INDEFINITE LINEARIZED AUGMENTED LAGRANGIAN METHOD FOR CONVEX PROGRAMMING WITH LINEAR INEQUALITY CONSTRAINTS*

BINGSHENG HE†, SHENGJIE XU‡, AND JING YUAN§

May 4, 2021

Abstract. The augmented Lagrangian method (ALM) is a benchmark for tackling the convex optimization problem with linear constraints; ALM and its variants for linearly equality-constrained convex minimization models have been well studied in the literatures. However, much less attention has been paid to ALM for efficiently solving the linearly inequality-constrained convex minimization model. In this paper, we exploit an enlightening reformulation of the most recent indefinite linearized (equality-constrained) ALM, and present a novel indefinite linearized ALM scheme for efficiently solving the convex optimization problem with linear inequality constraints. The proposed method enjoys great advantages, especially for large-scale optimization cases, in two folds mainly: first, it significantly simplifies the optimization of the challenging key subproblem of the classical ALM by employing its linearized reformulation, while keeping low complexity in computation; second, we prove that a smaller proximity regularization term is needed for convergence guarantee, which allows a bigger step-size and can largely reduce required iterations for convergence. Moreover, we establish an elegant global convergence theory of the proposed scheme upon its equivalent compact expression of prediction-correction, along with a worst-case $\mathcal{O}(1/N)$ convergence rate. Numerical results demonstrate that the proposed method can reach a faster converge rate for a higher numerical efficiency as the regularization term turns smaller, which confirms the theoretical results presented in this study.

Key words. augmented Lagrangian method, convex programming, convergence analysis, inequality constraints, image segmentation

AMS subject classifications. 90C25, 90C30, 65K05, 65K10

1. Introduction.

In this work, we study the canonical convex minimization problem with linear inequality constraints:

$$\min\{\theta(x) \mid Ax \geq b,\ x \in \mathcal{X}\}, \tag{1.1}$$

where $\theta(x)$: $\Re^n \to \Re$ is a closed proper convex but not necessarily smooth function, $\mathcal{X} \subset \Re^n$ is a nonempty closed convex set, $A \in \Re^{m\times n}$ and $b \in \Re^m$. Throughout our discussion, the solution set of (1.1) is assumed to be nonempty, and $\rho(\cdot)$ is used to stand for the spectrum radius of a matrix.

In practice, the model (1.1) finds many applications in, e.g. linear and nonlinear programming problems [1, 6, 25], variational image processing models [33, 34, 35, 36], and machine learning [9, 10, 32] etc. Especially, the classical linearly equality-constrained convex optimization problem

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}, \tag{1.2}$$

*Submitted to the editors DATE.

Funding: B.S. He acknowledges the support of National Natural Science Foundation of China (NSFC) under grant No. 11871029. S.J. Xu acknowledges the support of the Guangdong Basic and Applied Basic Research Foundation of China under grant No. 2018A0303130123. J. Yuan acknowledges the support of the NSFC under grant No. 61877047.

†Department of Mathematics, Nanjing University, Nanjing, China. (hebma@nju.edu.cn).

‡Department of Mathematics, Harbin Institute of Technology, Harbin, China, and Department of Mathematics, Southern University of Science and Technology, Shenzhen, China. (xsjnsu@163.com).

§Corresponding author. School of Mathematics and Statistics, Xidian University, Xi'an, China. (jyuan@xidian.edu.cn).

can be simply taken as its special case, for which the augmented Lagrangian method [22, 28] was developed as the fundamental tool by imposing the quadratic penalty term $\beta/2\|Ax-b\|^2$ on the linear equality constraints $Ax=b$ for constructing the augmented Lagrangian function w.r.t. (1.2):

$$\mathcal{L}_\beta(x,\lambda) = \theta(x) - \lambda^T(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2, \tag{1.3}$$

where $\beta > 0$ and $\lambda \in \Re^m$ is the associated Lagrange multiplier of (1.2); moreover, each k-th iteration of the classic ALM for the linearly equality-constrained minimization problem (1.2) explores the following two optimization steps:

$$\text{(Equality-ALM)} \left\{ \begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,\lambda^k) \mid x \in \mathcal{X}\}, & \text{(1.4a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}-b). & \text{(1.4b)} \end{array} \right.$$

The method (1.4) plays a significant role in both theoretical and algorithmic aspects for a number of convex minimization problems. We refer to, e.g. [3, 4, 5, 12, 13, 23] for the vast volume of related literatures. In particular, as analyzed in [29, 30], the classic ALM (1.4) is essentially an application of the proximal point algorithm (PPA) in [26] to the dual of (1.2).

Likewise, the classical way to solve the studied optimization problem (1.1) with linear inequality constraints is to apply the inequality-constrained version of ALM (see [3, 4] for more details), where its k-th iteration computes $w^{k+1} := (x^{k+1};\lambda^{k+1})$ simply by exploring the following two steps:

$$\text{(Inequality-ALM)} \left\{ \begin{array}{ll} x^{k+1} = \arg\min\left\{\mathcal{L}_\beta(x,\lambda^k) \mid x \in \mathcal{X}\right\}, & \text{(1.5a)} \\ \lambda^{k+1} = [\lambda^k - \beta(Ax^{k+1}-b)]_+, & \text{(1.5b)} \end{array} \right.$$

given the defined augmented Lagrangian function for (1.1):

$$\mathcal{L}_\beta(x,\lambda) := \theta(x) + \frac{1}{2\beta}\Big\{\|[\lambda - \beta(Ax-b)]_+\|^2 - \lambda^T\lambda\Big\}. \tag{1.6}$$

Here, $[\cdot]_+$ denotes the standard projection operator on the non-negative orthant in Euclidean space. Clearly, the essential step for implementing the inequality-constrained ALM (1.5) is to tackle the primal subproblem (1.5a). However, the gradient of the smoothness term over x

$$\nabla_x(\|[\lambda^k - \beta(Ax-b)]_+\|^2) = -2\beta A^T[\lambda^k - \beta(Ax-b)]_+$$

is non-smooth, which is the main obstacle for efficiently performing such an ALM scheme in practice.

Recent developments on the linearized (equality-constrained) ALM (see subsection 2.2 for more details), which adds an additional proximal term w.r.t. a matrix L (will be given in (2.8)) to linearize the classic augmented Lagrangian function so as to properly reduces the computational difficulty of the pivotal subproblem (1.4a), show that the positive definiteness of the added proximal term for linearization is not needed for convergence and result in an efficient indefinite linearized ALM. This allows a bigger step-size to speed-up convergence. Motivated by this discovery, the primary purpose of this paper is to develop a novel indefinite linearized ALM scheme for optimizing the convex minimization problem (1.1) with linear inequality constraints.

To our knowledge, this is the first work to study the indefinite linearized ALM framework for efficiently tackling the canonical linear-inequality constrained convex minimization problem. The proposed method enjoys great advantages in many folds: first, compared with the classical inequality-constrained ALM (1.5), the new method provides a much easier and more effective optimization strategy for solving its core subproblem, while still keeping the same computational complexities as the one in the linearized (equality-constrained) ALM; second, the introduced relaxation of the proximal regularization term allows a bigger step-size, and can largely reduce required iterations to achieve convergence; third, the novel proposed method covers the most recent studies on the indefinite linearized (equality-constrained) ALM, as its special case. In addition, we present an elegant equivalent prediction-correction expression of the proposed algorithm and establish the global convergence analysis upon it. We illustrate the numerical efficiency of the proposed method by extended experiments on the support vector machine for classification and continuous max-flow models for image segmentation.

The rest of the paper is organized as follows. In section 2, we summarize some preliminaries to motivate the proposed indefinite linearized ALM scheme for the convex minimization problem (1.1) with linear inequality constraints. In section 3, inspired by a new representation of the indefinite linearized (equality-constrained) ALM, we present a novel indefinite linearized ALM scheme for solving the studied problem (1.1) with linear inequality constraints. We turn to determine a smaller regularization term of the proposed method, which actually allows a considerable speed-up in computation, and show its global convergence in section 4. The numerical experiments are further conducted in section 5, which numerically confirms the theoretical results presented in this study. Finally, some conclusions and future studies are discussed in section 6.

2. Preliminaries.

In this section, we summarize some preliminary results which are useful for further discussion.

2.1. A variational characterization of (1.1).

The analysis of the paper is based on the following fundamental lemma of the variational inequality optimality condition (details can be found in, e.g., [2]).

LEMMA 2.1. *Let $\mathcal{X} \subset \Re^n$ be a nonempty closed convex set, $\theta(x)$ and $f(x)$ be convex functions. If $f(x)$ is differentiable on an open set which contains $\mathcal{X}$ and the solution set of the convex minimization problem $\min\{\theta(x)+f(x) \mid x \in \mathcal{X}\}$ is nonempty, then we have*

$$x^* \in \arg\min \left\{\theta(x) + f(x) \;\middle|\; x \in \mathcal{X}\right\}$$

if and only if

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

First of all, we show how to represent the optimality condition of (1.1) in the variational inequality context. By adding the Lagrange multiplier $\lambda \in \Re^m_+$, where $\Re^m_+ := \{x \in \Re^m \mid x \geq 0\}$, to the inequality constraints, the Lagrange function of (1.1) then reads as

$$L(x, \lambda) = \theta(x) - \lambda^T(Ax - b). \tag{2.1}$$

Let $(x^*, \lambda^*) \in \mathcal{X} \times \Re^m_+$ be a saddle point of (2.1). Then, for any $x \in \mathcal{X}$ and $\lambda \geq 0$, we have

$$L_{\lambda \in \Re^m_+}(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L_{x \in \mathcal{X}}(x, \lambda^*).$$

According to Lemma 2.1, these two inequalities can be alternatively rewritten as

$$\begin{cases} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \quad \forall\, x \in \mathcal{X}; \\ \lambda^* \in \Re^m_+, & (\lambda - \lambda^*)^T(Ax^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m_+, \end{cases} \tag{2.2}$$

or more compactly,

$$\text{VI}(\Omega, F, \theta): \quad w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.3}$$

where Ω^* denotes the solution set of $\text{VI}(\Omega, F, \theta)$ and is assumed to be nonempty,

$$\Omega = \mathcal{X} \times \Re^m_+, \quad w = \begin{pmatrix} x \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}. \tag{2.4}$$

Since the operator $F(\cdot)$ in (2.4) is affine with a skew-symmetric matrix, it implies that

$$(u - v)^T(F(u) - F(v)) \equiv 0, \quad \forall\, u, v \in \Omega, \tag{2.5}$$

which means F is monotone.

2.2. Motivations of indefinite linearized ALM. Clearly, the essential step for implementing the equality-constrained ALM (1.4) is to solve the problem (1.4a), for which the proximal augmented Lagrangian method (Prox-ALM) (see, e.g., [18, 37]) provides a significant strategy by considering the proximal regularized version of ALM:

$$\begin{cases} x^{k+1} = \arg\min\Big\{\mathcal{L}_\beta(x, \lambda^k) + \frac{1}{2}\|x - x^k\|_D^2 \;\Big|\; x \in \mathcal{X}\Big\}, & \text{(2.6a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). & \text{(2.6b)} \end{cases}$$

Here, we define $\|x\|_D^2 := x^T D x$ for any symmetric matrix D. As stated in [18], the proximal matrix $D \in \Re^{n\times n}$ given in (2.6) is generally required to be symmetric and positive-definite (or equivalently, $\|\cdot\|_D$ is a norm) to ensure convergence. The pivot subproblem (2.6a) which can be equivalently reformulated as

$$\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}}\Big\{\theta(x) - \lambda^k(Ax - b) + \frac{\beta}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x^k\|_D^2\Big\} \\ &= \arg\min_{x\in\mathcal{X}}\Big\{\theta(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|A(x - x^k) + Ax^k - b\|^2 + \frac{1}{2}\|x - x^k\|_D^2\Big\} \\ &= \arg\min_{x\in\mathcal{X}}\Big\{\theta(x) - x^TA^T[\lambda^k - \beta(Ax^k - b)] + \frac{\beta}{2}\|A(x - x^k)\|^2 + \frac{1}{2}\|x - x^k\|_D^2\Big\}, \end{aligned} \tag{2.7}$$

its solution in general, as discussed in e.g., [11, 18], can only be approximated by certain iterative approach.

On the other hand, if the positive definite matrix D in (2.6) is taken as

$$D := rI_n - \beta A^TA \quad \text{with} \quad r > \beta\rho(A^TA), \tag{2.8}$$

the subproblem (2.6a) or (2.7) can be essentially reduced to

$$\begin{aligned} x^{k+1} &= \arg\min\Big\{\theta(x) - x^TA^T[\lambda^k - \beta(Ax^k - b)] + \frac{r}{2}\|x - x^k\|^2 \;\Big|\; x \in \mathcal{X}\Big\}, \\ &= \arg\min\Big\{\theta(x) + \frac{r}{2}\|x - x^k - \frac{1}{r}A^T[\lambda^k - \beta(Ax^k - b)]\|^2 \;\Big|\; x \in \mathcal{X}\Big\}, \end{aligned} \tag{2.9}$$

i.e., the proximity operator of $\theta(x)$ when $\mathcal{X} = \Re^n$, which thus leads to the following so-called linearized ALM:

$$\begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}_\beta(x,\lambda^k) + \frac{1}{2}\|x - x^k\|_D^2 \mid x \in \mathcal{X}\right\}, & (2.10a)\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). & (2.10b)\end{cases}$$

Such linearization is particularly meaningful for the situation when the proximity operator of the objective function $\theta(x)$ defined via

$$\text{Prox}_{\theta,\beta}(y) = \arg\min\left\{\theta(x) + \frac{1}{2\beta}\|x - y\|^2 \mid x \in \Re^n\right\}, \tag{2.11}$$

has a closed-form solution. For instance, the solution of (2.11) can be given by the soft thresholding function (see, e.g., Example 6.8 in [2]) when $\theta(x) = \|x\|_1$. We also refer to the review paper [27] for more specific examples.

Meanwhile, it is easy to see that a smaller r in (2.9) results in a bigger step-size at each iteration, which potentially leads to less iterations, hence a faster convergence rate (see, e.g., [15, 17] for the related works). Recently, an interesting work [18] introduced the linearized ALM with the optimal limit of step size:

$$\text{(IDL-ALM)}\begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}_\beta(x,\lambda^k) + \frac{1}{2}\|x - x^k\|_{D_0}^2 \mid x \in \mathcal{X}\right\}, & (2.12a)\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b), & (2.12b)\end{cases}$$

in which

$$D_0 = \tau r I - \beta A^T A \ \text{ with } \ r > \beta\rho(A^T A) \ \text{ and } \ \tau \in (0.75, 1). \tag{2.13}$$

It is easy to see that the proximal matrix D_0 is indefinite when $\tau \in (0.75, 1)$ given $r = \beta\rho(A^T A)$, which implies that $\|\cdot\|_{D_0}$ is not necessarily a norm; the subproblem (2.12a) is also properly linearized as (2.9). With this regard, we call the scheme (2.12) *indefinite linearized augmented Lagrangian method* (abbreviated as IDL-ALM). In view of the positive definite matrix D defined in (2.8), D_0 can also be denoted by

$$D_0 = \tau D - (1-\tau)\beta A^T A. \tag{2.14}$$

2.3. New insights to IDL-ALM. Now we give an enlightening reformulation of the indefinite linearized ALM (2.12), which brings some new theoretical insights to the proposed numerical scheme for the studied convex minimization problem (1.1) with linear inequality constraints.

Recall (2.7) and set $\tilde{\lambda}^k := \lambda^k - \beta(Ax^k - b)$. The subproblem (2.12a) can be expressed as

$$\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}}\left\{\theta(x) - x^T A^T[\lambda^k - \beta(Ax^k - b)] + \frac{\beta}{2}\|A(x - x^k)\|^2 + \frac{1}{2}\|x - x^k\|_{D_0}^2\right\}\\ &= \arg\min_{x\in\mathcal{X}}\left\{\theta(x) - (\tilde{\lambda}^k)^T Ax + \frac{\beta}{2}\|A(x - x^k)\|^2 + \frac{1}{2}\|x - x^k\|_{(\tau r I - \beta A^T A)}^2\right\}\\ &= \arg\min_{x\in\mathcal{X}}\left\{\theta(x) - (\tilde{\lambda}^k)^T Ax + \frac{\tau r}{2}\|x - x^k\|^2\right\}, \end{aligned}$$

and the update λ^{k+1} in (2.12b) can be further rewritten as

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \beta(Ax^{k+1} - b) = \lambda^k - \beta(Ax^k - b) + \beta A(x^k - x^{k+1})\\ &= \tilde{\lambda}^k + \beta A(x^k - x^{k+1}). \end{aligned}$$

Consequently, we can equivalently represent the indefinite linearized ALM (2.12) as

$$\begin{cases} \tilde{\lambda}^k = \lambda^k - \beta(Ax^k - b), & \text{(2.15a)} \\ x^{k+1} = \arg\min\big\{\theta(x) - (\tilde{\lambda}^k)^T Ax + \frac{\tau r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(2.15b)} \\ \lambda^{k+1} = \tilde{\lambda}^k + \beta A(x^k - x^{k+1}). & \text{(2.15c)} \end{cases}$$

Indeed, this new representation (2.15) of the proposed indefinite linearized ALM (2.12) directly motivates the main study of this work.

3. Indefinite linearized ALM for (1.1).

In this work, we aim to efficiently solve the linearly inequality-constrained convex minimization problem (1.1). The essential step of the classical ALM approach for (1.1) is to tackle the subproblem (1.5a), which is, however, often complicated in practice and has no efficient solution in general. With this respect, we now introduce a novel efficient indefinite linearized ALM-based method for the convex minimization problem (1.1) with linear inequality constraints, which is motivated by the IDL-ALM formulation (2.15) and can properly avoid solving such a difficult optimization problem (1.5a) directly.

Observe the IDL-ALM representation (2.15). We first list the essential steps at k-th iteration of the new IDL-ALM-based method for (1.1), which begins with the given $w^k = (x^k; \lambda^k)$ and generates the new iterate $w^{k+1} = (x^{k+1}; \lambda^{k+1})$, as follows:

$$\begin{cases} \tilde{\lambda}^k = [\lambda^k - \beta(Ax^k - b)]_+, & \text{(3.1a)} \\ x^{k+1} = \arg\min\big\{\theta(x) - (\tilde{\lambda}^k)^T Ax + \frac{\tau r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(3.1b)} \\ \lambda^{k+1} = \tilde{\lambda}^k + \beta A(x^k - x^{k+1}), & \text{(3.1c)} \end{cases}$$

where $r > \beta\rho(A^T A)$ and $\tau \in (0.75, 1)$. Instead of $\tilde{\lambda}^k = \lambda^k - \beta(Ax^k - b)$ in (2.15a), we use $\tilde{\lambda}^k = [\lambda^k - \beta(Ax^k - b)]_+$ in (3.1a).

Clearly, the proposed method (3.1) reduces solving the difficult core subproblem (1.5a) at each iteration of the classical ALM to tackling a much easier proximal estimation which may have a direct close-form solver in many real-world applications. In particular, under the novel scheme (3.1), the original IDL-ALM (2.12) can be regarded as its special case when the free multiplier λ, no need of projection, for the corresponding linear equality constraints is considered. With this regard, we also name (3.1) *the inequality version of indefinite linearized augmented Lagrangian method* (abbreviated as I-IDL-ALM), which has the same computational complexity as the linearized (equality-constrained) ALM (2.10).

3.1. Prediction-correction interpretation of (3.1).

Now we show that the new I-IDL-ALM scheme (3.1) can be interpreted as a prediction-correction type method. Moreover, the associated variational inequality structure of its prediction step is deduced.

Let $\tilde{x}^k := x^{k+1}$, the novel method (3.1) can be explained as a prediction-correction scheme, whose predictor and corrector are

$$\text{(3.2)} \qquad \begin{cases} \tilde{\lambda}^k = [\lambda^k - \beta(Ax^k - b)]_+, \\ \tilde{x}^k = \arg\min\big\{\theta(x) - (\tilde{\lambda}^k)^T Ax + \frac{\tau r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, \end{cases}$$

and

$$\text{(3.3)} \qquad \begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ -\beta A & I_m \end{pmatrix} \begin{pmatrix} x^k - \tilde{x}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix},$$

respectively. To this end, we would emphasize that such a two-stage explanation (3.2)-(3.3) only serves for further theoretical analysis and there is no need to use it when implementing the novel method (3.1).

By the fundamental Lemma 2.1, the output of the x-subproblem in (3.2) satisfies the following variational inequality:

$$\tilde{x}^k \in \mathcal{X}, \;\; \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k + \tau r(\tilde{x}^k - x^k)\} \geq 0, \;\; \forall\, x \in \mathcal{X}. \tag{3.4}$$

For the λ-subproblem in (3.2), according to the basic property of the projection operator onto a convex set (see, e.g., Appendix B of [25]), we have

$$\tilde{\lambda}^k \in \Re^m_+, \;\; (\lambda - \tilde{\lambda}^k)^T\{\lambda^k - \beta(Ax^k - b) - \tilde{\lambda}^k\} \leq 0, \;\; \forall\, \lambda \in \Re^m_+,$$

which can be further rewritten as

$$\tilde{\lambda}^k \in \Re^m_+, \;\; (\lambda - \tilde{\lambda}^k)^T\big\{(A\tilde{x}^k - b) - A(\tilde{x}^k - x^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\big\} \geq 0, \;\; \forall\, \lambda \in \Re^m_+. \tag{3.5}$$

Observe (3.4)-(3.5) and the notations in (2.4). The variational inequalities w.r.t. (3.2) can be compactly written as

(Prediction step)

$$\tilde{w}^k \in \Omega, \;\; \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k) \geq (w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k), \;\; \forall\, w \in \Omega, \tag{3.6}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix} \in \Omega\,, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} \tau r I_n & 0 \\ -A & \frac{1}{\beta}I_m \end{pmatrix}. \tag{3.7}$$

And the output w^{k+1} in the corrector (3.3) can be succinctly given by

(Correction step)

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k), \tag{3.8}$$

where

$$M = \begin{pmatrix} I_n & 0 \\ -\beta A & I_m \end{pmatrix}. \tag{3.9}$$

3.2. Definitions and properties of basic matrices. In this subsection, we define some basic matrices which are useful to ease the convergence analysis of the proposed I-IDL-ALM scheme (3.1) in the following section.

Recall the matrices Q and M defined in (3.7) and (3.9), respectively. If we set $H = QM^{-1}$, then we have

$$H = QM^{-1} = \begin{pmatrix} \tau r I_n & 0 \\ -A & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ \beta A & I_m \end{pmatrix} = \begin{pmatrix} \tau r I_n & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{3.10}$$

It is easy to see that the matrix H is symmetric and positive-definite for any $\tau > 0$, $r > 0$ and $\beta > 0$.

Since $H = QM^{-1}$, we obtain $M^THM = M^TQ$, and thus

$$M^THM = M^TQ = \begin{pmatrix} I_n & -\beta A^T \\ 0 & I_m \end{pmatrix} \begin{pmatrix} \tau r I_n & 0 \\ -A & \frac{1}{\beta} I_m \end{pmatrix} = \begin{pmatrix} \tau r I_n + \beta A^T A & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix}.$$

Setting $G = Q^T + Q - M^THM$, it follows that

$$G = (Q^T + Q) - M^THM = \begin{pmatrix} 2\tau r I_n & -A^T \\ -A & \frac{2}{\beta} I_m \end{pmatrix} - \begin{pmatrix} \tau r I_n + \beta A^T A & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix}$$

$$= \begin{pmatrix} \tau r I_n - \beta A^T A & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \overset{(2.13)}{=} \begin{pmatrix} D_0 & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.11}$$

Note that D_0 defined in (2.13) is a symmetric but not necessarily positive-definite matrix. $\|\cdot\|_G$ (where $\|u\|_G^2 := u^TGu$) is not necessarily a norm.

4. Convergence analysis. In this section, we show the in-depth choice of the regularization parameter τ for the proximal subproblem of x^{k+1} in the proposed indefinite linearized ALM scheme (3.1), which is the main numerical load for (3.1), so as to reach its convergence efficiently with less iterations. Moreover, we establish an elegant global convergence theory for the proposed I-IDL-ALM method (3.1).

4.1. In-depth choice on τ. It is obvious that a smaller τ for the proximal optimization of x^{k+1} in the proposed method (3.1) presents a bigger step size for the proximity approximation of x^{k+1}, which actually saves iterations to achieve a faster convergence. Indeed, we show that the value of τ can be even less than 1, actually within $(0.75, 1)$, to achieve convergence, which exactly means that a positive indefinite matrix D, by (2.13), is taken for proximity approximation. We also prove some key inequalities which are used for establishing the global convergence theory of the proposed method (3.1) in subsection 4.2.

We first demonstrate the following lemma:

LEMMA 4.1. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prediction-correction formulation (3.2)-(3.3) of the proposed indefinite linearized ALM (3.1) for the optimization problem (1.1). Then we have*

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(w) \geq \frac{1}{2}\big\{\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2\big\} + \frac{1}{2}\|w^k - \tilde{w}^k\|_G^2, \ \ \forall\, w \in \Omega, \tag{4.1}$$

where G is given in (3.11).

Proof. Given (3.8), it is easy to see that $w^k - w^{k+1} = M(w^k - \tilde{w}^k)$; hence, in view of (3.10), we have $H(w^k - w^{k+1}) = Q(w^k - \tilde{w}^k)$ and

$$(w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k) = (w - \tilde{w}^k)^T H(w^k - w^{k+1}). \tag{4.2}$$

Recall the basic monotone property (2.5) of the operator F. It follows that

$$\begin{aligned} \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(w) &\overset{(2.5)}{=} \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\overset{(3.6)}{\geq} (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k) \\ &\overset{(4.2)}{=} (w - \tilde{w}^k)^T H(w^k - w^{k+1}). \end{aligned} \tag{4.3}$$

Applying the identity

$$(a-b)^T H(c-d) = \frac{1}{2}\big\{\|a-d\|_H^2 - \|a-c\|_H^2\big\} + \frac{1}{2}\big\{\|c-b\|_H^2 - \|d-b\|_H^2\big\}$$

to the right hand of (4.2) with $a = w,\ b = \tilde{w}^k,\ c = w^k$ and $d = w^{k+1}$, we obtain

$$\begin{aligned} &(w-\tilde{w}^k)^T H(w^k - w^{k+1}) \\ &= \frac{1}{2}\{\|w-w^{k+1}\|_H^2 - \|w-w^k\|_H^2\} + \frac{1}{2}\{\|w^k-\tilde{w}^k\|_H^2 - \|w^{k+1}-\tilde{w}^k\|_H^2\}. \end{aligned} \tag{4.4}$$

Meanwhile, since

$$\begin{aligned} &\|w^k-\tilde{w}^k\|_H^2 - \|w^{k+1}-\tilde{w}^k\|_H^2 \\ &= \|w^k-\tilde{w}^k\|_H^2 - \|(w^k-\tilde{w}^k) - (w^k - w^{k+1})\|_H^2 \\ &\overset{(3.8)}{=} \|w^k-\tilde{w}^k\|_H^2 - \|(w^k-\tilde{w}^k) - M(w^k-\tilde{w}^k)\|_H^2 \\ &= 2(w^k-\tilde{w}^k)^T HM(w^k-\tilde{w}^k) - (w^k-\tilde{w}^k)^T M^T HM(w^k-\tilde{w}^k) \\ &= (w^k-\tilde{w}^k)^T(Q^T + Q - M^T HM)(w^k-\tilde{w}^k) \\ &\overset{(3.11)}{=} \|w^k-\tilde{w}^k\|_G^2, \end{aligned} \tag{4.5}$$

the assertion of the lemma follows directly by substituting (4.4) and (4.5) into (4.3).□

Upon Lemma 4.1, if the matrix G defined in (3.11) is positive-definite (it holds naturally when $\tau \geq 1$), the core inequality (4.1) can essentially imply the global convergence and a worst-case convergence rate measured by iteration complexity for the proposed method. We refer to, e.g., [14, 16, 20, 21], for analytical details. However, we are interested in whether a smaller τ is feasible, which implies the matrix G is not necessarily positive-definite; with this respect, the fundamental result of Lemma 4.1 can not be used directly, which makes the convergence analysis for the novel method (3.1) more challenging.

Motivated by the works in, e.g., [17, 18, 19], our main goal in the following is to show the term $\|w^k-\tilde{w}^k\|_G^2$ in (4.1) satisfies

$$\|w^k-\tilde{w}^k\|_G^2 \geq \phi(w^k, w^{k+1}) - \phi(w^{k-1}, w^k) + \varphi(w^k, w^{k+1}), \tag{4.6}$$

where $\phi(\cdot,\cdot)$ and $\varphi(\cdot,\cdot)$ are both non-negative functions, and $\varphi(\cdot,\cdot)$ is used to measure how much w^{k+1} fails to be a solution point of (2.3). Once the above inequality is established, by substituting (4.6) into (4.1), we can immediately obtain

$$\begin{aligned} &\theta(x) - \theta(\tilde{x}^k) + (w-\tilde{w}^k)^T F(w) \\ &\geq \frac{1}{2}\big\{\|w-w^{k+1}\|_H^2 + \phi(w^k, w^{k+1})\big\} - \frac{1}{2}\big\{\|w-w^k\|_H^2 + \phi(w^{k-1}, w^k)\big\} \\ &\quad + \varphi(w^k, w^{k+1}), \quad \forall\, w \in \Omega. \end{aligned} \tag{4.7}$$

Furthermore, as will be shown in subsection 4.2, the inequality (4.7) is the essential property for establishing the convergence analysis of the proposed method (3.1).

To show the pivotal inequality (4.6), we prove first a lemma.

LEMMA 4.2. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prediction-correction formulation (3.2)-(3.3) of the proposed indefinite linearized ALM (3.1) for solving (1.1). Then we have*

$$\begin{aligned} \|w^k-\tilde{w}^k\|_G^2 = {}& \tau\|x^k-x^{k+1}\|_D^2 + \tau\beta\|A(x^k-x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2 \\ & + 2(\lambda^k-\lambda^{k+1})^T A(x^k-x^{k+1}). \end{aligned} \tag{4.8}$$

Proof. To begin with, it follows from (3.11) and (2.14) that

$$\begin{aligned}\|w^k - \tilde{w}^k\|_G^2 &\overset{(3.11)}{=} \|x^k - \tilde{x}^k\|_{D_0}^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \\ &\overset{(2.14)}{=} \tau\|x^k - \tilde{x}^k\|_D^2 - (1-\tau)\beta\|A(x^k - \tilde{x}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2. \end{aligned} \tag{4.9}$$

According to (3.3), we have

$$\tilde{x}^k = x^{k+1} \quad \text{and} \quad \lambda^k - \tilde{\lambda}^k = (\lambda^k - \lambda^{k+1}) + \beta A(x^k - x^{k+1}). \tag{4.10}$$

Taking (4.10) back into (4.9), it follows that

$$\begin{aligned}&\|w^k - \tilde{w}^k\|_G^2 \\ &= \tau\|x^k - x^{k+1}\|_D^2 - (1-\tau)\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|(\lambda^k - \lambda^{k+1}) + \beta A(x^k - x^{k+1})\|^2 \\ &= \tau\|x^k - x^{k+1}\|_D^2 + \tau\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\ &\quad +2(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1}),\end{aligned}$$

and the proof is complete. ∎

We now turn to deal with the coupled term $2(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})$ in the equality (4.8) and estimate a lower-bound in the quadratic forms. Two various lower-bounds for the cross-multiplication term $(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})$ will be given by the following two lemmas respectively.

LEMMA 4.3. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prediction-correction formulation (3.2)-(3.3) of the proposed indefinite linearized ALM (3.1) for solving (1.1). Then, for the cross-multiplication term in (4.8), we have*

$$\begin{aligned}(\lambda^k - &\lambda^{k+1})^T A(x^k - x^{k+1}) \\ &\geq \Big\{\frac{1}{2}\tau\|x^k - x^{k+1}\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^k - x^{k+1})\|^2\Big\} \\ &\quad - \Big\{\frac{1}{2}\tau\|x^{k-1} - x^k\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\Big\} \\ &\quad - 2(1-\tau)\beta\|A(x^k - x^{k+1})\|^2.\end{aligned} \tag{4.11}$$

Proof. Our first goal is to rewrite the inequality (3.4) as a form which does not contain $\tilde{x}^k$ and $\tilde{\lambda}^k$. It directly follows from (4.10) that

$$\tilde{x}^k = x^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^{k+1} - \beta A(x^k - x^{k+1}).$$

Given the matrix D_0 defined in (2.14), the term $-A^T\tilde{\lambda}^k + \tau r(\tilde{x}^k - x^k)$ in (3.4) then satisfies

$$\begin{aligned}-A^T&\tilde{\lambda}^k + \tau r(\tilde{x}^k - x^k) \\ &= -A^T(\lambda^{k+1} - \beta A(x^k - x^{k+1})) + \tau r(x^{k+1} - x^k) \\ &= -A^T\lambda^{k+1} + D_0(x^{k+1} - x^k).\end{aligned}$$

Therefore, the inequality (3.4) can be rewritten as

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + D_0(x^{k+1} - x^k)\} \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{4.12}$$

Note that (4.12) also holds for $k := k-1$. It follows that

$$\theta(x) - \theta(x^k) + (x - x^k)^T\{-A^T\lambda^k + D_0(x^k - x^{k-1})\} \geq 0, \ \ \forall\, x \in \mathcal{X}. \tag{4.13}$$

Set $x = x^k$ and $x = x^{k+1}$ in (4.12) and (4.13) respectively, and add them, then we have

$$\begin{aligned}&(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})\\ &\qquad \geq \|x^k - x^{k+1}\|^2_{D_0} + (x^k - x^{k+1})^T D_0(x^k - x^{k-1}).\end{aligned} \tag{4.14}$$

For the first part of the right side of (4.14), it follows that

$$\begin{aligned}\|x^k - x^{k+1}\|^2_{D_0} &\overset{(2.14)}{=} \|x^k - x^{k+1}\|^2_{(\tau D - (1-\tau)\beta A^T A)}\\ &= \tau\|x^k - x^{k+1}\|^2_D - (1-\tau)\beta\|A(x^k - x^{k+1})\|^2.\end{aligned} \tag{4.15}$$

For the second part of the right side of (4.14), by the Cauchy-Schwarz inequality, we also have

$$\begin{aligned}&(x^k - x^{k+1})^T D_0(x^k - x^{k-1})\\ &\quad = (x^k - x^{k+1})^T(\tau D - (1-\tau)\beta A^T A)(x^k - x^{k-1})\\ &\quad = \tau(x^k - x^{k+1})^T D(x^k - x^{k-1})\\ &\qquad -(1-\tau)\beta(A(x^k - x^{k+1}))^T A(x^k - x^{k-1})\\ &\quad \geq -\frac{1}{2}\tau\big\{\|x^k - x^{k+1}\|^2_D + \|x^{k-1} - x^k\|^2_D\big\}\\ &\qquad -\frac{1}{2}(1-\tau)\beta\big\{\|A(x^k - x^{k+1})\|^2 + \|A(x^{k-1} - x^k)\|^2\big\}.\end{aligned} \tag{4.16}$$

The assertion (4.11) follows immediately by substituting (4.15) and (4.16) into (4.14). ∎

Lemma 4.4. *Let $\{w^k\}$ be the sequence generated by the scheme (3.2)-(3.3) for solving (1.1). Then, for the cross-multiplication term in (4.8) and any $\tau \in [\frac{3}{4}, 1]$, we have*

$$\begin{aligned}&(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})\\ &\qquad \geq -(\tau - \frac{1}{2})\beta\|A(x^k - x^{k+1})\|^2 - (\frac{5}{2} - 2\tau)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2.\end{aligned} \tag{4.17}$$

Proof. Since $(\tau - \frac{1}{2}) > 0$, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})\\ &\qquad \geq -(\tau - \frac{1}{2})\beta\|A(x^k - x^{k+1})\|^2 - \frac{1}{4(\tau - \frac{1}{2})}\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2.\end{aligned}$$

Thus, it suffices to prove that

$$\frac{1}{4(\tau - \frac{1}{2})} \leq \frac{5}{2} - 2\tau, \quad \forall\, \tau \in [\frac{3}{4}, 1].$$

Let $h(\tau) := (\tau - \frac{1}{2})(\frac{5}{2} - 2\tau)$. It is easy to verify that $h(\tau)$ is a concave function and it reaches its minimum on the interval $[\frac{3}{4}, 1]$ at the endpoint $\tau = \frac{3}{4}$ or $\tau = 1$. Due to $h(\frac{3}{4}) = h(1) = \frac{1}{4}$, we have

$$h(\tau) = (\tau - \frac{1}{2})(\frac{5}{2} - 2\tau) \geq \frac{1}{4}, \quad \forall\, \tau \in [\frac{3}{4}, 1].$$

It immediately implies the assertion of the lemma. ∎

Based on two different lower-bounds (4.11) and (4.17) for the cross-multiplication term $(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})$, it follows that

$$
\begin{aligned}
&2(\lambda^k - \lambda^{k+1})^T A(x^k - x^{k+1})\\
&\quad \geq \big\{\frac{1}{2}\tau\|x^k - x^{k+1}\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^k - x^{k+1})\|^2\big\}\\
&\qquad - \big\{\frac{1}{2}\tau\|x^{k-1} - x^k\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\big\}\\
&\qquad + (\tau - \frac{3}{2})\beta\|A(x^k - x^{k+1})\|^2 - (\frac{5}{2} - 2\tau)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2,
\end{aligned} \tag{4.18}
$$

by adding (4.11) and (4.17). Then, we get the following theorem immediately.

THEOREM 4.5. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prediction-correction formulation (3.2)-(3.3) of the proposed indefinite linearized ALM (3.1) for solving (1.1). Then, for any $\tau \in [\frac{3}{4}, 1]$, we obtain*

$$
\begin{aligned}
&\|w^k - \tilde{w}^k\|_G^2\\
&\quad \geq \frac{1}{2}\big\{\tau\|x^k - x^{k+1}\|_D^2 + (1-\tau)\beta\|A(x^k - x^{k+1})\|^2\big\}\\
&\qquad -\frac{1}{2}\big\{\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\big\}\\
&\qquad + \tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})\big\{\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big\}.
\end{aligned} \tag{4.19}
$$

Proof. To begin with, by substituting (4.18) into (4.8), we obtain

$$
\begin{aligned}
\|w^k - \tilde{w}^k\|_G^2 \geq{}& \tau\|x^k - x^{k+1}\|_D^2 + \tau\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\\
&+ \big\{\frac{1}{2}\tau\|x^k - x^{k+1}\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^k - x^{k+1})\|^2\big\}\\
&- \big\{\frac{1}{2}\tau\|x^{k-1} - x^k\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\big\}\\
&+ (\tau - \frac{3}{2})\beta\|A(x^k - x^{k+1})\|^2 - (\frac{5}{2} - 2\tau)\frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned}
$$

By a further manipulation, we get

$$
\begin{aligned}
&\|w^k - \tilde{w}^k\|_G^2\\
&\quad \geq \frac{1}{2}\big\{\tau\|x^k - x^{k+1}\|_D^2 + (1-\tau)\beta\|A(x^k - x^{k+1})\|^2\big\}\\
&\qquad -\frac{1}{2}\big\{\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\big\}\\
&\qquad + \tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})\big\{\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\big\},
\end{aligned}
$$

and the proof is complete. ∎

It is obvious that the assertion (4.19) corresponds to the desirable inequality (4.6) by defining the non-negative functions $\phi(\cdot,\cdot)$ and $\varphi(\cdot,\cdot)$ as below:

$$
\phi(w^k, w^{k+1}) := \frac{1}{2}\big\{\tau\|x^k - x^{k+1}\|_D^2 + (1-\tau)\beta\|A(x^k - x^{k+1})\|^2\big\}
$$

and

$$\varphi(w^k, w^{k+1}) := \tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})\{\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\}.$$

4.2. Global convergence of (3.1). To show the global convergence of the proposed indefinite linearized ALM (3.1), we first prove the following lemma which paves the way to the convergence proof of (3.1) (see Theorem 4.7).

LEMMA 4.6. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prediction-correction representation (3.2)-(3.3) of the proposed indefinite linearized ALM (3.1) for solving (1.1). Then, for any $\tau \in [\frac{3}{4}, 1]$ and $w^* \in \Omega^*$, we have*

$$\begin{aligned}
&\|w^{k+1} - w^*\|_H^2 + \frac{1}{2}\{\tau\|x^k - x^{k+1}\|_D^2 + (1-\tau)\beta\|A(x^k - x^{k+1})\|^2\} \\
&\quad \le \|w^k - w^*\|_H^2 + \frac{1}{2}\{\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\} \\
&\quad - \{\tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})(\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2)\}.
\end{aligned} \tag{4.20}$$

Proof. To begin with, by substituting (4.19) into (4.1), it follows that

$$\begin{aligned}
&\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(w) \\
&\quad \ge \{\frac{1}{2}\|w - w^{k+1}\|_H^2 + \frac{1}{4}(\tau\|x^k - x^{k+1}\|_D^2 + (1-\tau)\beta\|A(x^k - x^{k+1})\|^2)\} \\
&\qquad - \{\frac{1}{2}\|w - w^k\|_H^2 + \frac{1}{4}(\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2)\} \\
&\qquad + \frac{\tau}{2}\|x^k - x^{k+1}\|_D^2 + (\tau - \frac{3}{4})\{\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\}.
\end{aligned} \tag{4.21}$$

Setting $w = w^*$ in (4.21), we further obtain

$$\begin{aligned}
&\|w^k - w^*\|_H^2 + \frac{1}{2}\{\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\} \\
&- \{\|w^{k+1} - w^*\|_H^2 + \frac{1}{2}(\tau\|x^k - x^{k+1}\|_D^2 + (1-\tau)\beta\|A(x^k - x^{k+1})\|^2)\} \\
&\qquad \ge \tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})\{\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\} \\
&\qquad\quad + 2\{\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(w^*)\}.
\end{aligned}$$

The assertion (4.20) follows immediately from $\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0$. ∎

By Lemma 4.6, it is easy to prove the convergence of the proposed indefinite linearized ALM (3.1) such that

THEOREM 4.7. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prediction-correction formulation (3.2)-(3.3) of the proposed method (3.1) for the problem (1.1). Then, for any $\tau \in (0.75, 1)$, the sequence $\{w^k\}$ converges to some $w^\infty \in \Omega^*$.*

Proof. First of all, summarizing the inequality (4.20) over $k = 1, 2, \ldots, \infty$, it follows that

$$\begin{aligned}
&\sum_{k=1}^{\infty}\{\tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})(\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2)\} \\
&\qquad \le \{\|w^1 - w^*\|_H^2 + \frac{1}{2}\tau\|x^0 - x^1\|_D^2 + \frac{1}{2}(1-\tau)\beta\|A(x^0 - x^1)\|^2\},
\end{aligned}$$

which further implies that

$$\lim_{k\to\infty} \tau\|x^k - x^{k+1}\|_D^2 + 2(\tau - \frac{3}{4})\{\beta\|A(x^k - x^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2\} = 0.$$

Thus, we have $\lim_{k\to\infty} \|x^k - x^{k+1}\|_D^2 = 0$ and $\lim_{k\to\infty} \|\lambda^k - \lambda^{k+1}\|^2 = 0$, which means

$$\lim_{k\to\infty} \|w^k - w^{k+1}\| = 0. \tag{4.22}$$

For arbitrary fixed $w^* \in \Omega^*$ and integer $k \geq 1$, it follows from (4.20) that

$$\begin{aligned} &\|w^{k+1} - w^*\|_H^2 \\ &\quad \leq \|w^k - w^*\|_H^2 + \frac{1}{2}\{\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\} \\ &\quad \leq \cdots \leq \|w^1 - w^*\|_H^2 + \frac{1}{2}\{\tau\|x^0 - x^1\|_D^2 + (1-\tau)\beta\|A(x^0 - x^1)\|^2\}, \end{aligned} \tag{4.23}$$

hence the sequence $\{w^k\}$ is bounded. Due to the matrix M is non-singular, it follows from (3.8) and (4.22) that the sequence $\{\tilde{w}^k\}$ is also bounded. Let w^∞ be some cluster point of $\{\tilde{w}^k\}$, and $\{\tilde{w}^{k_j}\}$ be the associated subsequence converging to w^∞. According to (4.3) and (4.22), the point w^∞ must satisfy

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

and thus $w^\infty \in \Omega^*$.

Therefore, in view of (4.23), we have

$$\|w^{k+1} - w^\infty\|_H^2 \leq \|w^k - w^\infty\|_H^2 + \frac{1}{2}\{\tau\|x^{k-1} - x^k\|_D^2 + (1-\tau)\beta\|A(x^{k-1} - x^k)\|^2\}.$$

In addition, note that w^∞ is also a cluster point of $\{w^{k_j}\}$; by (4.22), it is impossible that the sequence $\{w^k\}$ has more than one cluster point. Therefore, $\{w^k\}$ converges to w^∞ and the proof is complete. ∎

Remark. Refer to the analysis results in, e.g., [18, 20], it is trivial to prove a worst-case $\mathcal{O}(1/N)$ convergence rate measured by the iteration complexity of the proposed method (3.1). We omit its details here for succinctness.

4.3. An extension to indefinite linearized ALM. Now we introduce an extension of to the proposed indefinite linearized ALM (3.1) for the linearly inequality-constrained problem (1.1) when $\rho(A^TA)$ is difficult to estimate in practice. With this respect, we follow the basic indefinite linearized framework (2.15) and present the new ALM-based algorithm as follows:

$$\begin{cases} \tilde{\lambda}^k = [\lambda^k - \beta(Ax^k - b)]_+, & \text{(4.24a)} \\ x^{k+1} = \arg\min\{\theta(x) - (\tilde{\lambda}^k)^T Ax + (\tau+\delta)\frac{\beta}{2}\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\}, & \text{(4.24b)} \\ \lambda^{k+1} = \tilde{\lambda}^k + \beta A(x^k - x^{k+1}), & \text{(4.24c)} \end{cases}$$

in which $\beta > 0$, $\delta > 0$ and $\tau \in (0.75, 1)$.

Compared with the proposed indefinite linearized ALM method (3.1), it is obvious that the scheme (4.24) takes $(\tau+\delta)\beta\|A(x - x^k)\|^2$ instead of $\tau r\|x - x^k\|^2$ for the subproblem of x-approximation, which avoids solving the difficult subproblem (1.5a) directly and has the same computational complexity as the (equality-constrained) ALM (1.4).

We can prove that the introduced new algorithm (4.24) is convergent for any given $\beta > 0$, $\delta > 0$ and $\tau \in (0.75, 1)$. To see this, we first reformulate it in a prediction-correction way. Similar as in subsection 3.1, by setting $\tilde{x}^k = x^{k+1}$, the predictor of (4.24) then reads as

$$\text{(4.25)} \qquad \begin{cases} \tilde{\lambda}^k & = & [\lambda^k - \beta(Ax^k - b)]_+, \\ \tilde{x}^k & = & \arg\min\{\theta(x) - (\tilde{\lambda}^k)^T Ax + (\tau+\delta)\frac{\beta}{2}\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\}, \end{cases}$$

and its corresponding VI-structure satisfies

Prediction step:

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

where

$$\text{(4.26)} \qquad Q = \begin{pmatrix} (\tau+\delta)\beta A^T A & 0 \\ -A & \frac{1}{\beta} I_m \end{pmatrix}.$$

Furthermore, the corrector of (4.24) can be compactly given via

Correction step:

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k),$$

where

$$\text{(4.27)} \qquad M = \begin{pmatrix} I_n & 0 \\ -\beta A & I_m \end{pmatrix}.$$

Under the assumption that A is full column-rank, $\beta > 0$, $\delta > 0$ and $\tau \in (0.75, 1)$, we can easily verify that,

$$H = QM^{-1} = \begin{pmatrix} (\tau+\delta)\beta A^T A & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \succ 0,$$

and

$$G = Q^T + Q - M^T H M = \begin{pmatrix} \underbrace{\delta\beta A^T A - (1-\tau)\beta A^T A}_{D_0} & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}.$$

Clearly, by setting the novel matrices $D := \delta\beta A^T A/\tau$ and $D_0 := [\delta - (1-\tau)]\beta A^T A$, we can follow the same route of convergence analysis in subsection 4.1 and subsection 4.2. Therefore, we can easily show the global convergence and a worst-case $\mathcal{O}(1/N)$ convergence rate for the proposed scheme (4.24) for any given $\beta > 0$, $\delta > 0$ and $\tau \in (0.75, 1)$. A formal description in theory will be further exposed in a future work soon.

5. Numerical experiments. In this section, we evaluate the numerical performance of the proposed method (3.1) by solving the support vector machine for classification and continuous max-flow models for image segmentation, which can be well modelled in terms of convex minimization problems (1.1) with linear inequality constraints. The preliminary numerical results affirmatively illustrate that the proposed algorithm can converge efficiently with a smaller regularization term. Our algorithms were implemented in Python 3.9 and conducted in a Lenovo computer with a 2.20GHz Intel Core i7-8750H CPU and 16GB memory.

5.1. Linear support vector machine (SVM). Suppose the set of training data given by $\mathcal{D} = \big\{(x_1, y_1),\ (x_1, y_1),\ \cdots\ ,(x_m, y_m)\ \big|\ x_i \in \Re^n,\ y_i \in \{-1, 1\}\big\}$ is linearly separable, in which x and y denote the attribute and the label of a sample, respectively. The linear SVM (see, e.g., [10] for more details) is to find the maximum-margin hyperplane separating two classes of data as much as possible, i.e.

$$\begin{aligned} \min_{w\in\Re^n, a\in\Re} \quad & \frac{1}{2}\|w\|^2 \\ \text{s.t.} \quad & y_i(w^T x_i + a) \geq 1, \quad i = 1, \ldots, m. \end{aligned}$$

Obviously, such a model can be rewritten compactly as

$$\min\{\frac{1}{2}\|Hu\|^2 \mid Au \geq b\}, \tag{5.1}$$

by setting

$$u = \begin{pmatrix} w \\ a \end{pmatrix}, \quad H = \begin{pmatrix} I_{n\times n} & 0_{n\times 1} \\ 0_{1\times n} & 0 \end{pmatrix}, \quad A = \begin{pmatrix} y_1({x_1}^T, 1) \\ \vdots \\ y_m({x_m}^T, 1) \end{pmatrix} \text{ and } b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

When the proposed I-IDL-ALM scheme (3.1) is applied to (5.1), the resulting scheme is

$$\text{(I-IDL-ALM)} \begin{cases} \tilde{\lambda}^k &= [\lambda^k - \beta(Au^k - b)]_+, \\ u^{k+1} &= (H^T H + \tau r I_{n+1})^{-1}(A^T\tilde{\lambda}^k + \tau r u^k), \\ \lambda^{k+1} &= \tilde{\lambda}^k + \beta A(u^k - u^{k+1}). \end{cases} \tag{5.2}$$

In the experiments with $\{100, 150, 200, 500, 800\}$ training data respectively, we take $\beta = 0.01$ and $r = \beta\rho(A^T A) + 0.1$. The training data is generated by random numbers satisfying a normal distribution, and we set $n = 2$ for visualizing the classification results. In addition, the stopping criterion is defined by

$$\text{Aer}(k) := \|u^{k+1} - u^k\| < 10^{-6}.$$

TABLE 1
The numerical results of I-IDL-ALM (3.1) with various τ for solving the linear SVM (5.1).

Size	$\tau = 0.75$		$\tau = 0.80$		$\tau = 0.90$		$\tau = 1.00$	
	It.	Aer	It.	Aer	It.	Aer	It.	Aer
100	**7811**	1.00e-8	8235	1.00e-8	9066	1.00e-8	9875	1.00e-8
150	**11601**	1.00e-8	12214	1.00e-8	13414	1.00e-8	14579	1.00e-8
200	**10256**	9.94e-7	10911	9.97e-7	12210	9.97e-7	13508	9.98e-7
500	**14638**	9.93e-7	15588	9.95e-7	17483	9.93e-7	19372	9.94e-7
800	**14711**	8.16e-7	15214	9.12e-7	16179	9.06e-7	17097	7.73e-7

Table 1 shows the experiment results with different settings of total data number and regularization factor τ, which exactly demonstrates that the algorithm (5.2) obtains convergence while $\tau \geq 0.75$. Figure 1 illustrates the computed classification results with 100 and 150 data points with $\tau = 0.75$. As shown in Table 1, the convergence rate gets faster as τ turns smaller and saves up to 25% iterations by comparing the results between $\tau = 1$ and $\tau = 0.75$. We also plot errors w.r.t. iterations ("It.") for the method (5.2) with various settings of τ in Figure 1, from which it is clear that the novel method has a steeper convergence curve for a smaller factor τ.

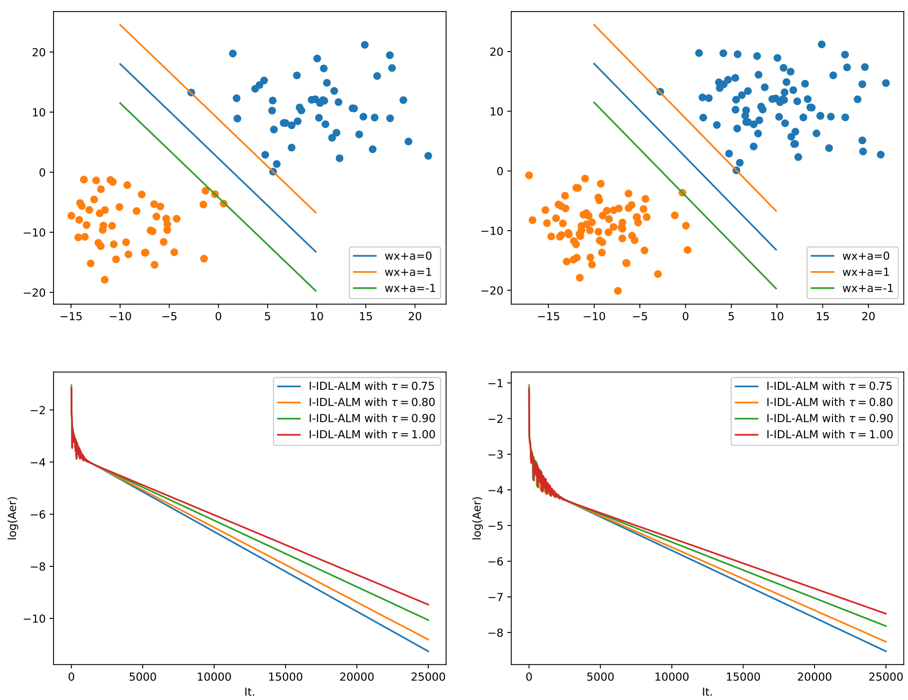

FIG. 1. *The above graphs visualize the numerical performance of the proposed I-IDL-ALM algorithm* (3.1) *for the cases of* 100 *and* 150 *training data respectively, i.e., left and right columns.*

5.2. Potts model-based image segmentation. During recent years, convex optimization was developed as an elegant and efficient optimization tool to image segmentation [7, 8, 24, 35] with guaranteed convergence. In this experiment, we consider the convex-relaxed Potts model for multiphase image segmentation

$$\min_{u_i(x)\geq 0} \Big\{ \sum_{i=1}^{m} \int_{\Omega} u_i(x)\rho(l_i,x) + \alpha|\nabla u_i(x)|dx \;\Big|\; \sum_{i=1}^{m} u_i(x) = 1 \Big\},$$

where $\alpha > 0$ is the weight parameter for the regularization term of the total perimeter of all segmented regions, and $\rho(l_i, x)$ $(i = 1, \dots, m)$ are used to evaluate the cost of assigning the label l_i to the specified position x. In order to avoid directly tackling the complicated pixel-wise simplex constraints on the labeling functions $u_i(x)$, $i = 1, \dots, m$, and the non-smooth total-variation (TV) term (see [31]) in the above convex relaxed Potts model, we take advantage of its equivalent dual optimization model introduced in [34], namely the *continuous max-flow model*:

$$
\begin{aligned}
&\max_{p_s,p,q} \int_\Omega p_s(x)dx \\
&\text{s.t.}\;\; \mathbb{D}iv\; q_i(x) - p_s(x) + p_i(x) = 0,\; i = 1, 2, \ldots, m; \\
&\qquad |q_i(x)| \le \alpha,\; p_i(x) \le \rho(l_i, x),\; i = 1, 2, \ldots, m,
\end{aligned} \tag{5.3}
$$

where $p_s(x)$ and $p_i(x)$ $(i = 1, \ldots, m)$ denote the source flow and the sink flow respectively, and the linear equality constraints are the so-called flow balance conditions of the continuous max-flow model. The optimized labeling functions $u_i(x)$ $(i = 1, \ldots, m)$ by the convex relaxed Potts model are just the multipliers to the linear equality constraints in (5.3). We follow the variational analysis as in [33, 34], where the linear inequality constraints $p_i(x) \le \rho(l_i, x)$ $(i = 1, \ldots, m)$ of (5.3) are nothing but associated with $u_i(x) \ge 0$, then we can omit both the flow variables $p_i(x)$ and their related linear inequalities so as to fix non-negative labeling functions $u_i(x) \ge 0$ directly. This largely saves computational complexities and therefore results in a new abbreviated version of the continuous max-flow model (5.3) as follows:

$$
\begin{aligned}
&\min_{p_s,q} \int_\Omega -p_s(x)dx \\
&\text{s.t.}\;\; \mathbb{D}iv\; q_i(x) - p_s(x) \ge -\rho(l_i, x),\; i = 1, 2, \ldots, m; \\
&\qquad |q_i(x)| \le \alpha,\; i = 1, 2, \ldots, m\,.
\end{aligned} \tag{5.4}
$$

It is easy to observe that the above simplified continuous max-flow model (5.4) can be equally written as (1.1) in that

$$
\begin{aligned}
&\min_{p_s,q} \int_\Omega -p_s(x)dx \,+\, I_C(q(x)) \\
&\text{s.t.}\;\; \mathbb{D}iv\; q_i(x) - p_s(x) \ge -\rho(l_i, x),\; i = 1, 2, \ldots, m,
\end{aligned} \tag{5.5}
$$

where $I_C(q(x))$ is the convex characteristic function of the convex set $C := \{|q_i(x)| \le \alpha,\; i = 1, 2, \ldots, m\}$.

After discretization, the linearly inequality-constrained convex minimization model (5.5) can be essentially written as

$$
\min_{p_s,q} \underbrace{-\mathbf{1}^T p_s \,+\, I_C(q)}_{\theta(p_s,q)}
$$

subject to the following linear inequality constraints

$$
\underbrace{\begin{pmatrix} -I \\ -I \\ \vdots \\ -I \end{pmatrix} p_s(x) + \begin{pmatrix} \mathbb{D}iv \\ 0 \\ \vdots \\ 0 \end{pmatrix} q_1(x) + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \mathbb{D}iv \end{pmatrix} q_m(x)}_{A(p_s;q_1;\ldots;q_m)} \;\ge\; - \underbrace{\begin{pmatrix} \rho(l_1, x) \\ \rho(l_2, x) \\ \vdots \\ \rho(l_m, x) \end{pmatrix}}_{b} .
$$

To this end, the associated matrix A of (5.5) w.r.t. the above linear inequality constraints is

$$
\begin{pmatrix} -I & \mathbb{D}iv & 0 & \cdots & 0 \\ -I & 0 & \mathbb{D}iv & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -I & 0 & 0 & \ldots & \mathbb{D}iv \end{pmatrix}, \text{ and } AA^T = \begin{pmatrix} I-\Delta & I & \cdots & I \\ I & I-\Delta & \cdots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \cdots & I-\Delta \end{pmatrix}.
$$

It is easy to verify that $\rho(A^TA) = \rho(AA^T) \le (8+m)$ for the 2D image, and $\rho(A^TA) \le (12+m)$ for the 3D image in the convex-relaxed Potts model (5.5). For the k-th iteration, given the initial point (p_s^k, q^k, u^k), the proposed linearized indefinite ALM scheme (3.1) explores the following steps:

$$(5.6)\qquad \begin{cases} \tilde{u_i}^k &= \left[u_i^k - \beta(-p_s^k + \rho(l_i) + \mathbb{D}iv\ q_i^k)\right]_+, \quad i = 1, \ldots, m, \\ p_s^{k+1} &= p_s^k + \frac{1}{\tau r}(\mathbf{1} - \sum_{i=1}^n \tilde{u_i}^k), \\ q_i^{k+1} &= \mathcal{P}_C(q_i^k + \frac{\mathbb{D}iv\ ^T\tilde{u_i}^k}{\tau r}), \quad i = 1, \ldots, m, \\ {u_i}^{k+1} &= \tilde{u_i}^k + \beta\{p_s^{k+1} - p_s^k + \mathbb{D}iv\ (q_i^k - q_i^{k+1})\}, \quad i = 1, \ldots, m. \end{cases}$$

Now, we explore the experiments of 2D and 3D image segmentation to show the numerical performance of the new method (3.1) (all image data can be downloaded in the package: https://www.mathworks.com/matlabcentral/fileexchange/34224-fast-continuous-max-flow-algorithm-to-2d-3d-multi-region-image-segmentation). Due to $\rho(A^TA) \le (8+m)$ for the 2D image and $\rho(A^TA) \le (12+m)$ for the 3D image in the convex relaxed Potts model (5.5), we then set $r = 9.1\beta$ and $r = 13.1\beta$ for foreground-background image segmentation in 2D and 3D respectively, i.e., m=2 labels; and $r = 12.1\beta$ and $r = 16.1\beta$ for image segmentation with $m = 4$ labels/phases in 2D and 3D respectively. In the following experiments, the stopping condition is defined via

$$\mathrm{Aer}(k) := \frac{\|\tilde{u}^k - u^k\|}{\mathrm{size}(u^k)} < \mathrm{Tol}.$$

5.2.1. Numerical experiments on foreground-background image segmentation. The foreground-background image segmentation aims to partition the given image into two regions of foreground and background, i.e., two labels.

Table 2

Numerical results of the proposed I-IDL-ALM approach for 2D foreground-background image segmentation, where the TV regularization parameter is $\alpha = 0.5$. The iteration is stopped when Aer(k) <Tol. Its convergence rates w.r.t. different settings of τ are illustrated in the left graph of Figure 4, and the computed segmentation results are shown in Figure 2.

Regularization factor τ	Tol=10^{-6}			Tol=10^{-7}		
	It.	Aer	CPU	It.	Aer	CPU
$\tau = 0.75$	79	8.19e-07	0.82	191	9.88e-08	2.24
$\tau = 0.80$	79	9.17e-07	0.85	195	9.94e-08	2.27
$\tau = 0.90$	80	9.30e-07	0.87	203	9.90e-08	2.33
$\tau = 1.00$	81	9.81e-07	0.93	209	9.96e-08	2.52

Table 2 and Table 3 show the numerical performance of the proposed I-IDL-ALM approach to foreground-background image segmentation in 2D and 3D respectively, and the convergence rates by different settings of τ are well demonstrated in Figure 4. It is easy to see that a smaller value of τ results in fewer iterations to reach convergence. The acceleration effect of the introduced I-IDL-ALM algorithm with the smallest choice $\tau = 0.75$ is clearly shown, which saves about 15% iterations comparing to the case $\tau = 1$. These preliminary numerical results affirmatively verify our theoretical result of the paper, i.e., the relaxation of the regularization term allows a bigger step size to potentially reduce required convergence iterations. Figure 2 and

Figure 3 illustrate the computed segmentation regions in 2D (image size: 238×238) and 3D (image size: $208 \times 112 \times 114$) respectively.

TABLE 3

Numerical results of the proposed I-IDL-ALM approach for 3D foreground-background image segmentation, where the TV regularization parameter is $\alpha = 0.2$. The iteration is stopped when $Aer(k) < Tol$. Its convergence rates w.r.t. different settings of τ are plotted in the right graph of Figure 4, and the segmented results are shown in Figure 3.

Regularization factor τ	Tol=10^{-6}			Tol=10^{-7}		
	It.	Aer	CPU	It.	Aer	CPU
$\tau = 0.75$	44	9.90e-07	50.03	102	9.70e-08	117.06
$\tau = 0.80$	46	9.65e-07	52.06	104	9.87e-08	119.38
$\tau = 0.90$	49	9.95e-07	55.69	110	9.99e-08	126.85
$\tau = 1.00$	53	9.33e-07	60.46	117	9.81e-08	133.55

FIG. 2. *From left to right: the original image and the computed segmentation result by the proposed I-IDL-ALM scheme ($\tau = 0.75$, $\beta = 0.3$ and $r = 9.1\beta$) after 191 iterations in convergence.*

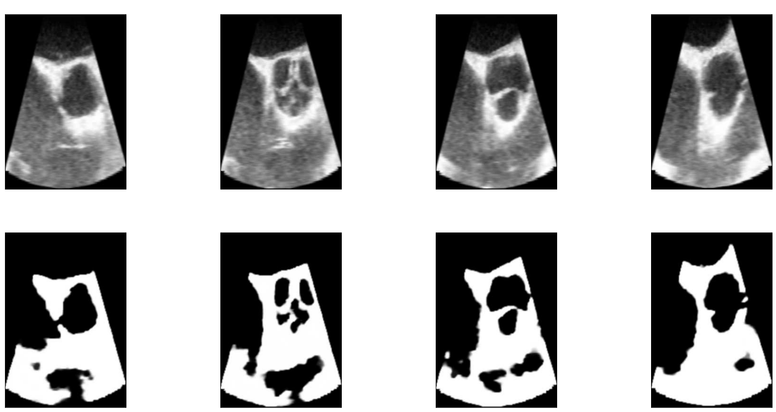

FIG. 3. *First row: the original images (the 40-th, 50-th, 60-th, 70-th slice in sagittal view); second row: the computed segmentation results by the proposed I-IDL-ALM scheme ($\tau = 0.75$, $\beta = 0.9$ and $r = 12.1\beta$) after 102 iterations in convergence.*

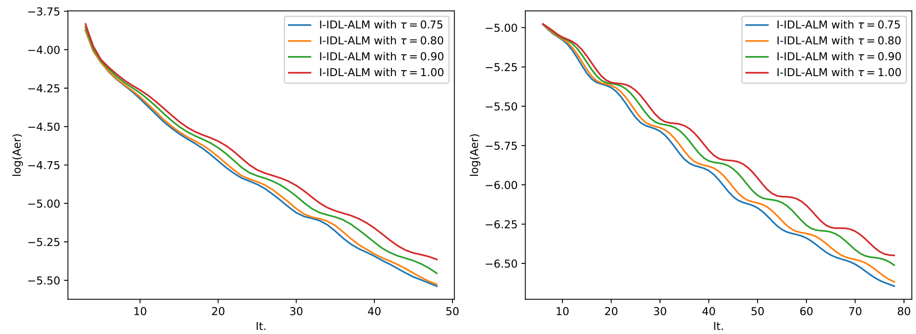

FIG. 4. *Convergence rates of the proposed I-IDL-ALM scheme to foreground-background image segmentation in 2D (left graph) and 3D (right graph), by different settings of τ, are illustrated above respectively.*

5.2.2. Experiments on multiphase image segmentation. Numerical experiments on multiphase image segmentation are also conducted in 2D and 3D cases.

Table 4 and Table 5 show the performance of the proposed I-IDL-ALM approach to multiphase image segmentation in 2D and 3D respectively, and the convergence rates by different settings of τ are clearly demonstrated in Figure 6. It is easy to see that a smaller value of τ results in fewer iterations to reach convergence for both 2D and 3D experiments. The acceleration effect of the introduced I-IDL-ALM scheme with the smallest choice $\tau = 0.75$ is obviously shown, which saves about 15% iterations comparing to the case $\tau = 1$. Figure 5 and Figure 7 illustrate the computed segmentation regions in 2D (image size: 256×256, 4 labels) and 3D (image size: $102 \times 101 \times 100$, 4 labels) respectively.

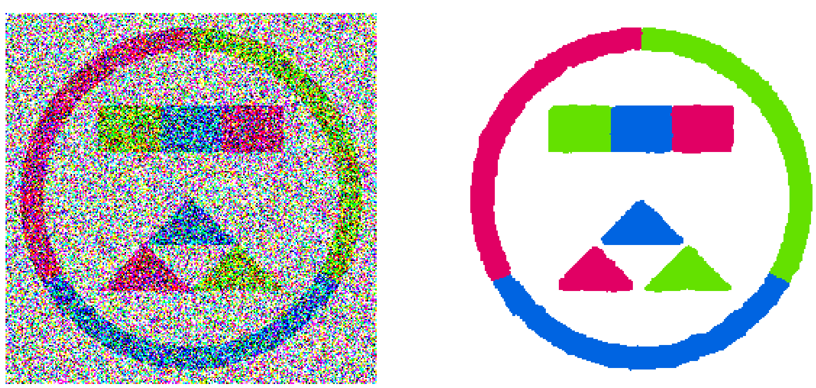

FIG. 5. *From left to right: the original image and the computed segmentation result by the proposed I-IDL-ALM algorithm ($\tau = 0.75$, $\beta = 0.10$ and $r = 12.1\beta$) after 70 iterations in convergence (4 labels, image size: 256×256).*

In summary, our observations from these numerical experiments affirmatively demonstrate that the novel method can converge faster with a smaller regularization term. This numerically confirms the theoretical results presented in this study. Meanwhile, notice that the proposed method (3.1) possesses a parallel splitting structure (parallel computing) for the primal variables. It may bring out certain accelerations when the image data is huge.

TABLE 4

Numerical results of the proposed I-IDL-ALM approach for 2D multiphase image segmentation (4 labels), where the TV regularization parameter is $\alpha = 1$. The algorithm converges when $Aer(k) < Tol$. Its convergence rates w.r.t. different settings of τ are illustrated in the left graph of Figure 6, and the computed segmentation results are shown in Figure 5.

Regularization factor τ	Tol=10^{-5}			Tol=10^{-6}		
	It.	Aer	CPU	It.	Aer	CPU
$\tau = 0.75$	34	9.73e-06	2.16	70	9.93e-07	4.29
$\tau = 0.80$	35	9.50e-06	2.23	73	9.62e-07	4.87
$\tau = 0.90$	36	9.76e-06	2.46	78	9.89e-07	5.19
$\tau = 1.00$	37	9.97e-06	2.64	87	9.71e-07	5.48

TABLE 5

The results listed in the following table demonstrate the numerical performance of the proposed I-IDL-ALM approach for 3D multiphase image segmentation (4 labels), where the TV regularization parameter is $\alpha = 1$. The algorithm converges when $Aer(k) < Tol$. Its convergence rates w.r.t. different settings of τ are illustrated in the right graph of Figure 6, and the computed segmentation results are shown in Figure 7.

Regularization factor τ	Tol=10^{-5}			Tol=10^{-6}		
	It.	Aer	CPU	It.	Aer	CPU
$\tau = 0.75$	34	9.88e-06	63.43	92	9.77e-07	164.34
$\tau = 0.80$	36	9.56e-06	65.63	95	9.73e-07	169.70
$\tau = 0.90$	38	9.90e-06	71.13	101	9.67e-07	179.80
$\tau = 1.00$	41	9.63e-06	75.09	107	9.62e-07	191.36

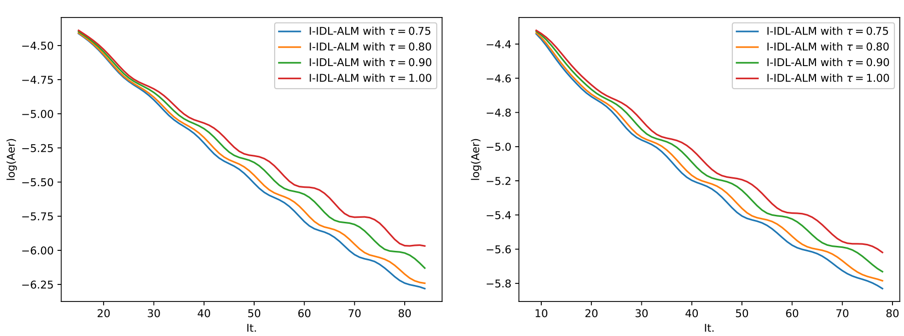

FIG. 6. *Convergence rates of the proposed I-IDL-ALM algorithm to multiphase image segmentation in 2D (left graph) and 3D (right graph), by different settings of τ, are illustrated above respectively.*

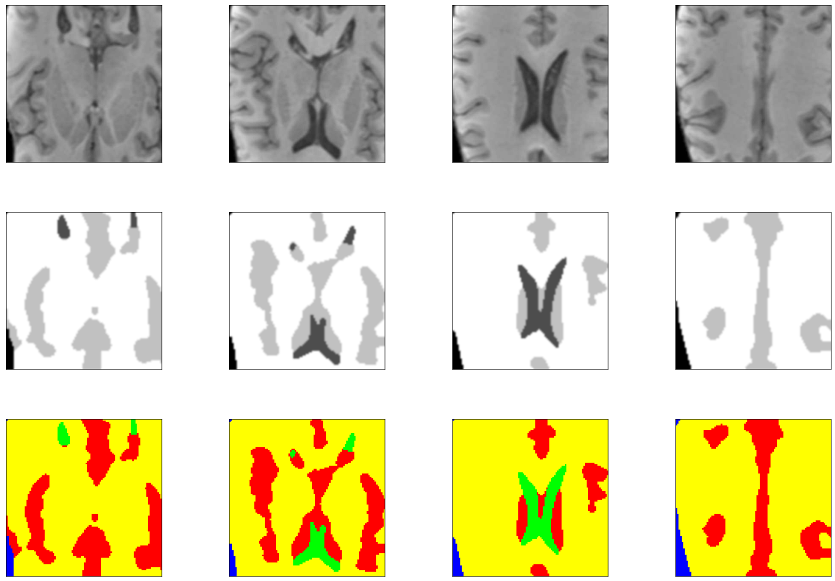

FIG. 7. *First row: the original images (70-th, 80-th, 90-th and 100-th slice in transverse view). Second row: the computed segmentation results by the proposed I-IDL-ALM algorithm ($\tau = 0.75$, $\beta = 0.15$ and $r = 16.1\beta$) after 92 iterations in convergence. Third row: the corresponding colorized regions for better visualization (4 labels, image size: $102 \times 101 \times 100$).*

6. Conclusions and future studies. We propose a novel indefinite linearized ALM scheme for the canonical convex minimization problem with linear inequality constraints, overcoming the calculation difficulty of the core subproblem in original inequality-constrained ALM. To our knowledge, this is the first work to introduce the indefinite linearized ALM scheme framework for efficiently tackling the canonical convex minimization problem with linear inequality constraints, along with an elegant theory of global convergence. Under the new algorithmic framework, the most recent indefinite linearized ALM for the linearly equality-constrained convex optimization problem can be regarded as its special case. The numerical tests on some application problems, including the support vector machine for classification and continuous max-flow models for image segmentation, demonstrate the proposed method can converge faster with a smaller regularization term. This study may enrich and extend our knowledge of the inequality-constrained version of ALM. An extensive research will focus on the case where $\rho(A^T A)$ is difficult to estimate for the proposed indefinite linearized ALM (3.1). We outline its theoretical ideas in subsection 4.3, and will present a formal study in both theory and experiments in a future work soon.

REFERENCES

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, Hoboken, NJ, 2006.

[2] A. Beck, *First-order Methods in Optimization*, SIAM, Philadelphia, 2017.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, Belmont, MA, 1996.

[4] D. P. Bertsekas, *Convex Optimization Algorithms*, Athena Scientific, Nashua, NH, 2015.

[5] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, 2014.
[6] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
[7] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145, https://doi.org/10.1007/s10851-010-0251-1.
[8] T. F. Chan, S. Esedoglu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM Journal on Applied Mathematics, 66 (2006), pp. 1632–1648, https://doi.org/10.1137/040615286.
[9] C. Cortes and V. Vapnik, *Support-vector networks*, Machine learning, 20 (1995), pp. 273–297.
[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2000.
[11] J. Eckstein and W. Yao, *Approximate ADMM algorithms derived from Lagrangian splitting*, Computational Optimization and Applications, 68 (2017), pp. 363–405, https://doi.org/10.1007/s10589-017-9911-z.
[12] M. Fortin and R. Glowinski, *Augmented Lagrangian methods: Applications to the Numerical Solution of Boundary-value Problems*, Elsevier, Stud. Math. Appl. 15, North-Holland, Amsterdam, 1983.
[13] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM, Philadelphia, 1989.
[14] B. He, *My 20 years research on alternating directions method of multipliers*, Oper. Res. Trans., 22 (2018), pp. 1–31.
[15] B. He, L.-Z. Liao, D. Han, and H. Yang, *A new inexact alternating directions method for monotone variational inequalities*, Mathematical Programming, 92 (2002), pp. 103–118, https://doi.org/10.1007/s101070100280.
[16] B. He, H. Liu, Z. Wang, and X. Yuan, *A strictly contractive Peaceman-Rachford splitting method for convex programming*, SIAM Journal on Optimization, 24 (2014), pp. 1011–1040, https://doi.org/10.1137/13090849X.
[17] B. He, F. Ma, and X. Yuan, *Convergence study on the symmetric version of ADMM with larger step sizes*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1467–1501, https://doi.org/10.1137/15M1044448.
[18] B. He, F. Ma, and X. Yuan, *Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems*, IMA Journal of Numerical Analysis, 40 (2020), pp. 1188–1216, https://doi.org/10.1093/imanum/dry092.
[19] B. He, F. Ma, and X. Yuan, *Optimally linearizing the alternating direction method of multipliers for convex programming*, Computational Optimization and Applications, 75 (2020), pp. 361–388, https://doi.org/10.1007/s10589-019-00152-3.
[20] B. He and X. Yuan, *On the O(1/n) convergence rate of the Douglas-Rachford alternating direction method*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 700–709, https://doi.org/10.1137/110836936.
[21] B. He and X. Yuan, *A class of ADMM-based algorithms for three-block separable convex programming*, Computational Optimization and Applications, 70 (2018), pp. 791–826, https://doi.org/10.1007/s10589-018-9994-1.
[22] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 4 (1969), pp. 303–320, https://doi.org/10.1007/BF00927673.
[23] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM, 2008.
[24] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr, *Convex multi-class image labeling by simplex-constrained total variation*, in International conference on scale space and variational methods in computer vision, Springer, 2009, pp. 150–162.
[25] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, vol. 28, Addison-wesley Reading, MA, 1973.
[26] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Fr. Inform. Rech. Oper., 4 (1970), pp. 154–158.
[27] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in optimization, 1 (2014), pp. 127–239, https://doi.org/10.1561/2400000003.
[28] M. J. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.
[29] R. T. Rockafellar, *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research, 1 (1976), pp. 97–116, https:

//doi.org/10.1287/moor.1.2.97.

[30] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp. 877–898, https://doi.org/10.1137/0314056.

[31] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268, https://doi.org/10.1016/0167-2789(92)90242-F.

[32] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*, Mit Press, Cambridge, MA, 2011.

[33] J. Yuan, E. Bae, and X.-C. Tai, *A study on continuous max-flow and min-cut approaches*, in Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2217–2224.

[34] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, *A continuous max-flow approach to Potts model*, in European Conference on Computer Vision, Springer, 2010, pp. 379–392.

[35] J. Yuan, E. Bae, X.-C. Tai, and Y. Boykov, *A spatially continuous max-flow and min-cut framework for binary labeling problems*, Numerische Mathematik, 126 (2014), pp. 559–587, https://doi.org/10.1007/s00211-013-0569-x.

[36] J. Yuan and A. Fenster, *Modern convex optimization to medical image analysis*, arXiv preprint arXiv:1809.08734, (2018).

[37] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, Journal of Scientific Computing, 46 (2011), pp. 20–46, https://doi.org/10.1007/s10915-010-9408-8.