y^k \\ z - z^k \end{pmatrix} \right\|_{D_0}^2 \mid (y, z) \in \mathcal{Y} \times \mathcal{Z} \right\}, & (3.5b) \\ \lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), & (3.5c) \end{cases}$$

Comparing (3.5b) with (1.5b) (note that $\beta = 1$), we see that the scheme (1.11) can be symbolically represented as a proximal version of (1.5) in which the (y, z) -subproblem is proximally regularized by a proximal term. But the difficulty is that D_0 defined in (3.4) is positive indefinite when $\tau \in (0.5, 1)$. Indeed, our analysis in [11] requires $\tau \geq 1$ and thus the positive semidefiniteness of D_0 is ensured. For this case, the convergence analysis is relatively easy because it can follow some techniques used for the proximal point algorithm which is originated from [17, 19]. For the case where τ is relaxed to $\tau > 0.5$ and hence the matrix D_0 in (3.4) is positive indefinite, the analysis in [11] and other literatures is not applicable and more sophisticated techniques are needed for proving the convergence of the scheme (1.11) with $\tau > 0.5$.

4 A prediction-correction explanation of (1.11)

In this section, we show that the scheme (1.11) 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 (1.11).

In the scheme (1.11), 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

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

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

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

where $(x^{k+1}, y^{k+1}, z^{k+1})$ is the iterate generated by the scheme (1.11) from the given one (y^k, z^k, λ^k) . Using these notations, we have

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) \\ &= [\lambda^k - (Ax^{k+1} + By^k + Cz^k - b)] + B(y^k - y^{k+1}) + C(z^k - z^{k+1}) \\ &= \tilde{\lambda}^k + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k). \end{aligned} \quad (4.3)$$

Now, we interpret the optimality conditions of the subproblems in (1.11) by using the auxiliary variables \tilde{w}^k . First, ignoring some constant terms, the subproblem (1.11a) is equivalent to

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

and its optimality condition can be rewritten as

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

Using (4.3), $y^{k+1} = \tilde{y}^k$ and $z^{k+1} = \tilde{z}^k$, the inequalities (3.2a) and (3.2b) can be written as

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

and

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{-C^T \tilde{\lambda}^k + (1 + \tau)C^T C(\tilde{z}^k - z^k)\} \geq 0, \quad \forall z \in \mathcal{Z},$$

respectively. Thus, the inequality (3.3) becomes $(\tilde{y}^k, \tilde{z}^k) \in \mathcal{Y} \times \mathcal{Z}$ and

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

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

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

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:

Prediction Step

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

where

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

We call the auxiliary variable $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ as the predictor. Using (4.3), the update form (1.11c) can be represented as

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

Recall we define by v in (4.1) the essential variables for the scheme (1.11). The new essential variables of (1.11), $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$, are updated by the following scheme:

Correction Step

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

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -B & -C & I_m \end{pmatrix}. \quad (4.6b)$$

Overall, the scheme (1.11) 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). 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 (1.11). Indeed, we can further investigate the inequality (4.5) and derive a new right-hand side that is more preferred for establishing the convergence. For this purpose, let us define a matrix as

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

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

$$Q = HM. \quad (4.8)$$

In the following lemma, we further analyze the right-hand side of (4.5) and show more explicitly the difference of the proof for the convergence of (1.11) with $\tau > 0.5$ from that with $\tau \geq 1$ in [11].

Theorem 4.1 *Let $\{w^k\}$ be the sequence generated by (1.11) for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then, $\tilde{w}^k \in \Omega$ and*

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

where

$$G = Q^T + Q - M^T H M. \quad (4.10)$$

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

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

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

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

For the last term of (4.12), we have

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

Substituting (4.13) into (4.12), we get

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

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. \square

When G given in (4.10) is positive definite, as shown in [11], it is relatively easier to use the assertion (4.9) to prove the global convergence and estimate its worst-case convergence rate in terms of iteration complexity, see, e.g., [7, 14] for details and [6] (Sections 4 and 5 therein) for a tutorial proof. For the matrix G given in (4.10), since $HM = Q$ and $M^T HM = M^T Q$, we have

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

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

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

By using the notation D_0 (see (3.4)), the matrix G can be rewritten as

$$G = \begin{pmatrix} D_0 & 0 \\ 0 & 0 & I \end{pmatrix}.$$

Obviously, the proximal matrix D_0 in (3.4) can be rewritten as

$$D_0 = (\tau - 1) \begin{pmatrix} B^T B & 0 \\ 0 & C^T C \end{pmatrix} + \begin{pmatrix} B^T \\ -C^T \end{pmatrix} (B, -C). \quad (4.16)$$

Therefore, for $\tau \in (\frac{1}{2}, 1)$, G is positive indefinite because the matrix D_0 is not so. The positive indefiniteness of G is indeed the main difficulty of proving the convergence of the scheme (1.11) with $\tau > 0.5$; and we need to look into the quadratic term $(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k)$ more intensively.

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

As mentioned, the key point of proving the convergence of the scheme (1.11) with $\tau > 0.5$ is to analyze the quadratic term $(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k)$ which is not guaranteed to be positive. In this section, we focus on investigating this term and show that

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

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

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

As we shall show, all the components of the right-hand side of (5.2) in parentheses should be positive to establish the convergence and convergence rate of (1.11). It is indeed this requirement that implies our restriction of $\tau > 0.5$. We show the details in Theorem 5.1, preceded by several lemmas. Similar techniques for the convergence analysis of the ADMM are referred to, e.g. [4, 9, 10, 12].

Lemma 5.1 *Let $\{w^k\}$ be the sequence generated by (1.11) for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then we have*

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

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

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

and thus

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

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, by a manipulation, we get

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

The lemma is proved. \square

For further analysis, we will divide the crossing term $2(\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1}))$ in the right-hand side of (5.3) into two parts and give their lower bounds by quadratic terms.

Lemma 5.2 Let $\{w^k\}$ be the sequence generated by (1.11) for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then we have

$$\begin{aligned} & (\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1})) \\ & \geq (\psi(v^k, v^{k+1}) - \psi(v^{k-1}, v^k)) - 2(1 - \tau)(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2), \end{aligned} \quad (5.4)$$

where

$$\psi(v^k, v^{k+1}) = \frac{1}{2} \left(\left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 + (1 - \tau)(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) \right) \quad (5.5)$$

with

$$D = \begin{pmatrix} B^T \\ -C^T \end{pmatrix} (B, -C). \quad (5.6)$$

Proof. Recall (3.3). It holds that

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

Analogously, for the previous iteration, we have

$$\begin{aligned} (y^k, z^k) \in \mathcal{Y} \times \mathcal{Z}, & \quad \begin{pmatrix} \theta_2(y) - \theta_2(y^k) \\ \theta_3(z) - \theta_3(z^k) \end{pmatrix} + \begin{pmatrix} y - y^k \\ z - z^k \end{pmatrix}^T \\ & \quad \left\{ \begin{pmatrix} -B^T \\ -C^T \end{pmatrix} \lambda^k + D_0 \begin{pmatrix} y^k - y^{k-1} \\ z^k - z^{k-1} \end{pmatrix} \right\} \geq 0, \quad \forall (y, z) \in \mathcal{Y} \times \mathcal{Z}. \end{aligned} \quad (5.8)$$

Setting $(y, z) = (y^k, z^k)$ and $(y, z) = (y^{k+1}, z^{k+1})$ in (5.7) and (5.8), 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

$$\begin{aligned} & (\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1})) \\ & \geq \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T D_0 \left[\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} - \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right]. \end{aligned} \quad (5.9)$$

From (3.4) and (5.6) we get

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

Thus, using Cauchy-Schwarz inequality, from (5.9) we obtain

$$\begin{aligned}
& (\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1})) \\
& \geq \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T \left\{ D - (1 - \tau) \begin{pmatrix} B^T B & 0 \\ 0 & C^T C \end{pmatrix} \right\} \left[\begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} - \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right] \\
& = \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 - \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T D \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \\
& \quad - (1 - \tau) (\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) \\
& \quad + (1 - \tau) \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}^T \begin{pmatrix} B^T B & 0 \\ 0 & C^T C \end{pmatrix} \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \\
& \geq \frac{1}{2} \left\| \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix} \right\|_D^2 - \frac{1}{2} \left\| \begin{pmatrix} y^{k-1} - y^k \\ z^{k-1} - z^k \end{pmatrix} \right\|_D^2 \\
& \quad - \frac{3}{2} (1 - \tau) (\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) \\
& \quad - \frac{1}{2} (1 - \tau) (\|B(y^{k-1} - y^k)\|^2 + \|C(z^{k-1} - z^k)\|^2), \tag{5.10}
\end{aligned}$$

where the last inequality is because of the Cauchy-Schwarz inequality. Manipulating the right-hand side of (5.10) recursively and using the notation of $\psi(\cdot, \cdot)$ (see (5.5)), we get (5.4) and the lemma is proved. \square

In addition to (5.4), we need to the term $(\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1}))$ by another quadratic terms. This is done by the following lemma.

Lemma 5.3 *Let $\{w^k\}$ be the sequence generated by (1.11) for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then, for $\tau \in (0.5, 1)$, we have*

$$\begin{aligned}
& (\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1})) \\
& \geq -\tau (\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) - \left(\frac{3}{2} - \tau\right) \|\lambda^k - \lambda^{k+1}\|^2. \tag{5.11}
\end{aligned}$$

Proof. Setting $\delta = \tau - \frac{1}{2}$. Because $\tau \in (0.5, 1)$, we have $\delta \in (0, 0.5)$. Using the Cauchy-Schwarz inequality twice, we get

$$\begin{aligned}
& (\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1})) \\
& \geq -\frac{1}{4(1 - \delta)} \|B(y^k - y^{k+1}) + C(z^k - z^{k+1})\|^2 - (1 - \delta) \|\lambda^k - \lambda^{k+1}\|^2 \\
& \geq -\frac{1}{2(1 - \delta)} (\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) - (1 - \delta) \|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned}$$

Since $\delta \in (0, 0.5)$, we have

$$\frac{1}{2(1 - \delta)} < \frac{1}{2} + \delta,$$

and thus

$$\begin{aligned}
& (\lambda^k - \lambda^{k+1})^T (B(y^k - y^{k+1}) + C(z^k - z^{k+1})) \\
& \geq -\left(\frac{1}{2} + \delta\right) (\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) - (1 - \delta) \|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned}$$

Substituting $\delta = \tau - \frac{1}{2}$ in the above inequality, we get (5.11) and the lemma is proved. \square

Recall that we want to bound the quadratic term $(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k)$ in the form of (5.2). Our previous analysis enables us to achieve it; and this is the basis of the convergence analysis to be shown soon.

Lemma 5.4 *Let $\{w^k\}$ be the sequence generated by (1.11) for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then, for $\tau \in (0.5, 1)$, we have*

$$(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \geq (\psi(v^k, v^{k+1}) - \psi(v^{k-1}, v^k)) + \varphi(v^k, v^{k+1}), \quad (5.12)$$

where $\psi(v^k, v^{k+1})$ is defined in (5.5) and

$$\varphi(v^k, v^{k+1}) = \left(\tau - \frac{1}{2}\right) (2\|B(y^k - y^{k+1})\|^2 + 2\|C(z^k - z^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2). \quad (5.13)$$

Proof. Substituting (5.4) and (5.11) into (5.3), we get

$$\begin{aligned} & (v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \\ & \geq (1 + \tau)\|B(y^k - y^{k+1})\|^2 + (1 + \tau)\|C(z^k - z^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 \\ & \quad + (\psi(v^k, v^{k+1}) - \psi(v^{k-1}, v^k)) - 2(1 - \tau)(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) \\ & \quad - \tau(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) - \left(\frac{3}{2} - \tau\right)\|\lambda^k - \lambda^{k+1}\|^2 \\ & = (\psi(v^k, v^{k+1}) - \psi(v^{k-1}, v^k)) \\ & \quad + (2\tau - 1)(\|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2) + \left(\tau - \frac{1}{2}\right)\|\lambda^k - \lambda^{k+1}\|^2. \end{aligned}$$

The assertion of this lemma follows from the definition of $\varphi(v^k, v^{k+1})$ directly. \square

Finally, substituting (5.12) into (4.9), we obtain the following theorem directly. This theorem plays a fundamental role in proving the convergence of (1.11) with $\tau > 0.5$.

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by (1.11) for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then we have*

$$\begin{aligned} \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) & \geq \frac{1}{2}(\|v - v^{k+1}\|_H^2 + \psi(v^k, v^{k+1})) - \frac{1}{2}(\|v - v^k\|_H^2 + \psi(v^{k-1}, v^k)) \\ & \quad + \frac{1}{2}\varphi(v^k, v^{k+1}), \quad \forall w \in \Omega, \end{aligned} \quad (5.14)$$

where $\psi(v^k, v^{k+1})$ and $\varphi(v^k, v^{k+1})$ are defined in (5.5) and (5.13), respectively.

6 Convergence

As mentioned, proving the convergence of the scheme (1.11) with $\tau > 0.5$ essentially relies on Theorem 5.1. With Theorem 5.1, the remaining part of the proof is subroutine. In this section, we present the convergence of the scheme (1.11) with $\tau > 0.5$; a lemma is first proved to show the contraction property of the sequence generated by (1.11).

Lemma 6.1 *Let $\{w^k\}$ be the sequence generated by (1.11) with $\tau > 0.5$ for the problem (1.1). Then we have*

$$(\|v^{k+1} - v^*\|_H^2 + \psi(v^k, v^{k+1})) \leq (\|v^k - v^*\|_H^2 + \psi(v^{k-1}, v^k)) - \varphi(v^k, v^{k+1}), \quad (6.1)$$

where $\psi(v^k, v^{k+1})$ and $\varphi(v^k, v^{k+1})$ are defined in (5.5) and (5.13), respectively.

Proof. Setting $w = w^*$ in (5.14) and using

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

we obtain the assertion (6.1) immediately. \square

Theorem 6.1 *Let $\{w^k\}$ be the sequence generated by (1.11) with $\tau > 0.5$ for the problem (1.1). Then the sequence $\{v^k\}$ converges to a $v^\infty \in \mathcal{V}^*$ when B and C are both full column rank.*

Proof. First, it follows from (6.1) and (5.13) that

$$\begin{aligned} & \left(\tau - \frac{1}{2}\right) \left(2\|B(y^k - y^{k+1})\|^2 + 2\|C(z^k - z^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\right) \\ & \leq (\|v^k - v^*\|_H^2 + \psi(v^{k-1}, v^k)) - (\|v^{k+1} - v^*\|_H^2 + \psi(v^k, v^{k+1})). \end{aligned}$$

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

$$\sum_{k=1}^{\infty} \left\{ \left(\tau - \frac{1}{2}\right) \left(2\|B(y^k - y^{k+1})\|^2 + 2\|C(z^k - z^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\right) \right\} \leq \|v^1 - v^*\|_H^2 + \psi(v^0, v^1)$$

and thus

$$\lim_{k \rightarrow \infty} \|B(y^k - y^{k+1})\|^2 + \|C(z^k - z^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2 = 0. \quad (6.2)$$

For an arbitrarily fixed $v^* \in \mathcal{V}^*$, it follows from (6.1) that, for any $k > 1$, we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 + \psi(v^{k-1}, v^k) \leq \|v^1 - v^*\|_H^2 + \psi(v^0, v^1). \quad (6.3)$$

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 (6.3) that

$$\|v^{k+1} - v^\infty\|_H^2 \leq \|v^k - v^\infty\|_H^2 + \psi(v^{k-1}, v^k). \quad (6.4)$$

Together with (6.2), 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. \square

Remark 6.1 *Note that the convergence of (1.11) with $\tau > 0.5$ in terms of the sequence $\{v^k\}$ is proved in Theorem 6.1 under the assumption that both B and C are full column rank. Without this assumption, weaker convergence results in terms of $\{By^k, Cz^k\}$ can be derived. We refer to Section 6 in [11] for details.*

7 The optimality of $\tau = 0.5$

We have proved the convergence of (1.11) with $\tau > 0.5$; the key is sufficiently ensuring the non-negativeness of the coefficients in the right-hand side of (5.14). In this section, we show by an example that any $\tau \in (0, 0.5)$ may yield divergence of (1.11). Hence, $\tau = 0.5$ is the watershed, or optimal value, to ensure the convergence of (1.11).

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

$$\min\{x + \frac{\epsilon}{2}y^2 + \frac{\epsilon}{2}z^2 \mid x + y + z = 0, x \in \{0\}, y \in \Re, z \in \Re\}, \quad (7.1)$$

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

The augmented Lagrangian function of the problem (7.1) 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 (1.11) for (7.1) is

$$\begin{cases} x^{k+1} &= \arg \min \{ \mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \{0\} \}, \\ y^{k+1} &= \arg \min \{ \mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|y - y^k\|^2 \mid y \in \mathfrak{R} \}, \\ z^{k+1} &= \arg \min \{ \mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|z - z^k\|^2 \mid z \in \mathfrak{R} \}, \\ \lambda^{k+1} &= \lambda^k - (x^{k+1} + y^{k+1} + z^{k+1}). \end{cases} \quad (7.2)$$

Since $\mathcal{X} = \{0\}$, we have $x^{k+1} \equiv 0$. Ignoring constant terms in the objective function of the subproblems, the recursion (7.2) becomes

$$\begin{cases} x^{k+1} &\equiv 0, \\ y^{k+1} &= \arg \min \{ \frac{\epsilon}{2}y^2 - y^T \lambda^k + \frac{1}{2}\|y + z^k\|^2 + \frac{\tau}{2}\|y - y^k\|^2 \mid y \in \mathfrak{R} \}, \\ z^{k+1} &= \arg \min \{ \frac{\epsilon}{2}z^2 - z^T \lambda^k + \frac{1}{2}\|y^k + z\|^2 + \frac{\tau}{2}\|z - z^k\|^2 \mid z \in \mathfrak{R} \}, \\ \lambda^{k+1} &= \lambda^k - (y^{k+1} + z^{k+1}). \end{cases} \quad (7.3)$$

Further, it follows from (7.3) that

$$\begin{cases} \epsilon y^{k+1} - \lambda^k + (y^{k+1} + z^k) + \tau(y^{k+1} - y^k) = 0, \\ \epsilon z^{k+1} - \lambda^k + (z^{k+1} + y^k) + \tau(z^{k+1} - z^k) = 0, \\ \lambda^{k+1} = \lambda^k - (y^{k+1} + z^{k+1}). \end{cases}$$

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

$$\begin{cases} (\tau + 1 + \epsilon)y^{k+1} = \tau y^k - z^k + \lambda^k, \\ (\tau + 1 + \epsilon)z^{k+1} = -y^k + \tau z^k + \lambda^k, \\ \lambda^{k+1} = \lambda^k - (y^{k+1} + z^{k+1}). \end{cases} \quad (7.4)$$

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

$$\begin{cases} \frac{3}{2}y^{k+1} = (\tau - 1)y^k + \lambda^k, \\ \lambda^{k+1} = \lambda^k - 2y^{k+1}. \end{cases} \quad (7.5)$$

With elementary manipulations, we obtain

$$\begin{cases} y^{k+1} = \frac{-2(1-\tau)}{3}y^k + \frac{2}{3}\lambda^k, \\ \lambda^{k+1} = \frac{4(1-\tau)}{3}y^k + \frac{-1}{3}\lambda^k, \end{cases} \quad (7.6)$$

which can be written as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = P(\tau) \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} \quad \text{with} \quad P(\tau) = \frac{1}{3} \begin{pmatrix} -2(1-\tau) & 2 \\ 4(1-\tau) & -1 \end{pmatrix}. \quad (7.7)$$

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

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

and

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

Certainly, the scheme (7.6) 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

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

Hence, the scheme (1.11) is not necessarily convergent for any $\tau \in (0, 0.5)$.

8 Convergence rate

In this section, we derive a worst-case $O(1/t)$ convergence rate in terms of iteration complexity for the scheme (1.11) with $\tau > 0.5$, where t is the iteration counter. Hence, although the condition $\tau \geq 1$ in [11] is now relaxed to $\tau > 0.5$, the same convergence rate result in [11] remains valid for the scheme (1.11). Similar analysis is referred to [11, 13].

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

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

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

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

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

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

where

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

In the following, we show that based on the first t iterates generated by the scheme (1.11) with $\tau > 0.5$, we can find an approximate solution of (2.3), denoted by $\tilde{w} \in \Omega$, such that

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \{ \theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \} \leq \epsilon, \quad (8.8)$$

where $\epsilon = O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate is established for the scheme (1.11) with $\tau > 0.5$. Theorem 5.1 is still the basis for the analysis in this section.

Theorem 8.1 Let $\{w^k\}$ be the sequence generated by (1.11) with $\tau > 0.5$ for the problem (1.1) and \tilde{w}^k be defined by (4.2). Then for any integer t , we have

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2t} \{ \|v - v^1\|_H^2 + \psi(v^0, v^1) \}, \quad (8.9)$$

where

$$\tilde{w}_t = \frac{1}{t} \left(\sum_{k=1}^t \tilde{w}^k \right) \quad (8.10)$$

and $\psi(v^0, v^1)$ is defined in (5.5) and thus

$$\psi(v^0, v^1) = \frac{1}{2} \left(\left\| \begin{array}{c} y^0 - y^1 \\ z^0 - z^1 \end{array} \right\|_D^2 + (1 - \tau) \left(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2 \right) \right).$$

Proof. First, it follows from (5.14) that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 + \psi(v^k, v^{k+1})) - \frac{1}{2} (\|v - v^k\|_H^2 + \psi(v^{k-1}, v^k)).$$

Thus, we have

$$\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) + \frac{1}{2} (\|v - v^{k+1}\|_H^2 + \psi(v^k, v^{k+1})) \leq \frac{1}{2} (\|v - v^k\|_H^2 + \psi(v^{k-1}, v^k)). \quad (8.11)$$

Summarizing the inequality (8.11) over $k = 1, 2, \dots, t$, we obtain

$$\sum_{k=1}^t \theta(\tilde{u}^k) - t\theta(u) + \left(\sum_{k=1}^t \tilde{w}^k - tw \right)^T F(w) \leq \frac{1}{2} (\|v - v^1\|_H^2 + \psi(v^0, v^1))$$

and thus

$$\frac{1}{t} \left(\sum_{k=1}^t \theta(\tilde{u}^k) \right) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2t} (\|v - v^1\|_H^2 + \psi(v^0, v^1)). \quad (8.12)$$

Since $\theta(u)$ is convex and

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

we have that

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

Substituting it into (8.12), the assertion of this theorem follows directly. \square

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

$$d := \sup \left\{ \|v - v^1\|_H^2 + \frac{1}{2} \left\| \begin{array}{c} y^0 - y^1 \\ z^0 - z^1 \end{array} \right\|_D^2 + \frac{1 - \tau}{2} \left(\|B(y^0 - y^1)\|^2 + \|C(z^0 - z^1)\|^2 \right) \mid w \in \mathcal{D}_{(\tilde{w})} \right\}$$

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 (1.11), the point $\tilde{w}_t \in \Omega$ defined in (8.10) satisfies

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

which means \tilde{w}_t is an approximate solution of $\text{VI}(\Omega, F, \theta)$ with an accuracy $O(1/t)$ (recall (8.8)). That is, a worst-case $O(1/t)$ convergence rate is established for the scheme (1.11) with $\tau > 0.5$. Since \tilde{w}_t defined in (8.10) is the average of all iterates of (1.11), this convergence rate is in the ergodic sense.

9 Conclusions

We revisit the splitting method proposed in [11] for solving separable convex minimization models; and show that its optimal proximal parameter is 0.5 when the objective function is the sum of three functions. This optimal proximal parameter offers the possibility of immediate numerical acceleration; which can be easily verified by the examples tested in [2, 11] and others (for succinctness, we omit the presentation of numerical results). Meanwhile, more sophisticated techniques are required for the convergence analysis because this optimal proximal parameter generates positive indefiniteness in the proximal regularization term as well. We establish the convergence and estimate the worst-case convergence rate in terms of iteration complexity for the improved version of the method in [11] with the optimal proximal parameter. This work is inspired by the analysis in our recent work [9, 10] for the augmented Lagrangian method and alternating direction method of multipliers.

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), 3239-3252, 2012.
- [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 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.
- [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, 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.
- [8] 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.
- [9] B. S. He, F. Ma and X. M. Yuan, Indefinite proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems, manuscript, 2017.
- [10] B. S. He, F. Ma and X. M. Yuan, Optimal linearized alternating direction method of multipliers for convex programming, manuscript, 2017.
- [11] 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.

- [12] 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.
- [13] 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.
- [14] 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.
- [15] M. R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.*, 4 (1969), pp. 303-320.
- [16] K. C. Kiwiel, C. H. Rosa and A. Ruszczyński, Proximal decomposition via alternating linearization, *SIAM J. Optim.*, 9 (1999), pp. 668 - 689.
- [17] B. Martinet, Regularisation, d'inéquations variationnelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, 4 (1970), pp. 154-159.
- [18] 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.
- [19] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, 14 (1976), pp. 877-898.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.