Linearized Alternating Direction Method of Multipliers via Positive-Indefinite Proximal Regularization for Convex Programming

Bingsheng He[1] **Feng Ma**[2] **Xiaoming Yuan** [3]

July 31, 2016

Abstract. The alternating direction method of multipliers (ADMM) is being widely used for various convex minimization models with separable structures arising in a variety of areas. In the literature, the proximal version of ADMM which allows ADMM's subproblems to be proximally regularized has been well studied. Particularly the linearized version of ADMM can be yielded when the proximal terms are appropriately chosen; and for some applications it could alleviate an ADMM subproblem as easy as estimating the proximity operator of a function in the objective. This feature is significant in easing the numerical implementation and it makes the linearized version of ADMM very popular in a broad spectrum of application domains. To ensure the convergence of the proximal version of ADMM, however, existing results conventionally need to guarantee the positive definiteness of the corresponding proximal matrix. For some cases, this essentially results in small step sizes (or, over-regularization effectiveness) for the subproblems and thus inevitably decelerates the overall convergence speed of the linearized version of ADMM. In this paper, we investigate the possibility of relaxing the positive definiteness requirement of the proximal version of ADMM and show affirmatively that it is not necessary to ensure the positive definiteness of the proximal matrix. A new linearized ADMM with larger step sizes is thus proposed via choosing a positive-indefinite proximal regularization term. The global convergence of the new linearized ADMM is proved; and its worst-case convergence rate measured by the iteration complexity is also established. Since the ADMM can be regarded as a splitting version of the augmented Lagrangian method (ALM), a byproduct of our analysis is a new linearized version of ALM generated by choosing a positive-indefinite proximal regularization term for its subproblems.

Keywords: Convex programming, augmented Lagrangian method, alternating direction method of multipliers, positive indefinite proximal, convergence rate

1 Introduction

We consider the convex minimization problem with linear constraints and an objective function in form of the sum of two functions without coupled variables

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

where $\theta_1(x) : \Re^{n_1} \to \Re$ and $\theta_2(y) : \Re^{n_2} \to \Re$ are convex (but not necessarily smooth) functions, $A \in \Re^{m\times n_1}$ and $B \in \Re^{m\times n_2}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$ and $\mathcal{Y} \subset \Re^{n_2}$ are given closed convex sets. The model (1.1) is general enough to capture a variety of applications; a particular case arising often in many scientific computing areas is where one function in its objective represents a data fidelity term while the other is a regularization term. Throughout, the solution set of (1.1) is assumed to be nonempty.

Let the augmented Lagrangian function of (1.1) be

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

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

[2]The High-Tech Institute of Xi'an, Xi'an, 710025, Shaanxi, China. Email: mafengnju@gmail.com

[3]Corresponding author. 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

with $\lambda \in \Re^m$ the Lagrange multiplier and $\beta > 0$ a penalty parameter. Then, a benchmark solver for (1.1) is the alternating direction method of multipliers (ADMM) that was originally proposed in [10]. With a given iterate (y^k, λ^k), the ADMM generates a new iterate $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ via the scheme

$$\text{(ADMM)} \quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^2(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(1.3a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^2(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(1.3b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.3c)} \end{cases}$$

A meaningful advantage of the ADMM is that the functions θ_1 and θ_2 are treated individually in its iterations and the subproblems in (1.3) are usually much easier than the original problem (1.1). We refer the reader to, e.g. [2, 8, 9, 11, 13], for some earlier study on the ADMM in the partial differential equations community. Recently, the ADMM has found successful applications in a broad spectrum of fields such as image processing, statistical learning, computer vision, wireless network, distributed network, etc. We refer to [1, 5, 12] for some review papers of the ADMM. Note that we assume the penalty parameter β is fixed throughout our discussion.

Among various research spotlights of the ADMM in the literature, a particular one is the investigation of how to solve ADMM's subproblems (i.e., the problems (1.3a) and (1.3b)) more efficiently for different scenarios where the functions θ_1 and θ_2, and/or the coefficient matrices A and B may have some special properties or structures that can help us better design application-tailored specific algorithms based on the prototype ADMM scheme (1.3); while theoretically the convergence should be still guaranteed. This principle accounts for the importance of how to effectively apply the ADMM to many specific applications arising in various areas. To further elaborate, let us take a closer look at the subproblem (1.3b) that can be written respectively as

$$y^{k+1} = \arg\min\{\theta_2(y) + \frac{\beta}{2}\|By + (Ax^{k+1} - b - \frac{1}{\beta}\lambda^k)\|^2 \mid y \in \mathcal{Y}\}. \tag{1.4}$$

Certainly, how to solve this subproblem depends on the function $\theta_2(y)$, matrix B and set $\mathcal{Y}$. We refer to, e.g., [4, 16, 26], for some detailed discussions on the generic case where the function, matrix and set are general and a solution of (1.4) can only be approximated by certain iterative processes. On the other hand, for concrete applications, the function, matrix and set in the subproblem (1.4) may be special enough to inspire us to consider more efficient ways to tackle it. When B is not an identity matrix and $\mathcal{Y} = \Re^{n_2}$, the subproblem (1.4) is specified as

$$y^{k+1} = \arg\min\{\theta_2(y) + \frac{\beta}{2}\|By + (Ax^{k+1} - b - \frac{1}{\beta}\lambda^k)\|^2 \mid y \in \Re^{n_2}\}. \tag{1.5}$$

We can further linearize the quadratic term $\|By + (Ax^{k+1} - b - \frac{1}{\beta}\lambda^k)\|^2$ in (1.5) and alleviate the subproblem (1.5) as an easier one

$$y^{k+1} = \arg\min\{\theta_2(y) + \frac{r}{2}\|y - (y^k + \frac{1}{r}q_k)\|^2 \mid y \in \Re^{n_2}\}, \tag{1.6}$$

where

$$q_k = B^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)] \tag{1.7}$$

is a constant vector and $r > 0$ is a constant. That is, the linearized subproblem (1.6) reduces to estimating the proximity operator of $\theta_2(y)$ given by

$$\text{prox}_{\gamma\theta_2}(y) := \text{argmin}\Big\{\theta_2(z) + \frac{1}{2\gamma}\|z - y\|^2 \mid z \in \Re^{n_2}\Big\}, \tag{1.8}$$

with $\gamma > 0$. A representative case where $\theta_2(y) = \|y\|_1$ arises often in sparsity-driven applications such as the problems of basis pursuit, total variational image restoration and variable selection for high-dimensional datasets. For this case, the proximity operator of $\|y\|_1$ has a closed-form that can be represented by the so-called shrinkage operator whose definition is given component-wisely by

$$(T_\gamma(y))_i := (|y_i| - \gamma)_+ \mathrm{sign}(y_i)$$

with $\gamma > 0$. Hence, replacing the original ADMM subproblem (1.3b) with its linearized surrogate, we obtain the linearized version of ADMM:

$$\text{(Linearized ADMM)} \quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}^2_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.9a) \\ y^{k+1} = \arg\min\{\theta_2(y) + \frac{r}{2}\|y - (y^k + \frac{1}{r}q_k)\|^2 \ y \in \mathcal{Y}\}, & (1.9b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (1.9c) \end{cases}$$

where q_k is given in (1.7). The scheme (1.9) has been widely used in areas such as compressive sensing statistical learning, image processing, etc. We refer to, e.g. [25, 28, 29, 30], for some references. Note that the linearized subproblem (1.9b), after ignoring some constants in the objective function, can be further rewritten as

$$y^{k+1} = \arg\min\{\mathcal{L}^2_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|^2_D \mid y \in \mathcal{Y}\}, \tag{1.10}$$

with $D \in \Re^{n_2 \times n_2} = rI_{n_2} - \beta B^T B$ and the quadratic term $\frac{1}{2}\|y - y^k\|^2_D$ serves as a proximal regularization term. Indeed, the more general proximal version of ADMM can be written as

$$\text{(PD-ADMM)} \quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}^2_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.11a) \\ y^{k+1} = \arg\min\{\mathcal{L}^2_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|^2_D \mid y \in \mathcal{Y}\}, & (1.11b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (1.11c) \end{cases}$$

where $D \in \Re^{n_2 \times n_2}$ is positive definite. It is easy to see that the linearized version of ADMM (1.9) is just a special case of the general proximal ADMM (1.11) with the particular positive-definite proximal regularization:

$$D = rI_{n_2} - \beta B^T B \quad \text{and} \quad r > \beta\|B^T B\|. \tag{1.12}$$

Since for many applications it suffices to linearize one subproblem of the ADMM (1.3), without loss of generality, we just discuss the case where only the y-subproblem is linearized/proximally regularized in (1.11). Technically, it is still possible to consider the case where both the subproblems are linearized/proximally regularized, see, e.g., [16].

As well shown in the literature, the positive definiteness of the proximal matrix D is crucial for ensuring the convergence of the proximal ADMM (1.11). This can also be easily observed by our analysis in Section 4. For the case where $\|B^T B\|$ is large (see [14] for such an application in image processing), the constant r in the linearized version of ADMM (1.11) with D chosen by (1.12) is also forced to be large and thus tiny step sizes inevitably occur. The overall convergence speed of (1.11) thus may be substantially decelerated. A large value of r in (1.12) can be regarded as a over-regularization for the subproblem (1.11b) because the proximal term has a too high weight in the objective function and thus it deviates the original objective function in (1.3b) too much. A practical strategy for implementing the linearized ADMM (1.9) is to choose r's value larger than but extremely close to the lower bound $\beta\|B^T B\|$, as empirically used in, e.g., [6, 18]. Therefore, for the linearized ADMM (1.9), there is a dilemma that theoretically the constant r should be large enough to ensure the convergence while numerically smaller values of r are preferred.

An important question is thus how to further relax the positive-definiteness requirement of the proximal matrix D in (1.11b) while the convergence can be still theoretically ensured — without additional assumptions on the model (1.1). The main purpose of this paper is to answer this question.

More specifically, we show that the matrix D in (1.11b) could be positive-indefinite, thus its positive-definiteness is not necessary. To remain the possibility of alleviating this subproblem as estimating the proximity operator of $\theta_2(y)$ for some applications, we are also interested in the particular choice of (1.12), i.e., the linearized version of ADMM, but with a smaller value of r. That is, we investigate the choice of

$$D_0 = \tau r I_{n_2} - \beta B^T B \quad \text{with} \quad r > \beta \|B^T B\| \tag{1.13}$$

and $0 < \tau < 1$ to replace the positive definite matrix D in (1.11b). Indeed, as analyzed in Section 4, we can choose any value of $\tau \in [0.8, 1)$ and 0.8 is an "optimal" choice in sense of the inequality (4.21) to be proved. Thus, we propose the scheme

$$\text{(IP-LADMM)} \quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}^2_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (1.14a) \\ y^{k+1} = \arg\min\{\mathcal{L}^2_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|^2_{D_0} \mid y \in \mathcal{Y}\}, & (1.14b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (1.14c) \end{cases}$$

where D_0 is given by (1.13) with $\tau = 0.8$. In this case, ignoring some constants, the subproblem (1.14b) can be written as

$$y^{k+1} = \arg\min\{\theta_2(y) + \frac{\tau r}{2}\|y - (y^k + \frac{1}{\tau r} q_k)\|^2 \mid y \in \mathcal{Y}\},$$

with q_k given in (1.7), which also reduces to estimating the proximity operator of $\theta_2(y)$ if $\mathcal{Y} = \Re^{n_2}$. Clearly the matrix D_0 defined in (1.13) with $\tau = 0.8$ is not necessarily positive-definite. Thus, the new linearized ADMM (1.14) can be generated by the proximal ADMM (1.11) but with a positive-indefinite proximal regularization term. We also call the scheme (1.14) more succinctly as "indefinite proximal linearized ADMM", or abbreviated as "IP-LADMM". Note that the subproblem (1.14b) is still convex even though the proximal matrix D_0 is positive indefinite. Also, we slightly abuse the notation $\|y\|^2_{D_0} := y^T D_0 y$ when D_0 is not positive definite.

The rest of this paper is organized as follows. We first summarize some preliminary results in Section 2. Then we reformulate the IP-LADMM (1.14) in a prediction-correction framework in Section 3 and discuss how to determine the value of τ in Section 4. Then, in Section 5 we provide some insights on how to choose τ and show by an example that it is not possible to shrink the value of τ as small as 0.75. In Section 6, the convergence of (1.14) is proved. Its worst-case convergence rate measured by the iteration complexity is established in Section 7. In Section 8, we extend our analysis to the augmented Lagrangian method (ALM) proposed in [23, 27] and propose a new linearized ALM via positive-indefinite proximal regularization, followed by some necessary details for establishing its convergence. Some conclusions are draw in Section 9.

2 Preliminaries

In this section, we recall some preliminaries and state some simple results that will be used in our analysis.

Let the Lagrangian function of (1.1) defined on $\mathcal{X} \times \mathcal{Y} \times \Re^m$ be

$$L^2(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b)$$

with $\lambda \in \Re^m$ the Lagrange multiplier. A pair $((x^*, y^*), \lambda^*)$ is called a saddle point of the Lagrangian function if it satisfies

$$L^2_{\lambda \in \Re^m}(x^*, y^*, \lambda) \leq L^2(x^*, y^*, \lambda^*) \leq L^2_{x \in \mathcal{X}, y \in \mathcal{Y}}(x, y, \lambda^*).$$

We can rewrite them as the variational inequalities:

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, \quad \forall\, y \in \mathcal{Y}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m, \end{cases} \tag{2.1}$$

or in the more compact form:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.2a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \tag{2.2b}$$

$$\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{2.2c}$$

We denote by Ω^* the solution set of (2.2). Note that the operator F defined in (2.2b) is affine with a skew-symmetric matrix. Thus we have

$$(w - \bar{w})^T(F(w) - F(\bar{w})) = 0, \quad \forall\, w, \bar{w}. \tag{2.3}$$

The following lemma will be frequently used later; its proof is elementary and thus omitted.

Lemma 2.1 *Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions. Assume that f is differentiable and the solution set of the problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then we have*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}, \tag{2.4a}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.4b}$$

3 A prediction-correction reformulation of the IP-LADMM (1.14)

In this section, we revisit the IP-LADMM (1.14) from the variational inequality perspective and show that it can be rewritten as a prediction-correction framework. The prediction-correction reformulation helps us discern the main difficulty in the convergence proof and plays a pivotal role in our analysis.

As mentioned in [1], for the ADMM schemes (1.3) (also (1.14)), only (y^k, λ^k) is used to generate the next iteration and x^k is just in an "intermediate" role in the iteration. This is also why the convergence result of ADMM is established in terms of only the variables (y, λ) in the literature, see, e.g., [1, 3, 10, 16, 19, 20]. Thus, the variables x and (y, λ) are called intermediate and essential variables, respectively. To distinguish the essential variables, parallel to the notation in (2.2), we further denote the following notation

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad \mathcal{V} = \mathcal{Y} \times \Re^m, \quad \text{and} \quad \mathcal{V}^* = \{(y^*, \lambda^*) \,|\, (x^*, y^*, \lambda^*) \in \Omega^*\}. \tag{3.1}$$

First, from the optimality conditions of the subproblems (1.14a) and (1.14b), we respectively have

$$x^{k+1} \in \mathcal{X},\ \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)\} \geq 0,\ \forall x \in \mathcal{X}, \tag{3.2}$$

and

$$y^{k+1} \in \mathcal{Y},\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T \begin{pmatrix} -B^T\lambda^k + \beta B^T(Ax^{k+1}+By^{k+1}-b) \\ +D_0(y^{k+1}-y^k) \end{pmatrix} \geq 0,\ \forall y \in \mathcal{Y}. \quad (3.3)$$

Recall that $D_0 = \tau r I_{n_2} - \beta B^T B$ (see (1.13)). The inequality (3.3) can be further written as

$$y^{k+1} \in \mathcal{Y},\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T \begin{pmatrix} -B^T\lambda^k + \beta B^T(Ax^{k+1}+By^{k}-b) \\ +\tau r(y^{k+1}-y^k) \end{pmatrix} \geq 0,\ \forall y \in \mathcal{Y}. \quad (3.4)$$

With the given (y^k, λ^k), let (x^{k+1}, y^{k+1}) be the output of the IP-LADMM (1.14). If we rename them as $\tilde{x}^k = x^{k+1}$ and $\tilde{y}^k = y^{k+1}$, respectively, and further define an auxiliary variable

$$\tilde{\lambda}^k := \lambda^k - \beta(Ax^{k+1} + By^k - b), \quad (3.5)$$

then accordingly we have $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ given by

$$\tilde{x}^k = x^{k+1},\ \ \tilde{y}^k = y^{k+1},\ \ \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), \quad (3.6)$$

and $\tilde{v}^k = (\tilde{y}^k, \tilde{\lambda}^k)$. Therefore, the inequalities (3.2) and (3.4) can be rewritten respectively as

$$\tilde{x}^k \in \mathcal{X},\quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0,\ \ \forall\, x \in \mathcal{X}, \quad (3.7a)$$

and

$$\tilde{y}^k \in \mathcal{Y},\quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big(-B^T\tilde{\lambda}^k + \tau r(\tilde{y}^k - y^k)\big) \geq 0,\ \ \forall\, y \in \mathcal{Y}. \quad (3.7b)$$

Note that $\tilde{\lambda}^k$ defined in (3.5) can be also written as the variational inequality

$$\tilde{\lambda}^k \in \Re^m,\ \ (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0,\ \ \forall \lambda \in \Re^m. \quad (3.7c)$$

Thus, using the notation of (2.2), we can rewrite the inequalities (3.7a)-(3.7c) as the variational inequality:

$$\tilde{w}^k \in \Omega,\ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\quad \forall\, w \in \Omega, \quad (3.8a)$$

where

$$Q = \begin{pmatrix} \tau r I_{n_2} & 0 \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad (3.8b)$$

Then, using the notation in (3.6), we further have

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

and

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

Recall $y^{k+1} = \tilde{y}^k$ and the notation in (3.1). The essential variables updated by the IP-LADMM (1.14) are given by

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad (3.9a)$$

where

$$M = \begin{pmatrix} I_{n_2} & 0 \\ -\beta B & I_m \end{pmatrix}. \quad (3.9b)$$

Overall, the iteration of IP-LADMM (1.14) can be conceptually explained by a two-stage manner, first generating a predictor satisfying the variational inequality (3.8) and then correcting it via the correction step (3.9). We would emphasize that this prediction-correction reformulation only serves for the theoretical analysis and there is no need to decompose the iterative scheme into these two stages separately when implementing the IP-LADMM (1.14). Indeed, we see that $\tilde{w}^k$ satisfying (3.8) is not a solution point of the variational inequality (2.2) unless it ensures $(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k)=0$ for all $w \in \Omega$ and this fact inspires us to intensively analyze the term $(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k)$ in convergence analysis.

4 How to determine τ

We have mentioned that it is interesting to consider the IP-LADMM (1.14) but the proximal matrix D_0 is given by (1.13) with $0<\tau<1$, instead of $\tau=1$ as (1.12). That is, we focus on shrinking the value of τ for the linearized ADMM (1.9) that can be generated by the proximal ADMM (1.11) with the positive-definite proximal regularization (1.12). Then, the central problem is investigating the restriction of τ can ensure the convergence of the IP-LADMM (1.14). In this section, we focus on the predictor $\tilde{w}^k$ characterized by (3.8) and conduct a more elaborated analysis; some inequalities regarding $\tilde{w}^k$ will be derived and they are the clue for determining the value of τ. These inequalities are also essential for the convergence analysis of the IP-LADMM (1.14). Thus, the results in this section are also the preparation of the main convergence results to be established in Sections 6 and 7.

First of all, for any given positive constants τ, r and β, we define a matrix

$$H=\begin{pmatrix} \tau r I_{n_2} & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.1}$$

Obviously, H is positive definite. For the matrices Q and M defined respectively in (3.8) and (3.9), we have

$$HM=Q. \tag{4.2}$$

Moreover, if we define

$$G=Q^T+Q-M^T HM, \tag{4.3}$$

then we have the following proposition.

Proposition 4.1 *For the matrices Q, M and H defined in* (3.8), (3.9) *and (4.1), respectively, the matrix G defined in (4.3) is not positive definite when* $0<\tau<1$.

Proof. Because of $HM=Q$ and $M^T HM=M^T Q$, it follows from (3.8b) that

$$M^T HM=\begin{pmatrix} I_{n_2} & -\beta B^T \\ 0 & I_m \end{pmatrix}\begin{pmatrix} \tau r I_{n_2} & 0 \\ -B & \frac{1}{\beta} I_m \end{pmatrix}=\begin{pmatrix} \tau r I_{n_2}+\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

Consequently, we have

$$\begin{aligned} G &= (Q^T+Q)-M^T HM=\begin{pmatrix} 2\tau r I_{n_2} & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}-\begin{pmatrix} \tau r I_{n_2}+\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} \tau r I_{n_2}-\beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \overset{(1.13)}{=} \begin{pmatrix} D_0 & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \end{aligned} \tag{4.4}$$

Note that we can rewrite D_0 as

$$D_0=\tau D-(1-\tau)\beta B^T B, \tag{4.5}$$

where D is given by (1.12). Thus, D_0 is not positive indefinite, nor is G. $\square$

Lemma 4.1 *Let* $\{w^k\}$ *be the sequence generated by* (1.14) *for the problem* (1.1) *and* $\tilde{w}^k$ *be defined in* (3.6). *Then we have* $\tilde{w}^k \in \Omega$ *and*

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\ &\geq \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k), \ \ \forall w \in \Omega, \end{aligned} \tag{4.6}$$

where G *is defined in (4.3).*

Proof. Using $Q = HM$ (see (4.2)) and the relation (3.9a), we can rewrite the right-hand side of (3.8a), i.e., $(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)$, as $(v - \tilde{v}^k)^T H(v^k - v^{k+1})$. Hence, (3.8a) can be written as

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \ \forall w \in \Omega. \tag{4.7}$$

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 (4.7) with

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

we obtain

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \tag{4.8}$$

For the last term of (4.8), we have

$$\begin{aligned} \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 &= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\overset{(3.9a)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &\overset{(4.2)}{=} (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ &\overset{(4.3)}{=} (v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k). \end{aligned} \tag{4.9}$$

Substituting (4.9) into (4.8), we get

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k). \tag{4.10}$$

It follows from (2.3) that

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

Using this fact, the assertion of this lemma follows from (4.7) and (4.10) directly. □

In existing literature of the linearized ADMM such as [25, 28, 29, 30], the matrix D in (1.11b) is chosen by (1.12) with $\tau = 1$. Thus, the corresponding matrix G defined by (4.3) is ensured to be positive definite and the inequality (4.6) essentially implies the convergence and its worst-case convergence rate. We refer to, e.g., [17, 22], for more details. A tutorial proof can also be found in [15] (Section 4.3 and Section 5 therein). Here, because we aim at smaller values of τ and the matrix G given by (4.3) is not necessarily positive-definite, the inequality (4.6) cannot be used directly to derive the convergence and convergence rate. This difficulty makes the convergence analysis for the IP-LADMM (1.14) more challenging than that for the existing linearized ADMM (1.9).

To tackle this difficulty caused by the positive-indefiniteness of the matrix G in (4.3), our idea is to bound the term $(v^k-\tilde{v}^k)^TG(v^k-\tilde{v}^k)$ as

$$(v^k-\tilde{v}^k)^TG(v^k-\tilde{v}^k)\ge\psi(v^k,v^{k+1})-\psi(v^{k-1},v^k)+\varphi(v^k,v^{k+1}), \tag{4.11}$$

where $\psi(\cdot,\cdot)$ and $\varphi(\cdot,\cdot)$ are both non-negative functions. The first two terms $\psi(v^k,v^{k+1})$ and $\psi(v^{k-1},v^k)$ in the right-hand side of (4.11) 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 (2.2). If we find such functions that guarantee the assertion (4.11), then we can substitute it into (4.6) and obtain

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(w)\\ \ge\ &\frac{1}{2}\big(\|v-v^{k+1}\|_H^2+\psi(v^k,v^{k+1})\big)-\frac{1}{2}\big(\|v-v^k\|_H^2+\psi(v^{k-1},v^k)\big)\\ &+\frac{1}{2}\varphi(v^k,v^{k+1}),\ \ \forall w\in\Omega.\end{aligned} \tag{4.12}$$

As we shall show, all the components of the right-hand side of (4.12) in parentheses should be positive to establish the convergence and convergence rate of the IP-LADMM (1.14). The following lemmas are for this purpose; and similar techniques can be referred to [11, 19].

Lemma 4.2 *Let $\{w^k\}$ be the sequence generated by (1.14) for the problem (1.1) and $\tilde{w}^k$ be defined by (3.6). Then we have*

$$(v^k-\tilde{v}^k)^TG(v^k-\tilde{v}^k)=\tau r\|y^k-y^{k+1}\|^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2+2(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}). \tag{4.13}$$

Proof. First, it follows from (4.4) and $\tilde{y}^k=y^{k+1}$ that

$$(v^k-\tilde{v}^k)^TG(v^k-\tilde{v}^k)=\tau r\|y^k-\tilde{y}^k\|^2+\frac{1}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2-\beta\|B(y^k-\tilde{y}^k)\|^2. \tag{4.14}$$

Because $\tilde{x}^k=x^{k+1}$ and $\tilde{y}^k=y^{k+1}$, we have

$$\lambda^k-\tilde{\lambda}^k=\beta(Ax^{k+1}+By^k-b)\quad\text{and}\quad Ax^{k+1}+By^{k+1}-b=\frac{1}{\beta}(\lambda^k-\lambda^{k+1}),$$

and further

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

Substituting it into the right-hand side of (4.14), the assertion of this lemma follows directly. □

Recall that $D=rI_{n_2}-\beta B^TB$ and D is positive definite when $r>\beta\|B^TB\|$. The inequality (4.13) can be rewritten as

$$\begin{aligned}(v^k-\tilde{v}^k)^TG(v^k-\tilde{v}^k)\ &=\ \tau\|y^k-y^{k+1}\|_D^2+\tau\beta\|B(y^k-y^{k+1})\|^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2\\ &\quad+2(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}).\end{aligned} \tag{4.15}$$

Now, we treat the crossing term of the right-hand side of (4.15).

Lemma 4.3 *Let $\{w^k\}$ be the sequence generated by* (1.14) *for the problem* (1.1) *and $\tilde{w}^k$ be defined by* (3.6)*. Then we have*

$$\begin{array}{rcl} 2(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}) & \geq & \tau\|y^k-y^{k+1}\|_D^2-\tau\|y^{k-1}-y^k\|_D^2 \\ & & -3(1-\tau)\beta\|B(y^k-y^{k+1})\|^2-(1-\tau)\beta\|B(y^{k-1}-y^k)\|^2 \end{array} \quad (4.16)$$

Proof. First, according to $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$, the inequality (3.3) can be written as

$$y^{k+1}\in\mathcal{Y},\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\big(-B^T\lambda^{k+1}+D_0(y^{k+1}-y^k)\big)\geq 0,\ \ \forall y\in\mathcal{Y}. \quad (4.17)$$

Analogously, for the previous iterate, we have

$$y^k\in\mathcal{Y},\ \theta_2(y)-\theta_2(y^k)+(y-y^k)^T\big(-B^T\lambda^k+D_0(y^k-y^{k-1})\big)\geq 0,\ \ \forall y\in\mathcal{Y}. \quad (4.18)$$

Setting $y=y^k$ and $y=y^{k+1}$ in (4.17) and (4.18), respectively, and adding them, we get

$$(y^k-y^{k+1})^T\Big(B^T(\lambda^k-\lambda^{k+1})+D_0\big[(y^{k+1}-y^k)-(y^k-y^{k-1})\big]\Big)\geq 0,$$

and thus

$$2(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\geq 2(y^k-y^{k+1})^TD_0\big((y^k-y^{k+1})-(y^{k-1}-y^k)\big).$$

Consequently, by using $D_0=\tau D-(1-\tau)\beta B^TB$ (see (4.5)) and Cauchy-Schwarz inequality, it follows from the above inequality that

$$\begin{array}{cl} & 2(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1}) \\ \geq & 2(y^k-y^{k+1})^T\Big[\tau D-(1-\tau)\beta B^TB\Big]\big((y^k-y^{k+1})-(y^{k-1}-y^k)\big) \\ \geq & 2\tau\|y^k-y^{k+1}\|_D^2-2\tau(y^k-y^{k+1})D(y^{k-1}-y^k) \\ & -2(1-\tau)\beta\|B(y^k-y^{k+1})\|^2+2(1-\tau)\beta(y^k-y^{k+1})^T\big(B^TB\big)(y^k-y^{k-1}) \\ \geq & \tau\|y^k-y^{k+1}\|_D^2-\tau\|y^{k-1}-y^k\|_D^2 \\ & -3(1-\tau)\beta\|B(y^k-y^{k+1})\|^2-(1-\tau)\beta\|B(y^{k-1}-y^k)\|^2. \end{array}$$

We get (4.16) and the lemma is proved. $\Box$

Now, it follows from (4.15) and (4.16) that

$$\begin{array}{rcl} (v^k-\tilde{v}^k)^TG(v^k-\tilde{v}^k) & \geq & 2\tau\|y^k-y^{k+1}\|_D^2-\tau\|y^{k-1}-y^k\|_D^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2 \\ & & +(4\tau-3)\beta\|B(y^k-y^{k+1})\|^2-(1-\tau)\beta\|B(y^{k-1}-y^k)\|^2. \end{array} \quad (4.19)$$

In order to obtain the desired form (4.11) based on the proved inequality (4.19), we collect the negative terms in the right-hand side of (4.19) and set

$$\psi(v^{k-1},v^k):=\tau\|y^{k-1}-y^k\|_D^2+(1-\tau)\beta\|B(y^{k-1}-y^k)\|^2.$$

This also means that

$$\psi(v^k,v^{k+1}):=\tau\|y^k-y^{k+1}\|_D^2+(1-\tau)\beta\|B(y^k-y^{k+1})\|^2.$$

Thus, if we define

$$\varphi(v^k,v^{k+1}):=\tau\|y^k-y^{k+1}\|_D^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2+(5\tau-4)\beta\|B(y^k-y^{k+1})\|^2,$$

then the inequality (4.19) is in the form of (4.11). We summarize the deduction in the following lemma.

Lemma 4.4 *Let* $\{w^k\}$ *be the sequence generated by* (1.14) *for the problem* (1.1) *and* $\tilde{w}^k$ *be defined by* (3.6). *Then we have*

$$\begin{aligned}(v^k-\tilde{v}^k)^T G(v^k-\tilde{v}^k) \;\geq\;& \big(\tau\|y^k-y^{k+1}\|_D^2+(1-\tau)\beta\|B(y^k-y^{k+1})\|^2\big)\\ &-\big(\tau\|y^{k-1}-y^k\|_D^2+(1-\tau)\beta\|B(y^{k-1}-y^k)\|^2\big)\\ &+\big(\tau\|y^k-y^{k+1}\|_D^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2+(5\tau-4)\beta\|B(y^k-y^{k+1})\|^2\big). \quad (4.20)\end{aligned}$$

As mentioned, we need to ensure that $\psi(\cdot,\cdot)$ and $\varphi(\cdot,\cdot)$ are both nonnegative in (4.11). Hence, the coefficients in (4.20) should be all nonnegative. This requirement indeed implies the restriction of τ as $[0.8,1)$. It is easy to see that our analysis in Sections 6 and 7 is valid for any value of $\tau\in[0.8,1)$. But, as mentioned, we prefer smaller value of τ as long as the convergence is ensured. Thus we suggest choosing $\tau=0.8$ for the IP-LADMM (1.14). For the case where $\tau=0.8$, the term $(v^k-\tilde{v}^k)^T G(v^k-\tilde{v}^k)$ is bounded by the specific bound:

$$\begin{aligned}(v^k-\tilde{v}^k)^T G(v^k-\tilde{v}^k) \;\geq\;& \Big(\frac{4}{5}\|y^k-y^{k+1}\|_D^2+\frac{1}{5}\beta\|B(y^k-y^{k+1})\|^2\Big)\\ &-\Big(\frac{4}{5}\|y^{k-1}-y^k\|_D^2+\frac{1}{5}\beta\|B(y^{k-1}-y^k)\|^2\Big)\\ &+\Big(\frac{4}{5}\|y^k-y^{k+1}\|_D^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2\Big). \quad (4.21)\end{aligned}$$

This result is summarized in the following theorem.

Theorem 4.1 *Let* $\{w^k\}$ *be the sequence generated by* (1.14) *for the problem* (1.1) *and* $\tilde{w}^k$ *be defined by* (3.6). *Setting* $\tau=\frac{4}{5}$ *in* (4.5), *we have*

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(w)\\ &\geq\; \Big(\frac{1}{2}\|v-v^{k+1}\|_H^2+\big(\frac{2}{5}\|y^k-y^{k+1}\|_D^2+\frac{1}{10}\beta\|B(y^k-y^{k+1})\|^2\big)\Big)\\ &\quad-\Big(\frac{1}{2}\|v-v^k\|_H^2+\big(\frac{2}{5}\|y^{k-1}-y^k\|_D^2+\frac{1}{10}\beta\|B(y^{k-1}-y^k)\|^2\big)\Big)\\ &\quad+\Big(\frac{2}{5}\|y^k-y^{k+1}\|_D^2+\frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2\Big). \quad (4.22)\end{aligned}$$

5 More comments on τ

In Section 4, we show that $\tau\in[0.8,1)$ is sufficient to ensure the convergence of the IP-LADMM (1.14) and we suggest choosing $\tau=0.8$ because of the inequality (4.20). It worths to mention that our analysis is based on performing some inequalities and the lower bound 0.8 is a sufficient condition to ensure the convergence of (1.14); it is thus of conservative nature and practically it could be further relaxed when implementing (1.14). Meanwhile, because of the philosophy of preferring smaller values of τ as long as the convergence of (1.14) can be theoretically ensured, it is interesting to ask whether or not 0.8 is the smallest value to ensure the convergence of the IP-LADMM (1.14). This question seems to be too challenging to be answered rigorously; but we can show by an extremely simple example that the smallest value of τ cannot be smaller than 0.75. Therefore, the bound 0.8 of τ established in Section 4 is nearly optimal, even it may not be the truly optimal value.

Let us consider the simplest equation $y=0$ in $\Re$; and show that the IP-LADMM (1.14) is not necessarily convergent when $\tau<0.75$. Obviously, $y=0$ is a special case of the model (1.1) as:

$$\min\{0\cdot x+0\cdot y \mid 0\cdot x+y=0,\ x\in\{0\}, y\in\Re\}. \quad (5.1)$$

Without loss of generality, we take $\beta = 1$ and thus the augmented Lagrangian function of the problem (5.1) is

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

The iterative scheme of the IP-LADMM (1.14) for (5.1) is

$$\text{(IP-LADMM)} \quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}^2(x, y^k, \lambda^k) \mid x \in \{0\}\}, & \text{(5.2a)} \\ y^{k+1} = \arg\min\{-y^T\lambda^k + \frac{1}{2}\|y\|^2 + \frac{1}{2}\|y - y^k\|^2_{D_0} \mid y \in \Re\}, & \text{(5.2b)} \\ \lambda^{k+1} = \lambda^k - (y^{k+1}). & \text{(5.2c)} \end{cases}$$

Since $\beta = 1$ and $B^T B = 1$, it follows from (4.5) and (1.12) that

$$D_0 = \tau D - (1 - \tau) \qquad \text{and} \qquad D = r - 1 > 0.$$

We thus have

$$D_0 = \tau r - 1, \quad \forall r > 1$$

and the recursion (5.2) becomes

$$\begin{cases} x^{k+1} = 0, \\ -\lambda^k + y^{k+1} + (\tau r - 1)(y^{k+1} - y^k) = 0, \\ \lambda^{k+1} = \lambda^k - y^{k+1}. \end{cases} \tag{5.3}$$

For any $k > 0$, we have $x^k = 0$. We thus just need to study the iterative sequence $\{v^k = (y^k, \lambda^k)\}$. For any given $\tau < 0.75$, there exists $r > 1$ such that $\tau r < 0.75$ holds. Setting $\alpha = \tau r$, the iterative scheme for $v = (y, \lambda)$ can be written as

$$\begin{cases} \alpha y^{k+1} = \lambda^k + (\alpha - 1)y^k \\ \lambda^{k+1} = \lambda^k - y^{k+1}. \end{cases} \tag{5.4}$$

With elementary manipulations, we get

$$\begin{cases} y^{k+1} = \dfrac{\alpha - 1}{\alpha} y^k + \dfrac{1}{\alpha}\lambda^k, \\ \lambda^{k+1} = \dfrac{1 - \alpha}{\alpha} y^k + \dfrac{\alpha - 1}{\alpha}\lambda^k, \end{cases} \tag{5.5}$$

which can be written as

$$v^{k+1} = P(\alpha)v^k \qquad \text{with} \qquad P(\alpha) = \frac{1}{\alpha}\begin{pmatrix} \alpha - 1 & 1 \\ 1 - \alpha & \alpha - 1 \end{pmatrix}. \tag{5.6}$$

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

$$f_1(\alpha) = \frac{(\alpha - 1) + \sqrt{1 - \alpha}}{\alpha}, \quad \text{and} \quad f_2(\alpha) = \frac{(\alpha - 1) - \sqrt{1 - \alpha}}{\alpha}.$$

For the function $f_2(\alpha)$, we have $f_2\big(0.75\big) = -1$ and

$$\begin{aligned} f_2'(\alpha) &= \frac{1}{\alpha^2}\Big(\big(1 - \frac{-1}{2\sqrt{1 - \alpha}}\big)\alpha - \big((\alpha - 1) - \sqrt{1 - \alpha}\big)\Big) \\ &= \frac{1}{\alpha^2}\Big(\big(\alpha + \frac{\alpha}{2\sqrt{1 - \alpha}}\big) + (1 - \alpha) + \sqrt{1 - \alpha}\Big) > 0, \quad \forall \alpha \in \big(0, 0.75\big). \end{aligned}$$

Therefore, we have

$$f_2(\alpha) = \frac{(\alpha-1)-\sqrt{1-\alpha}}{\alpha} < -1, \quad \forall \alpha \in \big(0, 0.75\big).$$

That is, for any $\alpha \in \big(0, 0.75\big)$, the matrix $P(\alpha)$ in (5.6) has an eigenvalue less than -1 and the iterative scheme (5.5) is divergent. Hence, the IP-LADMM (1.14) is not necessarily convergent for any $\tau \in (0, 0.75)$.

6 Convergence

In this section, we explicitly prove the convergence of the IP-LADMM (1.14). With the assertion in Theorem 4.1, the proof is subroutine.

Lemma 6.1 *Let $\{w^k\}$ be the sequence generated by* (1.14) *for the problem* (1.1)*. Then we have*

$$\begin{aligned} &\|v^{k+1}-v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^k-y^{k+1})\|^2 + 4\|y^k-y^{k+1}\|_D^2\big) \\ &\le \ \big(\|v^k-v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^{k-1}-y^k)\|^2 + 4\|y^{k-1}-y^k\|_D^2\big)\big) \\ &\quad -\big(\frac{4}{5}\|y^k-y^{k+1}\|_D^2 + \frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2\big). \end{aligned} \tag{6.1}$$

Proof. Setting $w = w^*$ in (4.22) and performing simple manipulations, we get

$$\begin{aligned} &\frac{1}{2}\|v^k-v^*\|_H^2 + \frac{1}{10}\big(\beta\|B(y^{k-1}-y^k)\|^2 + 4\|y^{k-1}-y^k\|_D^2\big) \\ &\ge \ \Big(\frac{1}{2}\|v^{k+1}-v^*\|_H^2 + \frac{1}{10}\big(\beta\|B(y^k-y^{k+1})\|^2 + 4\|y^k-y^{k+1}\|_D^2\big)\Big) \\ &\quad +\Big(\frac{2}{5}\|y^k-y^{k+1}\|_D^2 + \frac{1}{2\beta}\|\lambda^k-\lambda^{k+1}\|^2\Big) + \big(\theta(\tilde{u}^k)-\theta(u^*) + (\tilde{w}^k-w^*)^T F(w^*)\big). \end{aligned} \tag{6.2}$$

For a solution point of (2.2), we have

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

Thus, the assertion (6.1) follows from (6.2) directly. □

Theorem 6.1 *Let $\{w^k\}$ be the sequence generated by* (1.14) *for the problem* (1.1)*. Then the sequence $\{v^k\}$ converges to $v^\infty \in \mathcal{V}^*$.*

Proof. First, it follows from (6.1) that

$$\begin{aligned} &\frac{4}{5}\|y^k-y^{k+1}\|_D^2 + \frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2 \\ &\le \ \Big(\|v^k-v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^{k-1}-y^k)\|^2 + 4\|y^{k-1}-y^k\|_D^2\big)\Big) \\ &\quad -\Big(\|v^{k+1}-v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^k-y^{k+1})\|^2 + 4\|y^k-y^{k+1}\|_D^2\big)\Big). \end{aligned} \tag{6.3}$$

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

$$\sum_{k=1}^{\infty}\big(\frac{4}{5}\|y^k-y^{k+1}\|_D^2 + \frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2\big) \le \|v^1-v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^0-y^1)\|^2 + 4\|y^0-y^1\|_D^2\big).$$

Because D is positive definite, it follows from the above inequality that

$$\lim_{k\to\infty} \|v^k - v^{k+1}\| = 0. \tag{6.4}$$

For an arbitrarily fixed $v^* \in \mathcal{V}^*$, it follows from (6.1) that

$$\begin{aligned} \|v^{k+1} - v^*\|_H^2 &\le \|v^k - v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^{k-1} - y^k)\|^2 + 4\|y^{k-1} - y^k\|_D^2\big) \\ &\le \|v^1 - v^*\|_H^2 + \frac{1}{5}\big(\beta\|B(y^0 - y^1)\|^2 + 4\|y^0 - y^1\|_D^2\big), \quad \forall k \ge 1. \end{aligned} \tag{6.5}$$

Thus, the sequence $\{v^k\}$ is bounded. Because M is non-singular, according to (3.9), $\{\tilde{v}^k\}$ is also bounded. Let v^∞ be a cluster point of $\{\tilde{v}^k\}$ and $\{\tilde{v}^{k_j}\}$ be the subsequence converging to v^∞. Let x^∞ be the vector accompanied with $(y^\infty, \lambda^\infty) \in \mathcal{V}$. Then, it follows from (4.7) and (6.4) that

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

which means that w^∞ is a solution point of (2.2) and its essential part $v^\infty \in \mathcal{V}^*$. Since $v^\infty \in \mathcal{V}^*$, it follows from (6.5) that

$$\|v^{k+1} - v^\infty\|_H^2 \le \|v^k - v^\infty\|_H^2 + \frac{1}{5}\big(\beta\|B(y^{k-1} - y^k)\|^2 + 4\|y^{k-1} - y^k\|_D^2\big). \tag{6.6}$$

Note that v^∞ is also the limit point of $\{v^{k_j}\}$. Together with (6.4), this fact means that it is impossible for the sequence $\{v^k\}$ to have more than one cluster point. Therefore, the sequence $\{v^k\}$ converges to v^∞ and the proof is complete. □

7 Convergence rate

In this section, we establish the worst-case $O(1/t)$ convergence rate measured by the iteration complexity for the IP-LADMM (1.14), where t is the iteration counter. Recall that the worst-case $O(1/t)$ convergence rate for the original ADMM (1.3) and its linearized version (1.9) generated by the proximal ADMM (1.11) with the positive definite proximal regularization (1.12) (actually, the matrix D in (1.11) could be relaxed as positive semidefinite) has been established in [21].

We first elaborate on how to define an approximate solution of the variational inequality (2.2). According to (2.2), if $\tilde{w}$ satisfies

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

then $\tilde{w}$ is a solution point of (2.2). By using $(w-\tilde{w})^T F(\tilde{w}) = (w-\tilde{w})^T F(w)$ (see (2.3)), the solution point $\tilde{w}$ can be also characterized by

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

Hence, we can use this characterization to define an approximate solution of the variational inequality (2.2). More specifically, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of the variational inequality (2.2) if it satisfies

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

where

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

We refer to [7, 21] for more details of this definition. Below, we show that after t iterations of the IP-LADMM (1.14), we can find $\tilde{w}$ such that

$$\tilde{w} \in \Omega \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\} \le \epsilon = O(\frac{1}{t}). \tag{7.1}$$

Theorem 4.1 is again the starting point of the analysis.

Theorem 7.1 *Let* $\{w^k\}$ *be the sequence generated by* (1.14) *for the problem* (1.1) *and* $\tilde{w}^k$ *be defined by* (3.6)*. Then for any integer* t*, we have*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2t}\Big(\|v - v^1\|_H^2 + \frac{1}{5}\big(\beta\|B(y^0 - y^1)\|^2 + 4\|y^0 - y^1\|_D^2\big)\Big), \tag{7.2}$$

where

$$\tilde{w}_t = \frac{1}{t}(\sum_{k=1}^{t} \tilde{w}^k). \tag{7.3}$$

Proof. First, it follows from (4.22) that

$$\begin{aligned}
&\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\
&\quad \ge \frac{1}{2}\Big(\|v - v^{k+1}\|_H^2 + \frac{1}{5}\big(\beta\|B(y^k - y^{k+1})\|^2 + 4\|y^k - y^{k+1}\|_D^2\big)\Big) \\
&\qquad - \frac{1}{2}\Big(\|v - v^k\|_H^2 + \frac{1}{5}\big(\beta\|B(y^{k-1} - y^k)\|^2 + 4\|y^{k-1} - y^k\|_D^2\big)\Big).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) + \frac{1}{2}\Big(\|v - v^{k+1}\|_H^2 + \frac{1}{5}\big(\beta\|B(y^k - y^{k+1})\|^2 + 4\|y^k - y^{k+1}\|_D^2\big)\Big) \\
&\quad \le \frac{1}{2}\Big(\|v - v^k\|_H^2 + \frac{1}{5}\big(\beta\|B(y^{k-1} - y^k)\|^2 + 4\|y^{k-1} - y^k\|_D^2\big)\Big).
\end{aligned} \tag{7.4}$$

Summarizing the inequality(7.4) over $k = 1, 2, \ldots, t$, we obtain

$$\sum_{k=1}^{t} \theta(\tilde{u}^k) - t\theta(u) + (\sum_{k=1}^{t} \tilde{w}^k - tw)^T F(w) \le \frac{1}{2}\|v - v^1\|_H^2 + \frac{1}{10}\big(\beta\|B(y^0 - y^1)\|^2 + 4\|y^0 - y^1\|_D^2\big).$$

and thus

$$\frac{1}{t}(\sum_{k=1}^{t} \theta(\tilde{u}^k)) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2t}\Big(\|v - v^1\|_H^2 + \frac{1}{5}\big(\beta\|B(y^0 - y^1)\|^2 + 4\|y^0 - y^1\|_D^2\big)\Big). \tag{7.5}$$

Since $\theta(u)$ is convex and

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

we have that

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

Substituting it into (7.5), the assertion of this theorem follows directly. □

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

$$d := \sup\{\|v - v^1\|_H^2 + \frac{1}{5}\big(\beta\|B(y^0 - y^1)\|^2 + 4\|y^0 - y^1\|_D^2\big) \mid w \in \mathcal{D}_{(\tilde{w})}\}$$

with $v^0 = (y^0, \lambda^0)$ and $v^1 = (y^1, \lambda^1)$ the initial point and the first computed iterate, respectively. Then, after t iterations of the IP-LADMM (1.14), the point $\tilde{w}_t$ defined in (7.3) satisfies

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

which means $\tilde{w}_t$ is an approximate solution of the variational inequality (2.2) with an accuracy of $O(1/t)$ (recall (7.1)). Since $\tilde{w}_t$ is defined in (7.3), the worst-case $O(1/t)$ convergence rate in Theorem 7.1 is in the ergodic sense.

8 Linearized augmented Lagrangian method via positive-indefinite proximal regularization

The original ADMM (1.3) can be regarded as a splitting version of the ALM proposed in [23, 27]. Our analysis for the IP-LADMM (1.14) can be easily simplified to the ALM and thus a linearized ALM via choosing a positive-indefinite proximal regularization can be proposed. In this section, we provide some necessary details and skip those similar as the analysis in Sections 3-7. We purposively reuse some notation for a clearer comparison with the analysis in Sections 3-7. Explanations and definitions of some notation having the same meaning as those in previous sections are also skipped for succinctness.

8.1 Brief review of the ALM

To start the discussion of ALM, we consider the canonical convex minimization model with linear constraints:

$$\min\{\theta_2(y) \mid By = b,\ y \in \mathcal{Y}\}, \tag{8.1}$$

which can be regarded as a special case of the model (1.1) with $\theta_1(x) \equiv 0$ and $A \equiv 0$. Let the Lagrangian and augmented Lagrangian functions of the problem (8.1) be

$$L(y, \lambda) = \theta(y) - \lambda^T (By - b), \tag{8.2}$$

and

$$\mathcal{L}_\beta(y, \lambda) = \theta_2(y) - \lambda^T (By - b) + \frac{\beta}{2}\|By - b\|^2, \tag{8.3}$$

respectively. A saddle point of $L(y, \lambda)$, denoted by $v^* = (y^*, \lambda^*)$, can be characterized by the variational inequality

$$v^* \in \mathcal{V}^*, \quad \theta(y) - \theta(y^*) + (v - v^*)^T F(v^*) \ge 0, \quad \forall\, v \in \mathcal{V}, \tag{8.4}$$

where

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \quad F(v) = \begin{pmatrix} -B^T\lambda \\ By - b \end{pmatrix} \quad \text{and} \quad \mathcal{V} = \mathcal{Y} \times \Re^m. \tag{8.5}$$

The iteration of the ALM originally proposed in [23, 27] for (8.1) reads as

$$\text{(ALM)} \quad \begin{cases} y^{k+1} = \arg\min\left\{\mathcal{L}_\beta(y, \lambda^k) \bigm| y \in \mathcal{Y}\right\}, & \text{(8.6a)} \\ \lambda^{k+1} = \lambda^k - \beta(By^{k+1} - b). & \text{(8.6b)} \end{cases}$$

8.2 ALM with positive-definite proximal regularization

Similar as the proximal ADMM (1.11), the subproblem (8.6a) can be proximally regularized and the resulting scheme is

$$\text{(PD-ALM)} \quad \begin{cases} y^{k+1} = \arg\min\{\mathcal{L}_\beta(y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \mid y \in \mathcal{Y}\}, & (8.7a) \\ \lambda^{k+1} = \lambda^k - \beta(By^{k+1} - b), & (8.7b) \end{cases}$$

where $D \in \Re^{n_2 \times n_2}$ is required to be positive definite. The convergence of (8.7) can be easily derived. Indeed, similar as Section 4, skipping the details, we can rewrite the scheme (8.7) as the variational inequality:

$$\theta(y) - \theta(y^{k+1}) + \begin{pmatrix} y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -B^T \lambda^{k+1} \\ By^{k+1} - b \end{pmatrix} + \begin{pmatrix} D & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \right\} \ge 0, \ \forall \, v \in \mathcal{V}. \quad (8.8)$$

Setting $v = v^*$ in (8.8) and using the notations in (8.5), we get

$$\begin{pmatrix} y^{k+1} - y^* \\ \lambda^{k+1} - \lambda^* \end{pmatrix}^T \begin{pmatrix} D & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} \ge \theta(y^{k+1}) - \theta(y^*) + (v^{k+1} - v^*)^T F(v^{k+1}).$$

Because $(v^{k+1} - v^*)^T F(v^{k+1}) = (v^{k+1} - v^*)^T F(v^*)$ (see (8.5)), and v^* is a solution point, the right-hand side of the last inequality is non-negative. It is easy to see that we have

$$(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \ge 0, \quad (8.9)$$

where

$$H = \begin{pmatrix} D & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}.$$

The matrix H is positive definite because D is assumed to be so. Then, it follows from (8.9) that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \ \forall \, v^* \in \mathcal{V}^*, \quad (8.10)$$

which essentially implies the convergence of the proximal ALM (8.7) via positive-definite proximal regularization.

8.3 Linearized ALM via positive-definite proximal regularization

If the matrix D in (8.7) is chosen by

$$D = rI_{n_2} - \beta B^T B \quad \text{with} \quad r > \beta \|B^T B\|, \quad (8.11)$$

then the ALM with positive-definite proximal regularization (8.7) is specified as the linearized version of ALM whose y-subproblem reduces to

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

Again, this problem reduces to estimating the proximity operator of $\theta_2(y)$ when $\mathcal{Y} = \Re^{n_2}$. We refer to, e.g., [29] for some efficient applications of the linearized version of ALM to sparse- and low-rank-driven models.

8.4 Linearized ALM via positive-indefinite proximal regularization

Similar as Section 4, we relax the positive definiteness requirement of D in (8.7) and propose the linearized ALM via positive-indefinite proximal regularization:

$$\text{(IP-LALM)} \quad \begin{cases} y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_0}^2 \mid y \in \mathcal{Y}\big\}, & (8.12a) \\ \lambda^{k+1} = \lambda^k - \beta(By^{k+1} - b). & (8.12b) \end{cases}$$

where

$$D_0 = 0.8 \cdot r \cdot I_{n_2} - \beta B^T B \quad \text{with} \quad r > \beta\|B^T B\|. \tag{8.13}$$

We call the scheme (8.12) succinctly as "indefinite proximal linearized ALM", further abbreviated as "IP-LALM". It is trivial to see that with some constants ignored, the subproblem (8.12a) can be written as

$$y^{k+1} = \arg\min\big\{\theta_2(y) + \frac{\tau r}{2}\big\|y - \big[y^k + \frac{1}{\tau r}B^T(\lambda^k - \beta(By^k - b))\big]\big\|^2 \mid y \in \mathcal{Y}\big\}, \tag{8.14}$$

which also reduces to estimating the proximity operator of $\theta_2(y)$ when $\mathcal{Y} = \Re^{n_2}$.

8.5 Convergence analysis of the IP-LALM

Similar as Section 3, we define the artificial vector $\tilde{v}^k = (\tilde{y}^k, \tilde{\lambda}^k)$ by

$$\tilde{y}^k = y^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - \beta(By^k - b), \tag{8.15}$$

where y^{k+1} is the output of (8.12a) from the given (y^k, λ^k). Then, following the steps in Section 3, we can easily see that the IP-LALM (8.12) can be reformulated as the prediction-correction framework

$$\begin{cases} \tilde{v}^k \in \mathcal{V}, \;\; \theta(y) - \theta(\tilde{y}^k) + (v - \tilde{v}^k)^T F(\tilde{v}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall\, v \in \mathcal{V}, \\ v^{k+1} = v^k - M(v^k - \tilde{v}^k), \end{cases}$$

where Q and M are the same as (3.8b) and (3.9b), respectively. Then, the remaining part of establishing the global convergence and worst-case $O(1/t)$ convergence rate for the IP-LALM (8.12) immediately follows from the analysis in Sections 6 and 7. We omit the details for succinctness.

9 Conclusions

In this paper, we propose a new linearized version of the alternating direction method of multipliers (ADMM) via choosing a positive-indefinite proximal regularization for solving a convex minimization model with linear constraints and an objective function represented as the sum of two convex functions without coupled variables. Without any additional condition on the model itself, we relax the positive-definiteness requirement that is popularly required in existing literature of proximal versions of ADMM. Compared with well-studied linearized versions of ADMM in existing literature, the new linearized ADMM still can alleviate an ADMM subproblem as easy as estimating the proximity operator of a function for some applications while it allows larger step sizes. This is an important feature to ensure faster convergence for various applications. The global convergence and worst-case $O(1/t)$ convergence rate measured by the iteration complexity are established for the new linearized ADMM. As a byproduct, we also propose a linearized version of the augmented Lagrangian method via choosing positive-indefinite proximal regularization; this algorithm itself is useful for many applications.

Finally, we would mention that, to expose the main idea more clearly, our discussion only focuses on the original prototype ADMM scheme (1.3) with a constant parameter β; and only one subproblem is proximally regularized. Our discussion based on the proximal-indefinite proximal regularization can be further extended to many variants of ADMM such as the strictly contractive version of the symmetric ADMM (also known as the Peaceman-Rachford splitting method) in [17], the case with dynamically-adjusted penalty parameters in [20], the case where both the subproblems are proximally regularized in [16], the case where the proximal matrix can be dynamically adjusted in [16], and even some more complicated cases where the mentioned variants are merged such as [24, 25]. But the discussion in our current setting still represents the simplest yet most fundamental case that is the basis of discussing the possibility of relaxing the positive definiteness requirement for other more complicated cases.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato J. and Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learning, 3 (2010), pp. 1-122.

[2] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations*, technical report, Stanford University, 1978.

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

[4] J. Eckstein, *Approximate ADMM algorithms derived from Lagrangian splitting*, manuscript, 2016.

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

[6] 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.

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

[8] M. Fortin and R. Glowinski, Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems, Stud. Math. Appl. 15, NorthCHolland, Amsterdam, 1983.

[9] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983, pp. 299-331.

[10] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

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

[12] R. Glowinski, *On alternating direction methods of multipliers: A historical perspective*, in Modeling, Simulation and Optimization for Science and Technology, W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanm ki, and O. Pironneau, eds., Springer Netherlands, 2014, pp. 59–82.

[13] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[14] W. Hager, C. Ngo, M. Yashtini and H. Zhang, *An Alternating direction approximate Newton algorithm for ill-conditioned inverse problems with application to parallel MRI*, J. Oper. Res. Soc. China, 3 (2015), pp. 139-162.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] B. S. He, H. Yang and S. L. Wang, *Alternating direction method with selfadaptive penalty parameters for monotone variational inequalities*, J. Optim. Theory Appli., 106(2) (2000), pp. 337-356.

[21] 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.

[22] B. S. He and X. M. Yuan, *On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers*, Numer. Math., 130 (3) (2015), pp. 567-577.

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

[24] X. X. Li and X. M. Yuan, *A proximal strictly contractive Peaceman-Rachford splitting method for convex programming with applications to imaging*, SIAM J. Imag. Sci., 8(2) (2015), pp. 1332-1365.

[25] Z. Lin, R. Liu and H. Li, *Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning*, Machine Learning, 99 (2) (2015), pp. 287-325.

[26] M. K. Ng, F. Wang and X. M. Yuan, *Inexact alternating direction methods for image recovery*, SIAM J. Sci. Comput., 33(4) (2011), pp. 1643-1668.

[27] 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.

[28] X. F. Wang and X. M. Yuan, *The linearized alternating direction method of multipliers for Dantzig selector*, SIAM J. Sci. Comput., 34 (2012), pp. A2792–A2811.

[29] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301-329.

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