Optimal Linearized Alternating Direction Method of Multipliers for Convex Programming[1]

Bingsheng He[2] **Feng Ma**[3] **Xiaoming Yuan** [4]

October 14, 2017

Abstract. The alternating direction method of multipliers (ADMM) is being widely used in a variety of areas; its different variants tailored for different application scenarios have also been deeply researched in the literature. Among them, the linearized ADMM has received particularly wide attention from many areas because of its efficiency and easy implementation. To theoretically guarantee the convergence of the linearized ADMM, the step size for the linearized subproblems, or the reciprocal of the linearization parameter, should be sufficiently small. On the other hand, small step sizes decelerate the convergence numerically. Hence, it is crucial to determine an optimal (largest) value of the step size while the convergence of the linearized ADMM can be still ensured. Such an analysis seems to be lacked in the literature. In this paper, we show how to find this optimal step size for the linearized ADMM and hence propose the optimal linearized ADMM in the convex programming context. Its global convergence and worst-case convergence rate measured by the iteration complexity are proved as well.

Keywords: Convex programming, alternating direction method of multipliers, linearized, optimal step size, proximal regularization, 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

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

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

[1]This is an improved version of our earlier manuscript released on Optimization Online in July 2016 (#5569) entitled "Linearized Alternating Direction Method of Multipliers via Positive-Indefinite Proximal Regularization for Convex Programming", with the proved optimality of 0.75 for the parameter τ.

[2]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

[3]High-Tech Institute of Xi'an, Xi'an, 710025, Shaanxi, China. This author was supported by the NSFC Grant 11701564. Email: mafengnju@gmail.com

[4]Corresponding author. Department 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

[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

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

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 application domains such as image processing, statistical learning, computer vision, wireless network, and so on. We refer to [1, 5, 12] for some review papers of the ADMM. For simplicity, 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 the subproblems of ADMM (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 specific application-tailored 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 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 the function $\theta_2(y)$, the matrix B and the set $\mathcal{Y}$. We refer to, e.g., [4, 16, 26], for some detailed discussions on the generic case where the function and matrix 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 for considering 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 scalar. That is, the linearized subproblem (1.6) amounts to estimating the proximity operator of $\theta_2(y)$, which is defined as

$$
\operatorname{prox}_{\gamma\theta_2}(y) := \operatorname*{argmin}_z \Big\{\theta_2(z) + \frac{1}{2\gamma}\|z - y\|^2\Big\}, \tag{1.8}
$$

with $\gamma > 0$. A representative case is $\theta_2(y) = \|y\|_1$ which arises often in sparsity-driven problems such as compressive sensing, total variational image restoration and variable selection for high-dimensional data. For this case, the proximity operator of $\|y\|_1$ has a closed-form which is given by the so-called shrinkage operator defined as

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

Hence, replacing the original subproblem (1.3b) with its linearized surrogate, we obtain the linearized version of ADMM:

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

where q_k is given in (1.7). The scheme (1.9) has been widely used in areas; we refer to , e.g. [25, 27, 28, 29], for a few.

The linearized ADMM (1.9) can be explained via the proximal regularization perspective. Ignoring some constants in the objective function, we can rewrite the linearized subproblem (1.9b) as

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

with $D \in \Re^{n_2 \times n_2} = rI_{n_2} - \beta B^T B$; and the term $\frac{1}{2}\|y - y^k\|_D^2$ serves as a proximal regularization term. We call r the linearization parameter of the linearized ADMM (1.9), because as explained, the specific form of D linearizes the quadratic term $\|By + (Ax^{k+1} - b - \frac{1}{\beta}\lambda^k)\|^2$ when B is not an identity matrix. Accordingly, the reciprocal of r is referred to as the step size for solving the y-subproblem (1.9b). In the literature (see, e.g., [25, 27, 28, 29]), $r > \beta\|B^T B\|$ is required to ensure the positive definiteness of the matrix D and eventually the convergence of the linearized ADMM (1.9). Therefore, the linearized ADMM (1.9) is just a special case of the following proximal version of ADMM (1.11)

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

with the particular choice of

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

As well shown in the literature, the positive definiteness of the proximal matrix D plays an essential role in 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 linearization parameter r in (1.12) is forced to be large and thus tiny step sizes inevitably occur for solving the subproblem (1.9b). The overall convergence speed of the linearized ADMM (1.9) thus may be substantially decelerated. The proximal term in (1.11b) with a large value of r can be regarded as an over-regularization 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 proximal version of ADMM is to choose a value larger than but very close to the lower bound $\beta\|B^T B\|$, as empirically used in, e.g., [6, 19]. Therefore, 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 optimally relax the positive-definiteness requirement of the proximal matrix D in (1.12) and thus yield the largest step size for solving the linearized subproblems, while the convergence of the linearized ADMM (1.9) can be theoretically still ensured. The main purpose of this paper is answering this question. More specifically, instead of (1.12), we shall show that the convergence of the linearized ADMM (1.9) can be ensured by

$$D_0 = \tau r I - \beta B^T B \quad \text{with} \quad r > \beta\|B^T B\| \ \text{ and } \ \tau \in (0.75, 1). \tag{1.13}$$

That is, the linearization parameter r of the linearized ADMM (1.9) can be reduced by at most 25%. With this choice of τ, the matrix D_0 in (1.13) is positive indefinite; hence it is not necessary to require the positive definiteness or semi-definiteness for the matrix D as existing work in the literature. We shall also show that this is the optimal choice for τ because any value smaller than 0.75 can yield divergence of the linearized ADMM; hence it is not possible to guarantee the convergence of the linearized ADMM (1.9) with τ smaller than 0.75.

Overall, we propose the optimal linearized ADMM as following:

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

where D_0 is given by (1.13). In this case, ignoring some constants, the y-subproblem (1.14b) can be written as

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

with q_k given in (1.7). Note that the y-subproblem (1.14b) is still alleviated as estimating the proximity operator of $\theta_2(y)$ for some applications; hence the main feature of the linearized ADMM is kept by the OLADMM (1.14). Also, the subproblem (1.14b) is still convex even though the proximal matrix D_0 is positive indefinite. We slightly abuse the notation $\|y\|_{D_0}^2 := y^T D_0 y$ when D_0 is not positive definite.

Note that for many applications, it suffices to linearize one subproblem for the ADMM (1.3). Without loss of generality, we just discuss the case where only the y-subproblem is linearized/proximally regularized in (1.14). Technically, it is still possible to consider the case where both the subproblems are linearized or proximally regularized, see, e.g., [16]. Also, there are some works in the literature discussing how to relax the restriction on D as positive semi-definiteness (i.e., $r = \beta\|B^T B\|$) while posing more assumptions on the model (1.1) per se. Our analysis here indeed allows positive indefiniteness of D (i.e., $r < \beta\|B^T B\|$) without any additional assumptions on the model (1.1).

The rest of this paper is organized as follows. We first summarize some preliminary results in Section 2. Then we reformulate the OLADMM (1.14) in a prediction-correction framework in Section 3 and discuss how to determine the value of τ in Section 4. In Section 5, the convergence of the OLADMM (1.14) is proved with $\tau \in (0.75, 1)$. Then, in Section 6, we show by an example that any value of τ in $(0, 0.75)$ does not guarantee the convergence of the OLADMM (1.14) and thus illustrate that 0.75 is optimal for the linearization parameter. The worst-case convergence rate measured by the iteration complexity is established for the OLADMM (1.14) in Section 7. Some conclusions are draw in Section 8.

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

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

A pair $((x^*, y^*), \lambda^*)$ is called a saddle point of the Lagrangian function if it satisfies

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

We can rewrite them as the variational inequalities:

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

or in the more compact form:

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

where

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

and

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

We denote by Ω^* the solution set of (2.2). Note that the operator F 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

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

3 A prediction-correction reformulation of the OLADMM (1.14)

In this section, we revisit the OLADMM (1.14) from the variational inequality perspective and show that the scheme (1.14) 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) and (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, 20, 21]. Thus, the variables x and (y, λ) are called intermediate and essential variables, respectively. To distinguish the essential variables, 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

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

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 $D_0 = \tau r I - \beta B^T B$ (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 scheme (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

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

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

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

and

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

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

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

Then, using the notation (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 OLADMM (1.14) are given by

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

where

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I_m \end{pmatrix}. \tag{3.9b}$$

Overall, the new iterate of the OLADMM (1.14) can be 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 OLADMM (1.14). Indeed, according to (2.2), 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

In this section, we focus on the predictor $\tilde{w}^k$ in (3.8) and conduct a more elaborated analysis; some inequalities regarding $\tilde{w}^k$ will be derived. These inequalities are also essential for the convergence analysis of the scheme (1.14). Thus, the results in this section are also the preparation of the main convergence result in the next sections. In our analysis, how to choose τ becomes clear.

First of all, recall that we choose the matrix D_0 by (1.13) for the OLADMM (1.14). We can further rewrite D_0 as

$$D_0 = \tau D - (1-\tau)\beta B^T B, \tag{4.1}$$

where D is given by (1.12). In addition, for any given positive constants τ, r and β, we define a matrix

$$H = \begin{pmatrix} \tau r I & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.2}$$

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

$$HM = Q. \tag{4.3}$$

Moreover, if we define

$$G = Q^T + Q - M^T H M, \tag{4.4}$$

then we have the following proposition.

Proposition 4.1 *For the matrices Q, M and H defined in* (3.8)*,* (3.9) *and* (4.2)*, respectively, the matrix G defined in* (4.4) *is not positive definite when* $0 < \tau < 1$.

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

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

Consequently, we have

$$\begin{aligned} G &= (Q^T + Q) - M^T H M = \begin{pmatrix} 2\tau r I & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} \tau r I + \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} \tau r I - \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \overset{(1.13)}{=} \begin{pmatrix} D_0 & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \end{aligned} \tag{4.5}$$

Since D_0 is not positive definite (see (4.1)), nor is G. $\square$

Lemma 4.1 *Let $\{w^k\}$ be the sequence generated by the OLADMM* (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 (4.4).

Proof. Using $Q = HM$ (see (4.3)) 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 $a = v$, $b = \tilde{v}^k$, $c = v^k$ and $d = v^{k+1}$, 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

$$\begin{aligned} &\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\quad \overset{(3.9a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &\quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &\quad \overset{(4.3)}{=} \quad (v^k - \tilde{v}^k)^T(Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ &\quad \overset{(4.4)}{=} \quad (v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k). \end{aligned} \tag{4.9}$$

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, 27, 28, 29], $\tau = 1$ and $r > \beta\|B^T B\|$. Thus, the corresponding matrix G defined by (4.4) 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, 23], for more details. A tutorial proof can also be found in [15] (Sections 4.3 and 5 therein). Here, because we aim at smaller values of τ and the matrix G given by (4.4) 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 scheme (1.14) more challenging than that for the proximal version of ADMM (1.11) with a positive definite matrix D.

In the following we try to bound the term $(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k)$ as

$$(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \geq \psi(v^k, v^{k+1}) - \psi(v^{k-1}, v^k) + \varphi(v^k, v^{k+1}), \tag{4.11}$$

where $\psi(\cdot,\cdot)$ and $\varphi(\cdot,\cdot)$ are both non-negative functions. The first two terms $\psi(v^k,v^{k+1})-\psi(v^{k-1},v^k)$ in the right-hand side of (4.11) can be easily manipulated by consecutive 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

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

As we shall show, the inequality (4.12) with all positive components in its right-hand side is important for establishing the convergence and convergence rate of the OLADMM (1.14). The following lemmas are for this purpose; and similar techniques can be referred to [11, 20].

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

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

Proof. First, it follows from (4.5) and $\tilde{y}^k=y^{k+1}$ that

$$\begin{aligned} (v^k-\tilde{v}^k)^T G(v^k-\tilde{v}^k) &= \tau r\|y^k-\tilde{y}^k\|^2+\frac{1}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2-\beta\|B(y^k-\tilde{y}^k)\|^2 \\ &= \tau r\|y^k-y^{k+1}\|^2-\beta\|B(y^k-y^{k+1})\|^2+\frac{1}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2. \end{aligned} \tag{4.14}$$

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

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

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

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

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

Now, we treat the crossing term $2(\lambda^k-\lambda^{k+1})^T B(y^k-y^{k+1})$ in the right-hand side of (4.15) and estimate a lower-bound in the quadratic terms.

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

$$\begin{array}{ll} & (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \\ \geq & \big(\frac{\tau}{2}\|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^k - y^{k+1})\|^2\big) - \big(\frac{\tau}{2}\|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^{k-1} - y^k)\|^2\big) \\ & -2(1-\tau)\beta\|B(y^k - y^{k+1})\|^2. \end{array} \tag{4.16}$$

Proof. First, it follows from the equation $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$ that (3.3) can be written as

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

Analogously, for the previous iterate, we have

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

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

$$(y^k - y^{k+1})^T B^T(\lambda^k - \lambda^{k+1}) \geq (y^k - y^{k+1})^T D_0\big((y^k - y^{k+1}) - (y^{k-1} - y^k)\big).$$

Consequently, by using $D_0 = \tau D - (1-\tau)\beta B^T B$ (see (4.1)) and Cauchy-Schwarz inequality, it follows from the above inequality that

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

We get (4.16) from the above inequality immediately and the lemma is proved. □

Besides (4.16), we need to evaluate the term $(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1})$ in an another quadratic form as well.

Lemma 4.4 *Let $\{w^k\}$ be the sequence generated by the OLADMM* (1.14) *for the problem* (1.1) *and $\tilde{w}^k$ be defined by* (3.6)*. Then, for $\delta \in (0, 0.5)$, we have*

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq -(\frac{1}{4} + \frac{1}{2}\delta)\beta\|B(y^k - y^{k+1})\|^2 - (1-\delta)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2. \tag{4.19}$$

Proof. First, for $\delta \in (0, 1)$, by using the Cauchy-Schwarz inequality, we have

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq -\frac{1}{4(1-\delta)}\beta\|B(y^k - y^{k+1})\|^2 - (1-\delta)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2.$$

Notice that for $\delta \in (0, 0.5)$, we have

$$\frac{1}{4(1-\delta)} < \frac{1}{4} + \frac{1}{2}\delta,$$

and thus the proof of this lemma is complete. $\square$

We summarize the deduction in the following lemma.

Lemma 4.5 *Let $\{w^k\}$ be the sequence generated by the OLADMM* (1.14) *for the problem* (1.1) *and $\tilde{w}^k$ be defined by* (3.6). *Then, for $\delta = 2(\tau - \frac{3}{4})$, we have*

$$\begin{aligned} &(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \\ \geq\ & \big(\frac{\tau}{2}\|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^k - y^{k+1})\|^2\big) - \big(\frac{\tau}{2}\|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^{k-1} - y^k)\|^2\big) \\ & + \tau\|y^k - y^{k+1}\|_D^2 + 2\big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big). \end{aligned} \tag{4.20}$$

Proof. Since $\tau \in (0.75, 1)$ and $\delta = 2(\tau - \frac{3}{4})$, it follows that $\delta \in (0, 0.5)$ and thus (4.19) is valid. Using $\delta = 2(\tau - \frac{3}{4})$ and adding (4.16) and (4.19), we get

$$\begin{aligned} &2(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \\ \geq\ & \big(\frac{\tau}{2}\|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^k - y^{k+1})\|^2\big) - \big(\frac{\tau}{2}\|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^{k-1} - y^k)\|^2\big) \\ & -2(1-\tau)\beta\|B(y^k - y^{k+1})\|^2 - (\tau - \frac{1}{2})\beta\|B(y^k - y^{k+1})\|^2 - \big(1 - 2(\tau - \frac{3}{4})\big)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2. \end{aligned}$$

Substituting it into (4.15), we get

$$\begin{aligned} &(v^k - \tilde{v}^k)^T G(v^k - \tilde{v}^k) \\ \geq\ & \tau\|y^k - y^{k+1}\|_D^2 + \tau\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\ & + \big(\frac{\tau}{2}\|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^k - y^{k+1})\|^2\big) - \big(\frac{\tau}{2}\|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^{k-1} - y^k)\|^2\big) \\ & -2(1-\tau)\beta\|B(y^k - y^{k+1})\|^2 - (\tau - \frac{1}{2})\beta\|B(y^k - y^{k+1})\|^2 - \big(1 - 2(\tau - \frac{3}{4})\big)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\ =\ & \big(\frac{\tau}{2}\|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^k - y^{k+1})\|^2\big) - \big(\frac{\tau}{2}\|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^{k-1} - y^k)\|^2\big) \\ & + \tau\|y^k - y^{k+1}\|_D^2 + (2\tau - \frac{3}{2})\beta\|B(y^k - y^{k+1})\|^2 + 2(\tau - \frac{3}{4})\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\ =\ & \big(\frac{\tau}{2}\|y^k - y^{k+1}\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^k - y^{k+1})\|^2\big) - \big(\frac{\tau}{2}\|y^{k-1} - y^k\|_D^2 + \frac{1-\tau}{2}\beta\|B(y^{k-1} - y^k)\|^2\big) \\ & + \tau\|y^k - y^{k+1}\|_D^2 + 2\big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big). \end{aligned}$$

We then obtain the following theorem immediately; its proof is trivial based on the previous lemmas and propositions and thus omitted.

Theorem 4.1 *Let $\{w^k\}$ be the sequence generated by the OLADMM* (1.14) *for the problem* (1.1) *and $\tilde{w}^k$ be defined by* (3.6). *Setting $\tau \in (0.75, 1)$ in* (4.1), *we have*

$$\begin{aligned}
&\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) \\
\geq\ & \Big(\frac{1}{2}\|v - v^{k+1}\|_H^2 + \frac{1}{4}\big(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2\big)\Big) \\
&-\Big(\frac{1}{2}\|v - v^k\|_H^2 + \frac{1}{4}\big(\tau\|y^{k-1} - y^k\|_D^2 + (1-\tau)\beta\|B(y^{k-1} - y^k)\|^2\big)\Big) \\
&+\frac{1}{2}\tau\|y^k - y^{k+1}\|_D^2 + \big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big).
\end{aligned} \tag{4.21}$$

Note that this is a specific form of the inequality (4.12) with

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

and

$$\varphi(v^k, v^{k+1}) = \tau\|y^k - y^{k+1}\|_D^2 + 2\big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big).$$

5 Convergence

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

Lemma 5.1 *Let $\{w^k\}$ be the sequence generated by the OLADMM* (1.14) *for the problem* (1.1). *Then we have*

$$\begin{aligned}
&\|v^{k+1} - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2\big) \\
\leq\ & \|v^k - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^{k-1} - y^k\|_D^2 + (1-\tau)\beta\|B(y^{k-1} - y^k)\|^2\big) \\
&-\Big(\tau\|y^k - y^{k+1}\|_D^2 + 2\big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big)\Big).
\end{aligned} \tag{5.1}$$

Proof. Setting $w = w^*$ in (4.21) and performing simple manipulations, we get

$$\begin{aligned}
&\Big(\frac{1}{2}\|v^k - v^*\|_H^2 + \frac{1}{4}\big(\tau\|y^{k-1} - y^k\|_D^2 + (1-\tau)\beta\|B(y^{k-1} - y^k)\|^2\big)\Big) \\
\geq\ & \Big(\frac{1}{2}\|v^{k+1} - v^*\|_H^2 + \frac{1}{4}\big(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2\big)\Big) \\
&+\frac{1}{2}\tau\|y^k - y^{k+1}\|_D^2 + \big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big) \\
&+\big(\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*)\big).
\end{aligned} \tag{5.2}$$

For a solution point of (2.2), we have

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

Thus, the assertion (5.1) follows from (5.2) directly. □

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by the OLADMM* (1.14) *for the problem* (1.1). *Then the sequence $\{v^k\}$ converges to $v^\infty \in \mathcal{V}^*$.*

Proof. First, it follows from (5.1) that

$$\begin{aligned}
&\tau\|y^k - y^{k+1}\|_D^2 + 2\big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big) \\
&\quad \le \ \Big(\|v^k - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^{k-1} - y^k\|_D^2 + (1-\tau)\beta\|B(y^{k-1} - y^k)\|^2\big)\Big) \\
&\qquad -\Big(\|v^{k+1} - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^k - y^{k+1}\|_D^2 + (1-\tau)\beta\|B(y^k - y^{k+1})\|^2\big)\Big). \qquad (5.3)
\end{aligned}$$

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

$$\begin{aligned}
&\sum_{k=1}^{\infty}\Big(\tau\|y^k - y^{k+1}\|_D^2 + 2\big(\tau - \frac{3}{4}\big)\big(\beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big)\Big) \\
&\quad \le \ \|v^1 - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^0 - y^1\|_D^2 + (1-\tau)\beta\|B(y^0 - y^1)\|^2\big).
\end{aligned}$$

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

$$\lim_{k\to\infty}\|v^k - v^{k+1}\| = 0. \qquad (5.4)$$

For an arbitrarily fixed $v^* \in \mathcal{V}^*$, it follows from (5.1) that

$$\begin{aligned}
\|v^{k+1} - v^*\|_H^2 \ &\le \ \|v^k - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^{k-1} - y^k\|_D^2 + (1-\tau)\beta\|B(y^{k-1} - y^k)\|^2\big) \\
&\le \ \|v^1 - v^*\|_H^2 + \frac{1}{2}\big(\tau\|y^0 - y^1\|_D^2 + (1-\tau)\beta\|B(y^0 - y^1)\|^2\big), \quad \forall k \ge 1, \qquad (5.5)
\end{aligned}$$

and 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 (5.4) 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 (5.5) that

$$\|v^{k+1} - v^\infty\|_H^2 \ \le \ \|v^k - v^\infty\|_H^2 + \frac{1}{2}\big(\tau\|y^{k-1} - y^k\|_D^2 + (1-\tau)\beta\|B(y^{k-1} - y^k)\|^2\big). \qquad (5.6)$$

Note that v^∞ is also the limit point of $\{v^{k_j}\}$. Together with (5.4), this fact means that it is impossible that the sequence $\{v^k\}$ has more than one cluster point. Therefore, the sequence $\{v^k\}$ converges to v^∞ and the proof is complete. □

6 Optimality of τ

In Section 4, we show that $\tau \in (0.75, 1)$ is sufficient to ensure the convergence of the OLADMM (1.14). In this section, we show by an extremely simple example that the convergence of (1.14) is not guaranteed for any $\tau \in (0, 0.75)$. Hence, 0.75 is the optimal value for the linearization parameter τ of the OLADMM (1.14).

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

$$\min\{0 \cdot x + 0 \cdot y \mid 0 \cdot x + y = 0, \ x \in \{0\}, y \in \Re\}. \qquad (6.1)$$

Without loss of generality, we take $\beta = 1$ and thus the augmented Lagrangian function of the problem (6.1) is

$$\mathcal{L}_1(x, y, \lambda) = -\lambda^T y + \frac{1}{2}\|y\|^2.$$

The iterative scheme of the OLADMM (1.14) for (6.1) is

$$\text{(OLADMM)} \quad \begin{cases} x^{k+1} = \arg\min\big\{\mathcal{L}_1(x, y^k, \lambda^k) \;\big|\; x \in \{0\}\big\}, & \text{(6.2a)} \\ y^{k+1} = \arg\min\big\{-y^T\lambda^k + \frac{1}{2}\|y\|^2 + \frac{1}{2}\|y - y^k\|^2_{D_0} \;\big|\; y \in \Re\big\}, & \text{(6.2b)} \\ \lambda^{k+1} = \lambda^k - (y^{k+1}). & \text{(6.2c)} \end{cases}$$

Since $\beta = 1$ and $B^T B = 1$, it follows from (4.1) 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 (6.2) becomes

$$\begin{cases} x^{k+1} = 0, \\ -\lambda^k + y^{k+1} + (\tau r - 1)(y^{k+1} - y^k) = 0, \\ \lambda^{k+1} = \lambda^k - y^{k+1}. \end{cases} \tag{6.3}$$

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

$$\begin{cases} \alpha y^{k+1} = \lambda^k + (\alpha - 1)y^k, \\ \lambda^{k+1} = \lambda^k - y^{k+1}. \end{cases} \tag{6.4}$$

With elementary manipulations, we get

$$\begin{cases} y^{k+1} = \dfrac{\alpha - 1}{\alpha} y^k + \dfrac{1}{\alpha}\lambda^k, \\ \lambda^{k+1} = \dfrac{1 - \alpha}{\alpha} y^k + \dfrac{\alpha - 1}{\alpha}\lambda^k, \end{cases} \tag{6.5}$$

which can be written as

$$v^{k+1} = P(\alpha)v^k \qquad \text{with} \qquad P(\alpha) = \frac{1}{\alpha}\begin{pmatrix} \alpha - 1 & 1 \\ 1 - \alpha & \alpha - 1 \end{pmatrix}. \tag{6.6}$$

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 (6.6) has an eigenvalue less than -1. Hence, the iterative scheme (6.5), i.e., the application of the OLADMM (1.14) to the problem (6.1), is not necessarily convergent for any $\tau \in (0, 0.75)$.

7 Convergence rate

In this section, we establish the worst-case $O(1/t)$ convergence rate measured by the iteration complexity for the OLADMM (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 positive-definite linearized version (1.11)-(1.12) (actually, the matrix D in (1.11) could be positive semidefinite) has been established in [22].

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

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

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

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

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

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

where

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

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

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

Theorem 4.1 is again the starting point of the analysis.

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

$$\begin{aligned} &\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \\ &\quad \leq \; \frac{1}{2t}\Big(\|v - v^1\|_H^2 + \frac{1}{2}\big(\tau\|y^0 - y^1\|_D^2 + (1-\tau)\beta\|B(y^0 - y^1)\|^2\big)\Big). \end{aligned} \tag{7.2}$$

where

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

Proof. First, it follows from (4.21) that

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

Thus, we have

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

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

$$\begin{aligned}
&\sum_{k=1}^t \theta(\tilde{u}^k)-t\theta(u)+(\sum_{k=1}^t \tilde{w}^k-tw)^T F(w)\\
&\quad\leq\ \frac{1}{2}\|v-v^1\|_H^2+\frac{1}{4}\big(\tau\|y^0-y^1\|_D^2+(1-\tau)\beta\|B(y^0-y^1)\|^2\big).
\end{aligned}$$

and thus

$$\begin{aligned}
&\frac{1}{t}(\sum_{k=1}^t \theta(\tilde{u}^k))-\theta(u)+(\tilde{w}_t-w)^T F(w)\\
&\quad\leq\ \frac{1}{2t}\Big(\|v-v^1\|_H^2+\frac{1}{2}\big(\tau\|y^0-y^1\|_D^2+(1-\tau)\beta\|B(y^0-y^1)\|^2\big)\Big).
\end{aligned} \tag{7.5}$$

Since $\theta(u)$ is convex and

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

we have that

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

Substituting it into (7.5), the assertion of this theorem follows directly. $\square$

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

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

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

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

which means $\tilde{w}_t$ is an approximate solution of the variational inequality (2.2) 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 Conclusions

In this paper, we found the optimal step size, or linearization parameter, for the linearized version of alternating direction method of multipliers (ADMM) in the convex programming context. In the context of ADMM with proximal regularization, our optimal step size means it is not necessary to ensure the positive definiteness or semidefiniteness of the proximal regularization term. Without any additional assumptions on the model per se, the proposed optimal linearized ADMM allows larger step sizes for subproblems, theoretically keeps the convergence and convergence rate, and inherit the main feature of the original linearized ADMM which is tailored for applications as well. This work can be regarded as a more extensive analysis of [18] in which the optimal step size is shown for the proximal version of the augmented Lagrangian method. The efficiency of the optimal linearized ADMM can be immediately verified by various examples such as the sparse and low-rank models in [28]; only a very slight modification of multiplying the value of r by a coefficient of 0.75 in our own code can immediately yield about $20-30\%$ acceleration. We omit the detail of numerical experiments for succinctness; and we believe such an acceleration can be easily found by other well-studied applications of the linearized ADMM in the literature.

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 can be further extended to many other variants of the 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 [21], 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 setting still represents the simplest yet most fundamental case for finding the optimal step size of the linearized ADMM; and it is the basis of possible discussions 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, 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, 2016.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[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] 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.

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

[29] 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.