A family of accelerated inexact augmented Lagrangian methods with applications to image restoration*

Jianchao Bai † Yuxue Ma ‡

May 23, 2022

Abstract. In this paper, we focus on a class of convex optimization problems subject to equality or inequality constraints and have developed an Accelerated Inexact Augmented Lagrangian Method (AI-ALM). Different relative error criteria are designed to solve the subproblem of AI-ALM inexactly, and the popular used relaxation step is exploited to accelerate the convergence. By a unified variational analysis, we establish the global convergence of AI-ALM and its sublinear convergence rate in terms of the primal iterative residual, the objective function value gap and constraint violation, respectively. Numerical experiments on testing the image restoration problem with different types of images indicate that AI-ALM is effective and promising. In appendix, we also extend AI-ALM to solve a general multi-block problem and briefly discuss convergence of the extended method.

Key words: convex optimization, inexact augmented Lagrangian method, relaxation step, relative error criteria, convergence complexity, image restoration

Mathematics Subject Classification(2010): 65K10; 65Y20; 90C25

1 Introduction

We are interested in the following linearly constrained optimization problem

$$\min_{x\in\mathcal{R}^n} \Big\{\theta(\mathbf{x})|\ A\mathbf{x}=b\ (\text{or } A\mathbf{x}\geq b)\Big\}, \tag{1}$$

where $\theta(\mathbf{x}):\mathcal{R}^n\to\mathcal{R}$ is a proper closed convex function (but not necessarily strongly convex or smooth); $A\in\mathcal{R}^{m\times n}, b\in\mathcal{R}^m$ are given. Here and hereafter, the symbols $\mathcal{R},\mathcal{R}^m,\mathcal{R}^{m\times n}$ denote the sets of real numbers, m dimensional real column vectors, and $m\times n$ real matrices, respectively. The notation $Q\succ\mathbf{0}$ means Q is symmetric positive definite matrix, where $\mathbf{0}$ stands for a zero matrix/vector with proper dimensions. We also use $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ to denote the standard Euclidean norm and inner product, respectively. Throughout the context, the solution set of the problem (1) is assumed to be nonempty.

Motivated by the recent interesting work [4] which is a novel Augmented Lagrangian Method (ALM), we aims at developing an Accelerated Inexact ALM (AI-ALM, see Algorithm 1) for

*This work was supported by the National Natural Science Foundation of China (No. 12001430) and the China Postdoctoral Science Foundation (No. 2020M683545).

†`https://bjc1987.github.io`, School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China (`bjc1987@163.com`).

‡School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China (`Mayuxue708991@163.com`).

solving the problem (1), where the core subproblem is solved inexactly to satisfy one of the following relative error criteria:

$$2\left|\langle v^k - \widetilde{\mathbf{x}}^k, d^k\rangle\right| + \left\|d^k\right\|^2 \le (2-\gamma)\sigma\left\|\widetilde{\mathbf{x}}^k - \mathbf{x}^k\right\|_Q^2; \tag{C1}$$

$$2\left|\langle v^k - \widetilde{\mathbf{x}}^k, d^k\rangle\right| + \left\|d^k\right\|^2 \le (2-\gamma)\sigma\left\|\widetilde{\mathbf{x}}^{k-1} - \mathbf{x}^{k-1}\right\|_Q^2; \tag{C2}$$

$$2\left|\langle v^k - \widetilde{\mathbf{x}}^k, d^k\rangle\right| + \left\|d^k\right\|^2 \le \frac{(2-\gamma)\sigma}{2\beta\gamma^2}\left\|\lambda^k - \lambda^{k-1}\right\|^2. \tag{C3}$$

The reason why we design an inexact method is that most of algorithmic subproblems perhaps admit no closed-form solution, and it is not necessary to solve the involved subproblem exactly, especially in inverse imaging [5, 9, 11]. For more inexact absolute or relative error criteria, we refer to e.g. [6, 12, 13, 14, 17, 18] and the references therein.

Algorithm 1 [AI-ALM for solving Problem (1)]

---

1 **Initialize** $(\mathbf{x}^0, v^0, \lambda^0) \in \mathcal{R}^n \times \mathcal{R}^n \times \mathcal{R}^m$ and set $\beta > 0, Q \succ \mathbf{0}, \sigma \in [0,1), \gamma \in (0,2)$;

2 **While** stopping criteria is not satisfied **do**

3 Compute $\widetilde{\mathbf{x}}^k \approx \arg\min\limits_{\mathbf{x}\in\mathcal{R}^n} \left\{\theta(\mathbf{x}) - \langle \lambda^k, A\mathbf{x} - b\rangle + \beta\left\|A(\mathbf{x} - \mathbf{x}^k)\right\|^2 + \frac{1}{2}\left\|\mathbf{x} - \mathbf{x}^k\right\|_Q^2\right\}$ such that (C1), (C2) or (C3) holds with

$$d^k \in \partial\theta(\widetilde{\mathbf{x}}^k) - A^\mathsf{T}\lambda^k + \left(2\beta A^\mathsf{T}A + Q\right)\left(\widetilde{\mathbf{x}}^k - \mathbf{x}^k\right);$$

4 Update $\widetilde{\lambda}^k = \lambda^k - \beta\left[A(2\widetilde{\mathbf{x}}^k - \mathbf{x}^k) - b\right]$;

5 Update the auxiliary variable $v^{k+1} = v^k - d^k$;

6 Relaxation step: $\begin{pmatrix} \mathbf{x}^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}^k \\ \lambda^k \end{pmatrix} + \gamma \begin{pmatrix} \widetilde{\mathbf{x}}^k - \mathbf{x}^k \\ \widetilde{\lambda}^k - \lambda^k \end{pmatrix}$.

7 **End while**

Note that the objective function of the $\widetilde{\mathbf{x}}^k$-subproblem in Algorithm 1 differs from most of exact/inexact augmented Lagrangian methods, see e.g. [3, 7, 10, 11] to list a few. For example, the equivalent core subproblem of the standard ALM [7] is

$$\min_{\mathbf{x}\in\mathcal{R}^n} \left\{\theta(\mathbf{x}) - \left\langle \mathbf{x}, A^\mathsf{T}\lambda^k + \beta A^\mathsf{T}b\right\rangle + \frac{1}{2}\left\|\mathbf{x}\right\|_{A^\mathsf{T}A}^2\right\},$$

while the subproblem of Algorithm 1 amounts to

$$\min_{\mathbf{x}\in\mathcal{R}^n} \left\{\theta(\mathbf{x}) - \left\langle \mathbf{x}, A^\mathsf{T}\lambda^k + (Q + 2\beta A^\mathsf{T}A)\mathbf{x}^k\right\rangle + \frac{1}{2}\left\|\mathbf{x}\right\|_{Q+2\beta A^\mathsf{T}A}^2\right\}.$$

Obviously, the above subproblem does not depend on the known data b but the previous iteration $\mathbf{x}^k$, which means *our proposed algorithm is a novel inexact ALM*. Particularly, by taking $Q = \tau\mathbf{I} - 2\beta A^\mathsf{T}A$ with $\tau > 2\beta\|A^\mathsf{T}A\|$, the last subproblem will reduce to

$$\min_{\mathbf{x}\in\mathcal{R}^n} \left\{\theta(\mathbf{x}) + \frac{\tau}{2}\left\|\mathbf{x} - (A^\mathsf{T}\lambda^k + \tau\mathbf{x}^k)/\tau\right\|^2\right\}$$

and it is the proximity operator of $\theta(\mathbf{x})$ to be solved easier than the original. The bold symbol $\mathbf{I}$ used here and hereafter denotes an identity matrix with proper dimension. For the problem (1) subject to $A\mathbf{x} \ge b$, the dual update will read $\widetilde{\lambda}^k = \max\left\{\lambda^k - \beta\left[A(2\widetilde{\mathbf{x}}^k - \mathbf{x}^k) - b\right], \mathbf{0}\right\}$.

We have the following comments on the aforementioned criteria. By the steps 4 and 6 in Algorithm 1, the term in the right-hand-side of (C3) will become $\frac{(2-\gamma)\beta}{2}\sigma\left\|A(\widetilde{\mathbf{x}}^{k-1} - \mathbf{x}^{k-1}) +\right.$

$A\widetilde{\mathbf{x}}^{k-1} - b\|^2$. An obvious difference among these criteria is that the strategy (C1) employs a variable error since the term in the right-hand-side uses the current primal iterative residual, while both (C2) and (C3) enjoy an invariant error determined by the previous primal/dual iterative residual. If there exists an exact solution for the $\mathbf{x}$-subproblem, then we can take $d^k = \mathbf{0}$ such that (C1), (C2) or (C3) holds. If $\sigma = 0$, then the auxiliary variable v is invariant and hence Algorithm 1 reduces to an exact version of the previous P-rALM [4].

The paper is organized as follows. In Section 2, we provide a variational characterization for the primal-dual solution of the problem (1) and recall two basic lemmas for the sake of convergence analysis. In Section 3, we analyze the global convergence of the proposed algorithm and its sublinear convergence rate in a unified framework although different relative error criteria are involved. Section 4 investigates the numerical performance of the proposed algorithm for solving the image restoration problem with different kinds of images. We conclude the paper in Section 5 and finally discuss the convergence of an extended algorithms for solving the multi-block convex optimization problems.

2 Preliminary

We first characterize the solution of (1) as a variational inequality. By attaching the Lagrangian multiplier λ to the linear constraints, the Lagrangian function of (1) is given by

$$L(\mathbf{x}, \lambda) = \theta(\mathbf{x}) - \langle \lambda, A\mathbf{x} - b \rangle.$$

From the perspective of convex optimization, a point

$$\mathbf{w}^* = (\mathbf{x}^*; \lambda^*) \in \mathcal{M} := \mathcal{R}^n \times \Lambda, \quad \text{where} \quad \Lambda := \begin{cases} \mathcal{R}^m, & \text{if } A\mathbf{x} = b, \\ \mathcal{R}^m_+, & \text{if } A\mathbf{x} \geq b, \end{cases}$$

is called the primal-dual solution of (1) if and only if it is the saddle-point of $L(\mathbf{x}, \lambda)$:

$$L_{\lambda \in \Lambda}(\mathbf{x}^*, \lambda) \leq L(\mathbf{x}^*, \lambda^*) \leq L_{\mathbf{x} \in \mathcal{R}^n}(\mathbf{x}, \lambda^*).$$

These inequalities can be expressed as the following variational inequality form

$$\theta(\mathbf{x}) - \theta(\mathbf{x}^*) + \big\langle \mathbf{w} - \mathbf{w}^*, \mathcal{J}(\mathbf{w}^*) \big\rangle \geq 0, \ \forall \mathbf{w} \in \mathcal{M}.$$

where

$$\mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \lambda \end{pmatrix} \quad \text{and} \quad \mathcal{J}(\mathbf{w}) = \begin{pmatrix} -A^{\mathsf{T}}\lambda \\ A\mathbf{x} - b \end{pmatrix}.$$

Clearly, an equivalent expression of the above variational inequality reads

$$\mathrm{VI}(\theta, \mathcal{J}, \mathcal{M}): \quad \theta(\mathbf{x}) - \theta(\mathbf{x}^*) + \big\langle \mathbf{w} - \mathbf{w}^*, \mathcal{J}(\mathbf{w}) \big\rangle \geq 0, \ \forall \mathbf{w} \in \mathcal{M}, \tag{2}$$

because the affine mapping $\mathcal{J}(\mathbf{w})$ is skew symmetric and satisfies

$$\big\langle \mathbf{w} - \bar{\mathbf{w}}, \mathcal{J}(\mathbf{w}) - \mathcal{J}(\bar{\mathbf{w}}) \big\rangle = 0, \ \forall \mathbf{w}, \bar{\mathbf{w}} \in \mathcal{M}. \tag{3}$$

Notice that the solution set of $\mathrm{VI}(\theta, \mathcal{J}, \mathcal{M})$, denoted by $\mathcal{M}^*$, is nonempty by the previous assumption. Moreover, it is convex and can be expressed as

$$\mathcal{M}^* = \bigcap_{\mathbf{w} \in \mathcal{M}} \Big\{ \hat{\mathbf{w}} \mid \theta(\mathbf{x}) - \theta(\hat{\mathbf{x}}) + \big\langle \mathbf{w} - \hat{\mathbf{w}}, \mathcal{J}(\mathbf{w}) \big\rangle \geq 0 \Big\},$$

see e.g. [8, Theorem 2.1]. A straightforward conjecture is that convergence properties of our AI-ALM can be established if a similar inequality to (2) can be obtained with some additional residuals at the previous and current iterations and some extra terms converging to zero.

The forthcoming preliminary lemmas are used to prove the convergence rate of our proposed algorithm, which can be found in e.g. [19].

Lemma 2.1 *Given a function ϕ and a fixed point $\bar{x}$, if for any λ it holds that*

$$F(\bar{x}) - F(x^*) - \langle \lambda, A\bar{x} - b\rangle \le \phi(\lambda),$$

then for any $\rho > 0$ we have

$$F(\bar{x}) - F(x^*) + \rho\|A\bar{x} - b\| \le \sup_{\|\lambda\|\le\rho} \phi(\lambda).$$

Lemma 2.2 *For any $\epsilon \ge 0$, if*

$$F(\bar{x}) - F(x^*) + \rho\|A\bar{x} - b\| \le \epsilon,$$

then we have

$$\|A\bar{x} - b\| \le \frac{\epsilon}{\rho - \|\lambda^*\|} \quad \text{and} \quad -\frac{\|\lambda^*\|\epsilon}{\rho - \|\lambda^*\|} \le F(\bar{x}) - F(x^*) \le \epsilon,$$

where (x^*, λ^*) is a saddle point of (1), and we assume $\|\lambda^*\| \le \rho$.

3 Convergence analysis of AI-ALM

3.1 Variational characterization for the iterates

In this subsection, the iterates generated by AI-ALM will be characterized as a mixed variational inequality with the aid of the notation $\widetilde{\mathbf{w}}^k = (\widetilde{\mathbf{x}}^k; \widetilde{\lambda}^k)$ and the H-weighted norm defined as $\|\mathbf{w}\|_H = \sqrt{\langle \mathbf{w}, H\mathbf{w}\rangle}$, where H is a symmetric positive definite matrix.

Lemma 3.1 *The iterates generated by Algorithm 1 satisfy*

$$\widetilde{\mathbf{w}}^k \in \mathcal{M},\ \theta(\mathbf{x}) - \theta(\widetilde{\mathbf{x}}^k) + \left\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w})\right\rangle - \left\langle \mathbf{x} - \widetilde{\mathbf{x}}^k, d^k\right\rangle \ge \left\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, H(\mathbf{w}^k - \widetilde{\mathbf{w}}^k)\right\rangle \tag{4}$$

for any $\mathbf{w} \in \mathcal{M}$, where

$$\widetilde{\mathbf{w}}^k = \begin{pmatrix} \widetilde{\mathbf{x}}^k \\ \widetilde{\lambda}^k \end{pmatrix} \quad \text{and} \quad H = \begin{bmatrix} 2\beta A^\mathsf{T} A + Q & A^\mathsf{T} \\ A & \frac{1}{\beta}\mathbf{I} \end{bmatrix} \tag{5}$$

is symmetric positive definite for any $\beta > 0$ and $Q \succ -\beta A^\mathsf{T} A$. Moreover, we have

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\widetilde{\mathbf{x}}^k) + \left\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, \mathcal{J}(\mathbf{w})\right\rangle - \left\langle \mathbf{x} - \widetilde{\mathbf{x}}^k, d^k\right\rangle \\ \ge\ & \frac{1}{2\gamma}\Big\{\|\mathbf{w} - \mathbf{w}^{k+1}\|_H^2 - \|\mathbf{w} - \mathbf{w}^k\|_H^2 + \gamma(2-\gamma)\|\widetilde{\mathbf{w}}^k - \mathbf{w}^k\|_H^2\Big\}. \end{aligned} \tag{6}$$

Proof. For any $\beta > 0$ and $Q \succ -\beta A^\mathsf{T} A$, it holds

$$H \succ \begin{bmatrix} \beta A^\mathsf{T} A & A^\mathsf{T} \\ A & \frac{1}{\beta}\mathbf{I} \end{bmatrix} = \begin{pmatrix} \sqrt{\beta} A^\mathsf{T} \\ \frac{1}{\sqrt{\beta}}\mathbf{I} \end{pmatrix} \left(\sqrt{\beta} A, \frac{1}{\sqrt{\beta}}\mathbf{I}\right) \succeq \mathbf{0}.$$

Hence, H is a positive definite matrix.

By the third step in Algorithm 1, we have

$$\theta(\mathbf{x}) - \theta(\widetilde{\mathbf{x}}^k) + \left\langle \mathbf{x} - \widetilde{\mathbf{x}}^k, -A^\mathsf{T}\lambda^k + (2\beta A^\mathsf{T} A + Q)(\widetilde{\mathbf{x}}^k - \mathbf{x}^k) - d^k\right\rangle \ge 0,\ \forall \mathbf{x} \in \mathcal{R}^n,$$

or equivalently,

$$\begin{aligned} & \theta(\mathbf{x}) - \theta(\widetilde{\mathbf{x}}^k) + \left\langle \mathbf{x} - \widetilde{\mathbf{x}}^k, -A^\mathsf{T}\widetilde{\lambda}^k\right\rangle - \left\langle \mathbf{x} - \widetilde{\mathbf{x}}^k, d^k\right\rangle \\ \ge\ & \left\langle \mathbf{x} - \widetilde{\mathbf{x}}^k, (2\beta A^\mathsf{T} A + Q)(\mathbf{x}^k - \widetilde{\mathbf{x}}^k) + A^\mathsf{T}(\lambda^k - \widetilde{\lambda}^k)\right\rangle. \end{aligned}$$

The update of $\widetilde{\lambda}^k$ implies

$$\widetilde{\lambda}^k \in \Lambda, \quad \langle \lambda - \widetilde{\lambda}^k, A\widetilde{\mathbf{x}}^k - b \rangle = \Big\langle \lambda - \widetilde{\lambda}^k, A(\mathbf{x}^k - \widetilde{\mathbf{x}}^k) + \frac{1}{\beta}(\lambda^k - \widetilde{\lambda}^k) \Big\rangle, \quad \forall \lambda \in \Lambda.$$

Combining the last inequality and equality together with the structure of H given by (5) and the property in (3), the desired inequality (4) is confirmed.

The sixth step in Algorithm 1 shows

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \gamma(\widetilde{\mathbf{w}}^k - \mathbf{w}^k), \tag{7}$$

which makes the term in the right-hand-side of (4) become

$$\begin{aligned}
&\big\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, H(\mathbf{w}^k - \widetilde{\mathbf{w}}^k) \big\rangle = \frac{1}{\gamma} \langle \widetilde{\mathbf{w}}^k - \mathbf{w}, H(\mathbf{w}^{k+1} - \mathbf{w}^k) \rangle \\
= \; & \frac{1}{2\gamma} \left\{ \left\|\widetilde{\mathbf{w}}^k - \mathbf{w}^k\right\|_H^2 - \left\|\widetilde{\mathbf{w}}^k - \mathbf{w}^{k+1}\right\|_H^2 + \left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 \right\} \\
= \; & \frac{1}{2\gamma} \left\{ \gamma(2-\gamma)\left\|\widetilde{\mathbf{w}}^k - \mathbf{w}^k\right\|_H^2 + \left\|\mathbf{w} - \mathbf{w}^{k+1}\right\|_H^2 - \left\|\mathbf{w} - \mathbf{w}^k\right\|_H^2 \right\}.
\end{aligned} \tag{8}$$

Here, the third equality of (8) uses (7) and the second uses the identity

$$\big\langle \mathbf{p} - \mathbf{q}, H(\mathbf{u} - \mathbf{v}) \big\rangle = \frac{1}{2} \Big\{ \left\|\mathbf{p} - \mathbf{v}\right\|_H^2 - \left\|\mathbf{p} - \mathbf{u}\right\|_H^2 + \left\|\mathbf{q} - \mathbf{u}\right\|_H^2 - \left\|\mathbf{q} - \mathbf{v}\right\|_H^2 \Big\} \tag{9}$$

with specifications

$$\mathbf{p} := \widetilde{\mathbf{w}}^k, \;\; \mathbf{q} = \mathbf{w}, \;\; \mathbf{u} = \mathbf{w}^{k+1} \text{ and } \mathbf{v} := \mathbf{w}^k.$$

Then, plug (8) into (4) to obtain the inequality (6). ■

To obtain a more tight result (as shown in the following Corollary 3.1) from Lemma 3.1 and for the sake of simplicity, let us now denote

$$\begin{cases} \Phi_k(\mathbf{w}) = \dfrac{1}{\gamma}\|\mathbf{w} - \mathbf{w}^k\|_H^2 + \left\|\mathbf{x} - \upsilon^k\right\|^2, \\ \varphi(\sigma, \gamma) = \min\{1, 1-\sigma\}\dfrac{2-\gamma}{\gamma^2}, \end{cases} \tag{10}$$

and

$$\Psi_k = \begin{cases} 0, & \text{if (C1) is satisfied,} \\ \sigma\frac{2-\gamma}{\gamma^2}\left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|_Q^2, & \text{if (C2) is satisfied,} \\ \sigma\frac{2-\gamma}{2\beta\gamma^2}\left\|\lambda^k - \lambda^{k-1}\right\|^2, & \text{if (C3) is satisfied.} \end{cases} \tag{11}$$

Corollary 3.1 *Let Φ_k and Ψ_k be defined in (10) and (11), respectively. For any $\beta > 0, \gamma \in (0,2), \sigma \in [0,1)$, the iterates generated by Algorithm 1 satisfy*

$$\begin{aligned}
&2\Big\{\theta(\widetilde{\mathbf{x}}^k) - \theta(\mathbf{x}) + \big\langle \widetilde{\mathbf{w}}^k - \mathbf{w}, \mathcal{J}(\mathbf{w}) \big\rangle \Big\} + \Phi_{k+1}(\mathbf{w}) + \Psi_{k+1} \\
\le & \Phi_k(\mathbf{w}) + \Psi_k - \varphi(\sigma,\gamma)\Big\{ \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_Q^2 + \frac{1}{2\beta}\|\lambda^{k+1} - \lambda^k\|^2 \Big\}.
\end{aligned} \tag{12}$$

Moreover, we have

$$\Phi_{k+1}(\mathbf{w}^*) + \Psi_{k+1} \le \Phi_k(\mathbf{w}^*) + \Psi_k - \varphi(\sigma,\gamma)\Big\{ \left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|_Q^2 + \frac{1}{2\beta}\|\lambda^{k+1} - \lambda^k\|^2 \Big\}. \tag{13}$$

Proof. It follows from (6), the fifth step of Algorithm 1 and (9) that

$$
\begin{aligned}
&2\Big\{\theta(\mathbf{x})-\theta(\widetilde{\mathbf{x}}^k)+\big\langle \mathbf{w}-\widetilde{\mathbf{w}}^k,\mathcal{J}(\mathbf{w})\big\rangle\Big\}-2\big\langle v^k-\widetilde{\mathbf{x}}^k,d^k\big\rangle-(2-\gamma)\|\widetilde{\mathbf{w}}^k-\mathbf{w}^k\|_H^2+\|d^k\|^2\\
\geq&\frac{1}{\gamma}\left\{\|\mathbf{w}-\mathbf{w}^{k+1}\|_H^2-\|\mathbf{w}-\mathbf{w}^k\|_H^2\right\}-2\big\langle v^k-\mathbf{x},d^k\big\rangle+\|d^k\|^2\\
=&\frac{1}{\gamma}\left\{\|\mathbf{w}-\mathbf{w}^{k+1}\|_H^2-\|\mathbf{w}-\mathbf{w}^k\|_H^2\right\}+2\big\langle \mathbf{x}-v^k,v^k-v^{k+1}\big\rangle+\|v^k-v^{k+1}\|^2\\
=&\frac{1}{\gamma}\left\{\|\mathbf{w}-\mathbf{w}^{k+1}\|_H^2-\|\mathbf{w}-\mathbf{w}^k\|_H^2\right\}+\left\|\mathbf{x}-v^{k+1}\right\|^2-\left\|\mathbf{x}-v^k\right\|^2,
\end{aligned}
$$

in other words,

$$
\begin{aligned}
&2\Big\{\theta(\widetilde{\mathbf{x}}^k)-\theta(\mathbf{x})+\big\langle \widetilde{\mathbf{w}}^k-\mathbf{w},\mathcal{J}(\mathbf{w})\big\rangle\Big\} \qquad\qquad (14)\\
\leq\;&\Big\{\frac{1}{\gamma}\|\mathbf{w}-\mathbf{w}^k\|_H^2+\left\|\mathbf{x}-v^k\right\|^2\Big\}-\Big\{\frac{1}{\gamma}\|\mathbf{w}-\mathbf{w}^{k+1}\|_H^2+\left\|\mathbf{x}-v^{k+1}\right\|^2\Big\}\\
&-2\big\langle v^k-\widetilde{\mathbf{x}}^k,d^k\big\rangle-(2-\gamma)\|\widetilde{\mathbf{w}}^k-\mathbf{w}^k\|_H^2+\|d^k\|^2\\
=\;&\Phi_k(\mathbf{w})-\Phi_{k+1}(\mathbf{w})-2\big\langle v^k-\widetilde{\mathbf{x}}^k,d^k\big\rangle+\|d^k\|^2\\
&-(2-\gamma)\Big\{\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\|_{2\beta A^\intercal A+Q}^2+2\big\langle A(\widetilde{\mathbf{x}}^k-\mathbf{x}^k),\widetilde{\lambda}^k-\lambda^k\big\rangle+\frac{1}{\beta}\|\widetilde{\lambda}^k-\lambda^k\|^2\Big\}\\
\leq\;&\Phi_k(\mathbf{w})-\Phi_{k+1}(\mathbf{w})-2\big\langle v^k-\widetilde{\mathbf{x}}^k,d^k\big\rangle+\|d^k\|^2\\
&-(2-\gamma)\Big\{\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\|_Q^2+\frac{1}{2\beta}\|\widetilde{\lambda}^k-\lambda^k\|^2\Big\},
\end{aligned}
$$

where the equality uses the notation of $\Phi_k(\mathbf{w})$ given by (10) and the structure of the matrix H, the final inequality uses the following Cauchy-Schwartz inequality

$$
-2\big\langle A(\widetilde{\mathbf{x}}^k-\mathbf{x}^k),\widetilde{\lambda}^k-\lambda^k\big\rangle\leq 2\beta\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\|_{A^\intercal A}^2+\frac{1}{2\beta}\|\widetilde{\lambda}^k-\lambda^k\|^2.
$$

Next, we estimate the upper bound of the term in the right-hand-side of the final inequality in (14), denoted by RHS, under the aforementioned relative error criteria.

- For Algorithm 1 with (C1), we have

$$
\begin{aligned}
\text{RHS}\leq&\Phi_k(\mathbf{w})-\Phi_{k+1}(\mathbf{w})+\sigma(2-\gamma)\left\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\right\|_Q^2\\
&-(2-\gamma)\Big\{\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\|_Q^2+\frac{1}{2\beta}\|\widetilde{\lambda}^k-\lambda^k\|^2\Big\}\\
=&\Phi_k(\mathbf{w})-\Phi_{k+1}(\mathbf{w})-(1-\sigma)(2-\gamma)\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\|_Q^2-\frac{2-\gamma}{2\beta}\|\widetilde{\lambda}^k-\lambda^k\|^2\\
=&\Phi_k(\mathbf{w})-\Phi_{k+1}(\mathbf{w})-\frac{(1-\sigma)(2-\gamma)}{\gamma^2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|_Q^2-\frac{(2-\gamma)}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2.
\end{aligned}
$$

- For Algorithm 1 with (C2), it follows that

$$
\begin{aligned}
\text{RHS}\leq&\Phi_k(\mathbf{w})-\Phi_{k+1}(\mathbf{w})+\sigma(2-\gamma)\left\|\widetilde{\mathbf{x}}^{k-1}-\mathbf{x}^{k-1}\right\|_Q^2\\
&-(2-\gamma)\Big\{\|\widetilde{\mathbf{x}}^k-\mathbf{x}^k\|_Q^2+\frac{1}{2\beta}\|\widetilde{\lambda}^k-\lambda^k\|^2\Big\}\\
=&\Big\{\Phi_k(\mathbf{w})+\frac{\sigma(2-\gamma)}{\gamma^2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|_Q^2\Big\}-\Big\{\Phi_{k+1}(\mathbf{w})+\frac{\sigma(2-\gamma)}{\gamma^2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|_Q^2\Big\}\\
&-\frac{(1-\sigma)(2-\gamma)}{\gamma^2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|_Q^2-\frac{(2-\gamma)}{2\beta\gamma^2}\|\lambda^{k+1}-\lambda^k\|^2.
\end{aligned}
$$

- For Algorithm 1 with (C3), it follows that

$$
\begin{aligned}
\text{RHS} \leq & \Phi_k(\mathbf{w}) - \Phi_{k+1}(\mathbf{w}) + \frac{\sigma(2-\gamma)}{2\beta\gamma^2}\|\lambda^k - \lambda^{k-1}\|^2 \\
& - (2-\gamma)\Big\{\|\widetilde{\mathbf{x}}^k - \mathbf{x}^k\|_Q^2 + \frac{1}{2\beta}\|\widetilde{\lambda}^k - \lambda^k\|^2\Big\} \\
= & \Big\{\Phi_k(\mathbf{w}) + \frac{\sigma(2-\gamma)}{2\beta\gamma^2}\|\lambda^k - \lambda^{k-1}\|^2\Big\} - \Big\{\Phi_{k+1}(\mathbf{w}) + \frac{\sigma(2-\gamma)}{2\beta\gamma^2}\|\lambda^{k+1} - \lambda^k\|^2\Big\} \\
& - \frac{2-\gamma}{\gamma^2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_Q^2 - \frac{(1-\sigma)(2-\gamma)}{2\beta\gamma^2}\|\lambda^{k+1} - \lambda^k\|^2.
\end{aligned}
$$

Combining the above three cases and (14) together with the notations in (10) and (11), the result (12) is confirmed.

Finally, set $\mathbf{w} = \mathbf{w}^*$ in (12) and use (2) to complete the proof of (13). ■

Remark 3.1 *From the analysis of Corollary 3.1, the term* $\big(2|\langle v^k - \widetilde{\mathbf{x}}^k, d^k\rangle| + \|d^k\|^2\big)$ *can be also upper bounded by a convex combination of the two or three terms in the right-hand-side of (C1)-(C3). And similar convergence results will be achieved. Besides, comparing (4) to VI*$(\theta, \mathcal{J}, \mathcal{M})$*, the iterate* $\widetilde{\mathbf{w}}^k$ *will be a solution point of VI*$(\theta, \mathcal{J}, \mathcal{M})$ *if both* d^k *and* $(\mathbf{w}^k - \widetilde{\mathbf{w}}^k)$ *are zero, equivalently,* $(v^k - v^{k+1})$ *and* $(\mathbf{w}^k - \mathbf{w}^{k+1})$ *are zero. For this case, Algorithm 1 will reduce to an exact version without relaxation step.*

Remark 3.2 *Recalling the previous analysis in Section 3.1, one may use the following general iterate to replace the fourth step of Algorithm 1:*

$$
\widetilde{\lambda}^k = \lambda^k - s\beta\big[A(2\widetilde{\mathbf{x}}^k - \mathbf{x}^k) - b\big],
$$

where $s \in (0,1]$ *denotes a stepsize parameter. For this case, Lemma 3.1 still hold but with the lower-upper block of* H *given by (5) being replaced by* $\frac{1}{s\beta}\mathbf{I}$*. And the resulting matrix is still positive definite for any* $Q \succ \mathbf{0}$ *and* $s \in (0,1]$*. By multiplying the right-hand-side of (C3) by* $\frac{2-s}{s}$*, we can obtain the inequalities (12) and (13) but with the term* $\frac{1}{2\beta}\|\lambda^{k+1} - \lambda^k\|^2$ *being replaced by* $\frac{2-s}{2s\beta}\|\lambda^{k+1} - \lambda^k\|^2$*.*

3.2 Global convergence and sublinear convergence rate

Based on the key Corollary 3.1, this section aims to establish the global convergence of Algorithm 1 and its sublinear convergence rate in the sense of the primal iterative residual, the objective function value gap and constraint violation.

Theorem 3.1 *For Algorithm 1 with inexact criteria (C1), (C2) or (C3), we have*

(i) $\lim_{k\to\infty} \|\mathbf{w}^{k+1} - \mathbf{w}^k\| = 0$ *and* $\lim_{k\to\infty} \|A\mathbf{x}^{k+1} - b\| = 0$*;*

(ii) The sequence $\{\mathbf{w}^k\}$ *converges to a solution point* $\mathbf{w}^\infty \in \mathcal{M}^*$*;*

(iii) For any $k \geq 1$*, there exists an integer* $t \leq k$ *such that*

$$
\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 \leq \frac{1}{k}\frac{1}{\varphi(\sigma,\gamma)}\big[\Phi_1(\mathbf{w}^*) + \Psi_1\big].
$$

Proof. Sum (13) over $k = 1, 2, \cdots, \infty$ to obtain

$$
\sum_{k=1}^{\infty}\left(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_Q^2 + \frac{1}{2\beta}\|\lambda^{k+1} - \lambda^k\|^2\right) \leq \frac{1}{\varphi(\sigma,\gamma)}\big[\Phi_1(\mathbf{w}^*) + \Psi_1\big] < \infty. \tag{15}
$$

Together with $Q \succ \mathbf{0}$, the inequality (15) shows

$$\lim_{k\to\infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|\lambda^{k+1} - \lambda^k\| = 0.$$

So, the first result in (i) is confirmed. Combining the last two equations together with the sixth step of Algorithm 1 indicates

$$\lim_{k\to\infty} \|\widetilde{\mathbf{x}}^k - \mathbf{x}^k\|^2 = 0 \quad \text{and} \quad \lim_{k\to\infty} \|\widetilde{\lambda}^k - \lambda^k\|^2 = 0. \tag{16}$$

Together with a variant of the fourth step of Algorithm 1, that is,

$$\frac{\lambda^k - \widetilde{\lambda}^k}{\beta} = A\widetilde{\mathbf{x}}^k - b + A(\widetilde{\mathbf{x}}^k - \mathbf{x}^k) = A\mathbf{x}^k - b + 2A(\widetilde{\mathbf{x}}^k - \mathbf{x}^k),$$

and the sixth step of Algorithm 1, we further deduce

$$\lim_{k\to\infty} \left\|A\mathbf{x}^{k+1} - b\right\| = \lim_{k\to\infty} \left\|\frac{\lambda^k - \widetilde{\lambda}^k}{\beta} + (\gamma - 2)A(\widetilde{\mathbf{x}}^k - \mathbf{x}^k)\right\| = 0.$$

That is, the second result in (i) is ture.

For any $\mathbf{w}^* \in \mathcal{M}^*$, it follows from (13) that

$$\Phi_{k+1}(\mathbf{w}^*) \le \Phi_{k+1}(\mathbf{w}^*) + \Psi_{k+1} \le \Phi_k(\mathbf{w}^*) + \Psi_k \le \Phi_1(\mathbf{w}^*) + \Psi_1,$$

which, by using the definition of $\Phi_k(\mathbf{w})$ and $\gamma \in (0,2)$, shows that the sequences $\{\mathbf{w}^k\}, \{v^k\}$ are bounded. In addition, $\{\widetilde{\mathbf{w}}^k\}$ is also bounded by (7) and the boundedness of $\{\mathbf{w}^k\}$. Let $\mathbf{w}^\infty = (\mathbf{x}^\infty, \lambda^\infty)$ be a cluster point of $\{\widetilde{\mathbf{w}}^k\}$ and $\{\widetilde{\mathbf{w}}^{k_j}\}$ be the subsequence converging to $\mathbf{w}^\infty$. Combine (16) with (C1), (C2) or (C3) to achieve

$$\lim_{k\to\infty} \langle v^k - \widetilde{\mathbf{x}}^k, d^k\rangle = 0 \quad \text{and} \quad \lim_{k\to\infty} \|d^k\| = 0.$$

For any fixed $\mathbf{w} \in \mathcal{M}$, taking $\widetilde{\mathbf{w}}^{k_j}$ in (4) together with (3), letting j go to ∞ we have

$$\theta(\mathbf{x}) - \theta(\mathbf{x}^\infty) + \left\langle \mathbf{w} - \mathbf{w}^\infty, \mathcal{J}(\mathbf{w}^\infty)\right\rangle \ge 0,$$

which means that $\mathbf{w}^\infty \in \mathcal{M}^*$ is a solution point of VI$(\theta, \mathcal{J}, \mathcal{M})$. Since $\mathbf{w}^\infty \in \mathcal{M}^*$, for all $l \ge k_j$ it follows from (13) again that

$$\Phi_l(\mathbf{w}^\infty) + \Psi_l \le \Phi_{k_j}(\mathbf{w}^\infty) + \Psi_{k_j}.$$

This together with (i), $\lim_{j\to\infty} \widetilde{\mathbf{w}}^{k_j} = \mathbf{w}^\infty$ and the positive definiteness of H shows $\lim_{l\to\infty} \mathbf{w}^l = \mathbf{w}^\infty$. Namely, the sequence $\{\mathbf{w}^k\}$ cannot have another cluster point and thus it converges to the solution $\mathbf{w}^* = \mathbf{w}^\infty \in \mathcal{M}^*$.

Finally, let $k \ge 1$ be any fixed constant and $t \le k$ be a positive integer such that

$$\left\|\mathbf{w}^{t+1} - \mathbf{w}^t\right\| = \min\left\{\left\|\mathbf{w}^{l+1} - \mathbf{w}^l\right\| \mid l = 1, \dots, k\right\}.$$

Then, by (15) again we deduce $\left\|\mathbf{w}^{t+1} - \mathbf{w}^t\right\|^2 \le \frac{1}{k}\frac{1}{\varphi(\sigma,\gamma)}\left[\Phi_1(\mathbf{w}^*) + \Psi_1\right]$. ■

The results in Theorem 3.1 requires $Q \succ \mathbf{0}$ that is the condition we explained in Algorithm 1. According to the first conclusion in Theorem 3.1, one may use $\max\left\{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|, \|\lambda^{k+1} - \lambda^k\|\right\} \le \epsilon$ or the associated relative error as a simple stopping criterion for Algorithm 1, where $0 < \epsilon < 1$ is a given tolerance error. The second conclusion shows that $\mathbf{x}^\infty, \lambda^\infty$ are the solution point of the primal problem (1) and its dual respectively, while the third shows that Algorithm 1 has a sublinear convergence rate in terms of the primal iterative residual.

Next, we establish its sublinear convergence rate in terms of the objective function value gap and constraint violation for the following ergodic iterates (it was firstly appeared in [2])

$$\bar{\mathbf{w}}_T := \frac{1}{T}\sum_{k=\kappa}^{T+\kappa} \widetilde{\mathbf{w}}^k \quad \text{and} \quad \bar{\mathbf{x}}_T := \frac{1}{T}\sum_{k=\kappa}^{T+\kappa} \widetilde{\mathbf{x}}^k, \quad \forall \kappa, T > 0. \tag{17}$$

Theorem 3.2 *Let $\bar{\mathbf{x}}_T$ be defined in (17). Then, for any integer $T > 0$, we have*

$$\left|\theta(\bar{\mathbf{x}}_T) - \theta(\mathbf{x}^*)\right| \le \frac{1}{2T}\left(\frac{2\rho_0^2}{\gamma\beta} + c_0\right) \quad \text{and} \quad \left\|A\bar{\mathbf{x}}_T - b\right\| \le \frac{1}{2T}\left(\frac{2\rho_0^2}{\gamma\beta} + c_0\right), \tag{18}$$

where $\beta > 0, \gamma \in (0, 2)$, c_0 is a certain positive constant and

$$\rho_0 = \max\left\{1 + \|\lambda^*\|, 2\|\lambda^*\|\right\}. \tag{19}$$

Proof. Summing (12) over k between κ and $\kappa + T$ yields

$$\frac{1}{T}\sum_{k=\kappa}^{T+\kappa}\left[\theta(\widetilde{\mathbf{x}}^k) - \theta(\mathbf{x}) + \left\langle \widetilde{\mathbf{w}}^k - \mathbf{w}, \mathcal{J}(\mathbf{w})\right\rangle\right] \le \frac{1}{2T}\Big(\Phi_\kappa(\mathbf{w}) + \Psi_\kappa\Big).$$

By making use of the convexity of θ, the definitions in (17) and the structure of H, take $\mathbf{w} = (\mathbf{x}^*, \lambda)$ in the last inequality to obtain

$$\begin{aligned}
&\theta(\bar{\mathbf{x}}_T) - \theta(\mathbf{x}^*) - \lambda^\mathsf{T}(A\bar{\mathbf{x}}_T - b)\\
\le&\frac{1}{2T\gamma}\left(\|\mathbf{x}^* - \mathbf{x}^\kappa\|^2_{2\beta A^\mathsf{T}A+Q} + 2(\lambda - \lambda^\kappa)^\mathsf{T}A(\mathbf{x}^* - \mathbf{x}^\kappa) + \frac{1}{\beta}\|\lambda - \lambda^\kappa\|^2\right)\\
&+\frac{1}{2T}\left(\left\|\mathbf{x}^* - v^\kappa\right\|^2 + \Psi_\kappa\right)\\
\le&\frac{1}{2T}\left\{\frac{1}{\gamma}\left(\|\mathbf{x}^* - \mathbf{x}^\kappa\|^2_{3\beta A^\mathsf{T}A+Q} + \frac{2}{\beta}\|\lambda - \lambda^\kappa\|^2\right) + \left\|\mathbf{x}^* - v^\kappa\right\|^2 + \Psi_\kappa\right\}\\
=&\frac{1}{2T}\Big(\frac{2}{\gamma\beta}\|\lambda - \lambda^\kappa\|^2 + c_0\Big),
\end{aligned}$$

where the second inequality follows from the Cauchy-Schwartz inequality

$$2(\lambda - \lambda^\kappa)^\mathsf{T}A(\mathbf{x}^* - \mathbf{x}^\kappa) \le \left\|\mathbf{x}^* - \mathbf{x}^\kappa\right\|^2_{\beta A^\mathsf{T}A} + \frac{1}{\beta}\left\|\lambda - \lambda^\kappa\right\|^2$$

and $c_0 = \frac{1}{\gamma}\|\mathbf{x}^* - \mathbf{x}^\kappa\|^2_{3\beta A^\mathsf{T}A+Q} + \|\mathbf{x}^* - v^\kappa\|^2 + \Psi_\kappa$ is certain constant depending on the problem data and the parameters of Algorithm 1. For any $\rho > 0$, according to Lemma 2.1 we have

$$\theta(\bar{\mathbf{x}}_T) - \theta(\mathbf{x}^*) + \rho\|A\bar{\mathbf{x}}_T - b\| \le \frac{1}{2T}\Big(c_0 + \frac{2}{\gamma\beta}\sup_{\|\lambda\|\le\rho}\|\lambda - \lambda^\kappa\|^2\Big).$$

Moreover, letting $\rho > \|\lambda^*\|$ and applying Lemma 2.2 to above inequality, we have

$$\left\|A\bar{\mathbf{x}}_T - b\right\| \le \frac{1}{2T}\frac{2\rho^2/(\gamma\beta) + c_0}{\rho - \|\lambda^*\|},$$

and

$$-\frac{1}{2T}\frac{\|\lambda^*\|}{\rho - \|\lambda^*\|}\left(\frac{2\rho^2}{\gamma\beta} + c_0\right) \le \theta(\bar{\mathbf{x}}_t) - \theta(\mathbf{x}^*) \le \frac{1}{2T}\left(\frac{2\rho^2}{\gamma\beta} + c_0\right).$$

Taking $\rho = \rho_0$ in the above two relations with ρ_0 given by (19), we can get (18) immediately. This completes the whole proof. ∎

Remark 3.3 *According to the analysis of Corollary 3.1, one may exploit the following simple error criteria*

$$\|d^k\|^2 \le 2\sigma\big|\langle v^k - \tilde{\mathbf{x}}^k, d^k\rangle\big|, \quad \forall \sigma \in (0,1] \tag{C4}$$

when solving the subproblem of Algorithm 1 inexactly. And the convergence of Algorithm 1 under (C4) can be established similar to the previous analysis. In practice, combining (14) and (C4) is to obtain

$$\begin{aligned}&2\Big\{\theta(\tilde{\mathbf{x}}^k) - \theta(\mathbf{x}) + \big\langle \tilde{\mathbf{w}}^k - \mathbf{w}, \mathcal{J}(\mathbf{w})\big\rangle\Big\}\\ \le& \Phi_k(\mathbf{w}) - \Phi_{k+1}(\mathbf{w}) - \frac{1-\sigma}{\sigma}\|d^k\|^2 - (2-\gamma)\Big\{\|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_Q^2 + \frac{1}{2\beta}\|\tilde{\lambda}^k - \lambda^k\|^2\Big\},\end{aligned}$$

which, by setting $\mathbf{w} = \mathbf{w}^*$ together with (2), gives

$$\Phi_{k+1}(\mathbf{w}^*) \le \Phi_k(\mathbf{w}^*) - \frac{1-\sigma}{\sigma}\|d^k\|^2 - (2-\gamma)\Big\{\|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_Q^2 + \frac{1}{2\beta}\|\tilde{\lambda}^k - \lambda^k\|^2\Big\}.$$

Then, Algorithm 1 under (C4) converges globally with a sublinear convergence rate. A surprising observation in experiments is that AI-ALM with the criteria (C4) performs very well.

4 Numerical experiments

In this section, we apply our proposed method to solve a class of image restoration problems and then we report some related numerical results. All experiments are coded in Matlab R2017a on a PC with Intel Core i5 processor (2.5HZ) and 16GB memory.

Consider the following total variation (TV) uniform noise removal model in [16]:

$$\min\{\big\||\nabla \mathbf{x}|\big\|_1 \mid \big\|H\mathbf{x} - \bar{\mathbf{x}}\big\|_\infty \le \delta\},$$

where $\bar{\mathbf{x}} \in \mathcal{R}^l$ is an observed image corrupted by a zero-mean uniform noise; $H \in \mathcal{R}^{l\times l}$ denotes a blurring operator matrix; $\big\||\nabla\cdot|\big\|_1$ is the TV norm of the divergence operator $\nabla\cdot$ (see e.g. [13]), $\delta > 0$ is a parameter indicating the uniform noise level and $\big\|\mathbf{x}\big\|_\infty = \max_{1\le i\le l}|\mathbf{x}_i|$.

By denoting

$$A = \begin{pmatrix} H \\ -H \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} \bar{\mathbf{x}} - \delta\mathbf{1} \\ -\bar{\mathbf{x}} - \delta\mathbf{1} \end{pmatrix},$$

with $\mathbf{1} = (1,1,\ldots,1)^\mathsf{T} \in \mathcal{R}^l$, the above model can be rewritten as

$$\min\{\big\||\nabla \mathbf{x}|\big\|_1 \mid A\mathbf{x} \ge b\} \tag{20}$$

which is a special case of (1). Thus, the proposed AI-ALM can be applied to solve it.

As explained in Introduction, we take $Q = \tau\mathbf{I} - 2\beta A^\mathsf{T}A$ with $\tau > 2\beta\|A^\mathsf{T}A\|$ in experiments. Then, the core subproblem of AI-ALM is

$$\tilde{\mathbf{x}}^k \approx \arg\min_{\mathbf{x}\in\mathcal{R}^l} \big\||\nabla\mathbf{x}|\big\|_1 + \frac{\tau}{2}\Big\|\mathbf{x} - \frac{A^\mathsf{T}\lambda^k + \tau\mathbf{x}^k}{\tau}\Big\|^2.$$

In the following experiments, the method in [5] is used to solve the above subproblem to obtain an inexact solution per iteration, and AI-ALM terminates when the previous mentioned inexact criterions are satisfied or execute 10 iterations. For simplicity, we denote AI-ALM by AI-ALM1, AI-ALM2, AI-ALM3 when the inexact criterions (C1), (C2) and (C3) are adopted respectively. We also denote AI-ALM as AI-ALM4 when (C4) is utilized as we discussed in Remark 3.3. These methods are compared with the specialized algorithm C-PPA [9] whose main subproblem still use the method in [5]. We don't compare our algorithm with other image restoration algorithms

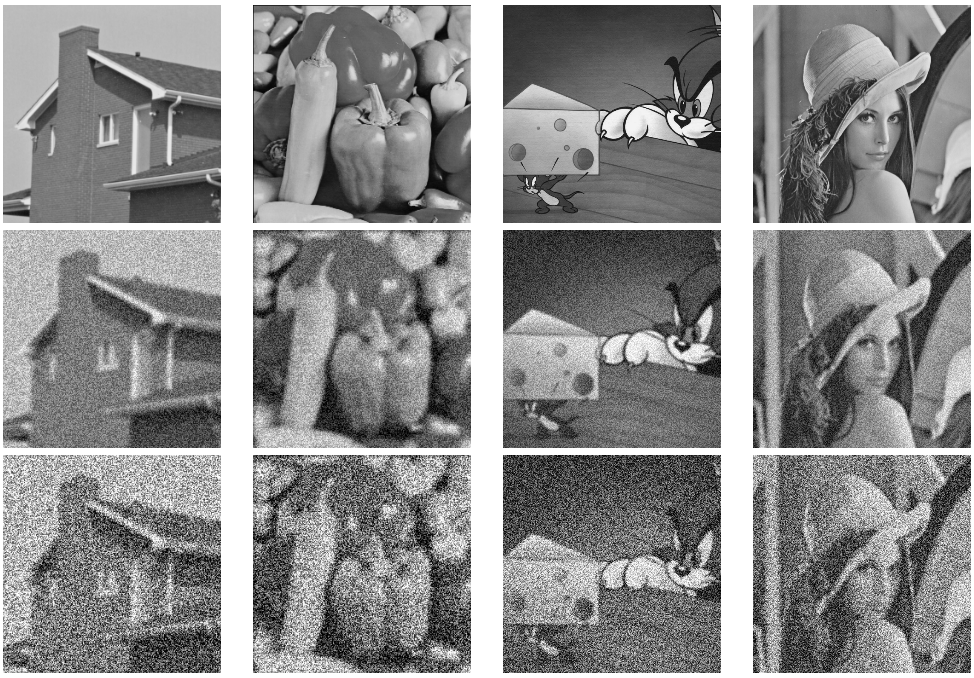

Figure 1: The first row is the clean images: House, Pepper, TomAndJerry and Lena, while the last two rows are degraded images with $\delta = 0.2$ and 0.5, respectively.

since C-PPA had been shown better than some state-of-the-art algorithms, see [9, Section 5]. The parameters of C-PPA use the default settings as mentioned therein.

We take $(\mathbf{x}^0, v^0, \lambda^0) = (\bar{\mathbf{x}}, \bar{\mathbf{x}}, \mathbf{0})$ as the initial point and use the tuned parameters $\beta = 12, \tau = 50, \sigma = 0.99, \gamma = 1.8$ for our proposed algorithms. The following Signal-to-Noise Ration (SNR) in decibel (dB) defined by

$$\mathrm{SNR} = 20 \log_{10} \frac{\|\bar{\mathbf{x}}\|}{\|\bar{\mathbf{x}} - \mathbf{x}^{\dagger}\|}$$

is used to measure the quality of a restored image, where $\bar{\mathbf{x}}$ denotes the origin clean image and $\mathbf{x}^{\dagger}$ denotes the restored image obtained by an algorithm. All mentioned algorithms will terminate when SNR no longer changes, which means that SNR value difference between the current iteration and the last iteration is controlled within 10^{-2}.

We test different type of images: House (256×256), Pepper (256×256), TomAndJerry (256×256) and Lena (512×512), as shown in the first row of Fig 1. The origin images are degraded by the Gaussian filter of size 9 with standard deviation 2.5, and then a zero-mean uniform noise with $\delta = 0.2$ or 0.5 is added to the degraded images. Those degraded images are shown respectively in the last two rows of Fig 1. Tables 1-2 report some numerical results of solving (20) with $\delta \in \{0.2, 0.5\}$, including the number of outer iterations (Oiter), the total number of inner iterations (Iiter)("-" for C-PPA means it executes 10 iterations for each outer iteration), the CPU time in seconds (Time). The notations "SNR(input)" and "SNR(final)" denote the SNR value of the corrupted images and the final restored images, respectively. Note that, the larger the obtained SNR value, the better the quality of a restored image. From Tables 1-2, we can see that the outer iteration number of our AI-ALM is smaller than C-PPA, and the CPU time of AI-ALM4 is of the smallest among the compared algorithms to obtain a relatively better SNR finally.

Fig 2 depicts the tendency of SNR versus the CPU time, showing intuitionally the advantage of our algorithms, especially the algorithm AI-ALM4 (it can obtain a large SNR value with relatively cheaper CPU time). Of course, it is undeniable that C-PPA may recover the image of average quality quickly when the image is seriously polluted. Fig 3 and Fig 4 visualize the images restored by all compared algorithms, from which we can see all methods can restore the degraded images with good qualities.

Table 1: Numerical results of different algorithms for solving (20) with $\delta = 0.2$.

Image	SNR(input)	Method	Oiter	Iiter	Time	SNR(final)
House	13.23	AI-ALM1	17	17	0.47	22.39
		AI-ALM2	17	26	0.61	22.39
		AI-ALM3	17	26	0.44	22.39
		AI-ALM4	17	17	0.24	22.39
		C-PPA	51	-	1.04	21.41
Pepper	11.81	AI-ALM1	16	16	0.48	17.85
		AI-ALM2	16	25	0.56	17.85
		AI-ALM3	16	25	0.43	17.85
		AI-ALM4	16	16	0.23	17.85
		C-PPA	35	-	0.72	17.10
TomAndJerry	11.18	AI-ALM1	21	21	3.43	19.03
		AI-ALM2	21	30	3.95	19.03
		AI-ALM3	21	30	2.89	19.03
		AI-ALM4	21	21	1.71	19.03
		C-PPA	44	-	6.72	18.24
Lena	12.56	AI-ALM1	16	16	2.59	22.23
		AI-ALM2	16	25	3.22	22.23
		AI-ALM3	16	25	2.30	22.23
		AI-ALM4	16	16	1.34	22.23
		C-PPA	43	-	6.61	21.53

Table 2: Numerical results of different algorithms for solving (20) with $\delta = 0.5$.

Image	SNR(input)	Method	Oiter	Iiter	Time	SNR(final)
House	5.80	AI-ALM1	25	25	0.70	20.64
		AI-ALM2	25	34	0.80	20.65
		AI-ALM3	25	34	0.60	20.65
		AI-ALM4	25	25	0.35	20.64
		C-PPA	45	-	0.93	20.09
Pepper	4.94	AI-ALM1	26	26	0.73	16.81
		AI-ALM2	26	35	0.82	16.81
		AI-ALM3	26	35	0.62	16.81
		AI-ALM4	26	26	0.35	16.81
		C-PPA	36	-	0.75	16.31
TomAndJerry	4.02	AI-ALM1	27	27	4.30	17.80
		AI-ALM2	27	36	4.92	17.80
		AI-ALM3	27	36	3.55	17.80
		AI-ALM4	27	27	2.20	17.80
		C-PPA	42	-	6.41	17.34
Lena	5.03	AI-ALM1	28	28	4.44	20.76
		AI-ALM2	28	37	5.09	20.76
		AI-ALM3	28	37	3.64	20.76
		AI-ALM4	28	28	2.27	20.76
		C-PPA	37	-	5.62	20.30

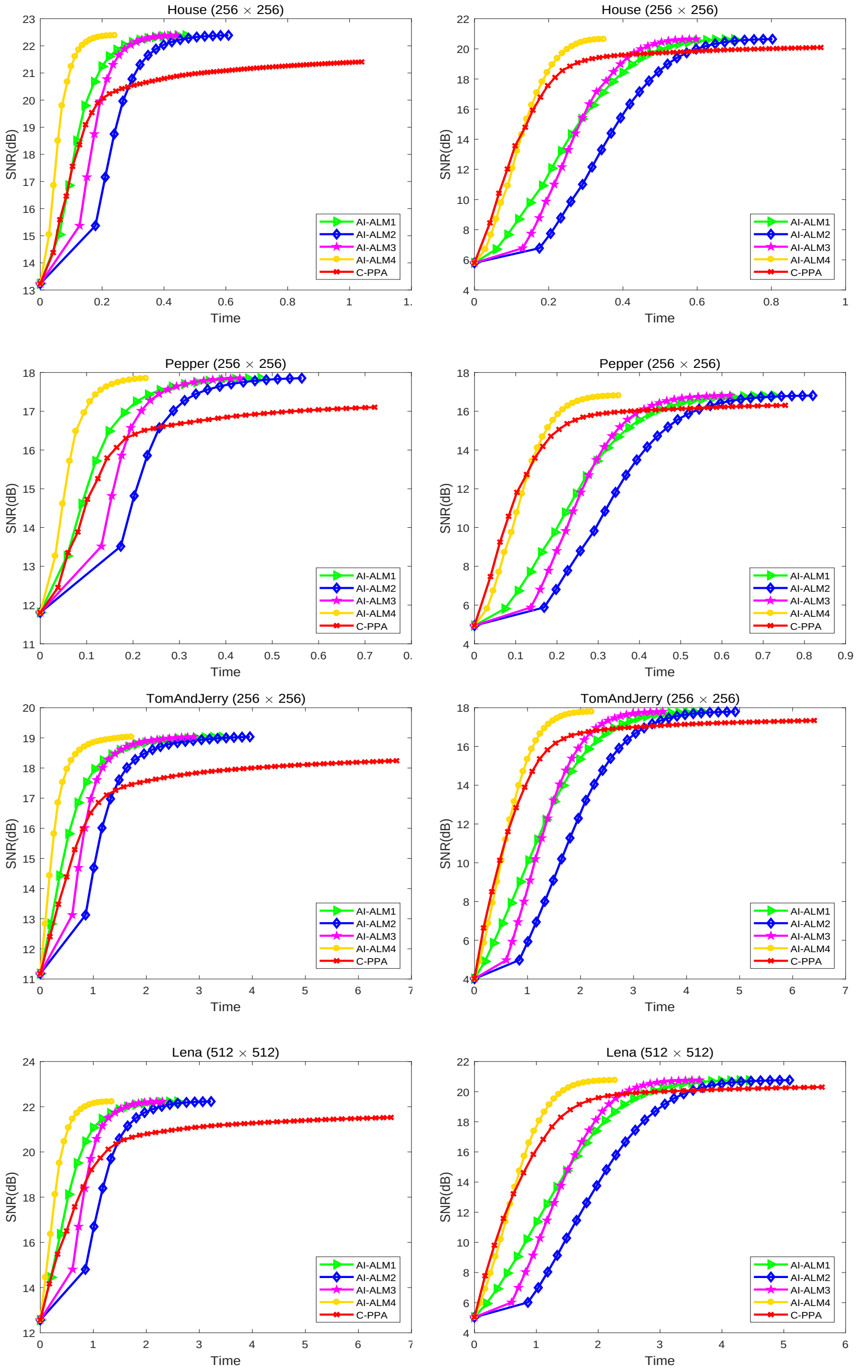

Figure 2: Comparison curves of the SNR vs the CPU time by different algorithms for solving (20) with $\delta = 0.2$ (the left column) and $\delta = 0.5$ (the right column), respectively.

5 Brief conclusion

In this paper, we proposed an accelerated inexact augmented Lagrangian method to solve the convex optimization problems with equality or inequality constraints. We designed several different inexact criteria to control the accuracy of the solution to the core subproblem and we proved the global convergence of the new method and its sublinear convergence rate in a unified framework. To investigate the numerical efficiency of the proposed algorithm under different criteria, we applied it to solve a class of image restoration problems. For the theoretical interests, we also extended the proposed algorithm to solve the multi-block problem in the appendix and briefly discussed the related convergence results. Notice that in the basic AI-ALM and its extended algorithm, a positive definite matrix Q is exploited to simplify the solving difficulty of involved subproblem. So, whether a positive indefinite matrix can be utilized is an interesting topic in the future work.

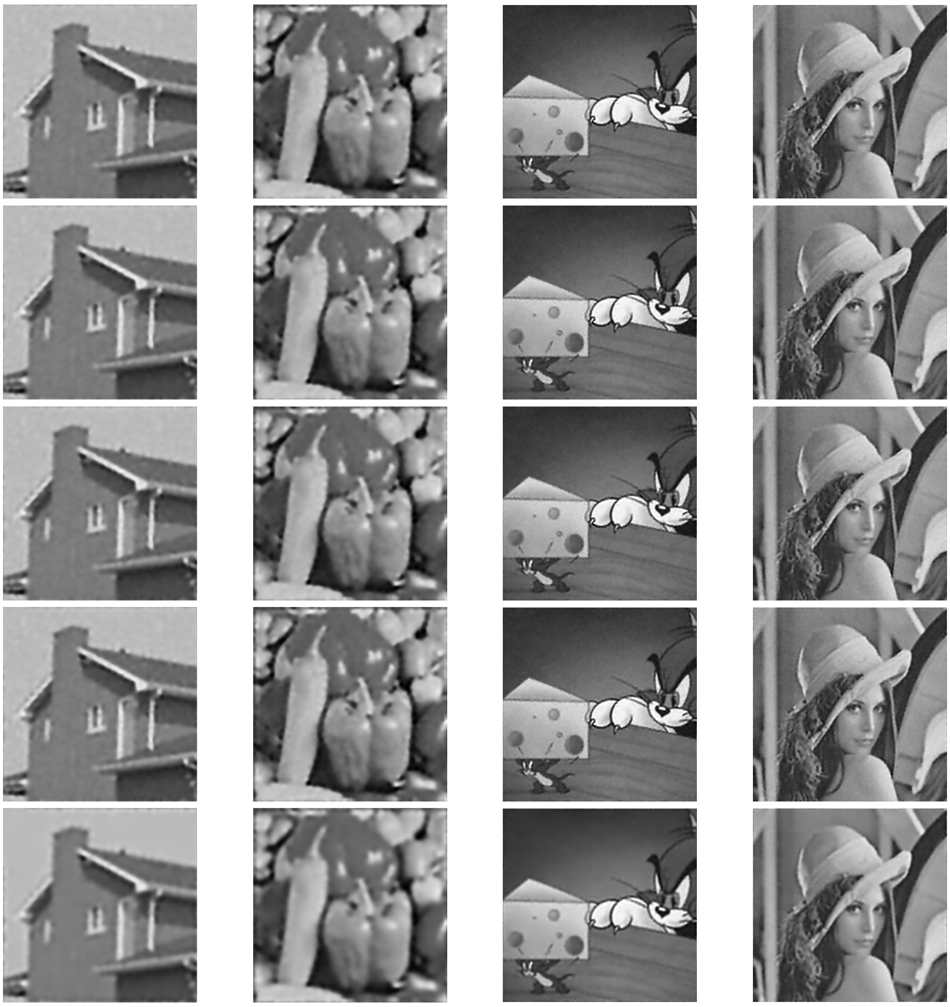

Figure 3: The restored images of different algorithms for solving (20) with $\delta = 0.2$: (from top to bottom) AI-ALM1, AI-ALM2, AI-ALM3, AI-ALM4 and C-PPA, respectively.

Appendix: Extension of AI-ALM to multi-block case

This appendix aims to briefly discuss the convergence of an extended AI-ALM (eAI-ALM, see Algorithm 2) for solving a multi-block convex programming in the form of

$$\min_{\mathbf{x}_i \in \mathcal{R}^{n_i}} \Big\{\theta(\mathbf{x}) := \sum_{i=1}^{p} \theta_i(\mathbf{x}_i) |\ \sum_{i=1}^{p} A_i \mathbf{x}_i = b \text{ (or } \geq b)\Big\}, \tag{A1}$$

where $\theta_i(\mathbf{x}_i) : \mathcal{R}^{n_i} \to \mathcal{R}, i = 1, 2, \cdots, p$ are proper closed convex functions; $A_i \in \mathcal{R}^{m \times n_i}$ and $b \in \mathcal{R}^m$ are given data. Problems in the form of (A1) arise in many practical applications such as statistical learning [1], video surveillance [15] and so on.

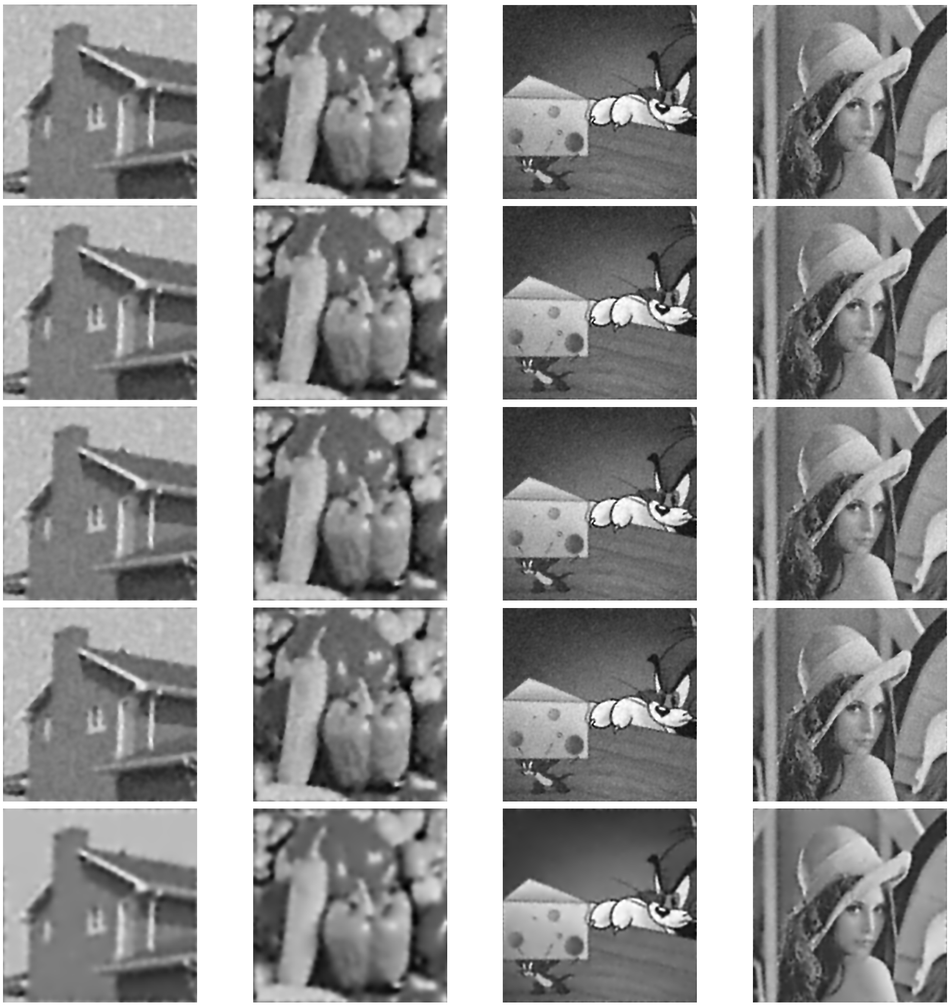

Figure 4: The restored images of different algorithms for solving (20) with $\delta = 0.5$: (from top to bottom) AI-ALM1, AI-ALM2, AI-ALM3, AI-ALM4 and C-PPA, respectively.

Note that, if some of the involved subproblems have exact solution to be solved easily, then we can only focus on the solving of other subproblems inexactly. To simplify the convergence of Algorithm 2, let us define

$$\mathcal{M} := \prod_{i=1}^{p} \mathcal{R}^{n_i} \times \Lambda, \quad \text{where } \Lambda := \begin{cases} \mathcal{R}^m, & \text{if } \sum_{i=1}^{p} A_i \mathbf{x}_i = b, \\ \mathcal{R}^m_+, & \text{if } \sum_{i=1}^{p} A_i \mathbf{x}_i \geq b, \end{cases}$$

and

Algorithm 2 [eAI-ALM for solving Problem (A1)]

Initialize $(\mathbf{x}_1^0,\dots,\mathbf{x}_p^0,v_1^0,\dots,v_p^0,\lambda^0)$ and set $\beta_i>0, Q_i\succ\mathbf{0},\sigma\in[0,1),\gamma\in(0,2)$;
While stopping criteria is not satisfied **do**
    For $i=1,\cdots,p$, parallelly update
        $\widetilde{\mathbf{x}}_i^k \approx \arg\min\limits_{\mathbf{x}_i\in\mathcal{R}^{n_i}} \theta_i(\mathbf{x}_i)-\left\langle\lambda^k,A_i\mathbf{x}_i-b\right\rangle+\beta_i\left\|A_i(\mathbf{x}_i-\mathbf{x}_i^k)\right\|^2+\frac{1}{2}\left\|\mathbf{x}_i-\mathbf{x}_i^k\right\|_{Q_i}^2$ such that (C1), (C2) or (C3) holds with

$$d_i^k\in\partial\theta_i(\widetilde{\mathbf{x}}_i^k)-A_i^{\mathsf{T}}\lambda^k+\left(2\beta_iA_i^{\mathsf{T}}A_i+Q_i\right)\left(\widetilde{\mathbf{x}}_i^k-\mathbf{x}_i^k\right);$$

    End for
    Update $\widetilde{\lambda}^k=\lambda^k-\beta\left[\sum_{i=1}^p A_i(2\widetilde{\mathbf{x}}_i^k-\mathbf{x}_i^k)-b\right]$ where $\beta=1/\sum_{i=1}^p\frac{1}{\beta_i}$;
    Update the auxiliary variable $v_i^{k+1}=v_i^k-d_i^k$ for $i=1,\cdots,p$;
    Relaxation step: $\begin{pmatrix}\mathbf{x}_1^{k+1}\\ \vdots\\ \mathbf{x}_p^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix}\mathbf{x}_1^{k}\\ \vdots\\ \mathbf{x}_p^{k}\\ \lambda^{k}\end{pmatrix}+\gamma\begin{pmatrix}\widetilde{\mathbf{x}}_1^{k}-\mathbf{x}_1^k\\ \vdots\\ \widetilde{\mathbf{x}}_p^{k}-\mathbf{x}_p^k\\ \widetilde{\lambda}^{k}-\lambda^k\end{pmatrix}$.
End while

$$\mathbf{w}=\begin{pmatrix}\mathbf{x}\\ \lambda\end{pmatrix}=\begin{pmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_p\\ \lambda\end{pmatrix},\ \widetilde{\mathbf{w}}^k=\begin{pmatrix}\widetilde{\mathbf{x}}^k\\ \widetilde{\lambda}^k\end{pmatrix}=\begin{pmatrix}\widetilde{\mathbf{x}}_1^k\\ \vdots\\ \widetilde{\mathbf{x}}_p^k\\ \widetilde{\lambda}^k\end{pmatrix},\ \mathcal{J}(\mathbf{w})=\begin{pmatrix}-A_1^{\mathsf{T}}\lambda\\ \vdots\\ -A_p^{\mathsf{T}}\lambda\\ \sum\limits_{i=1}^p A_i\mathbf{x}_i-b\end{pmatrix}.$$

We provide a similar lemma to Lemma 3.1 to establish the convergence of eAI-ALM.

Lemma A.1 *The iterates generated by eAI-ALM satisfy (4) and (6) where*

$$d^k=\begin{pmatrix}d_1^k\\ d_2^k\\ \vdots\\ d_p^k\end{pmatrix},\ H=\begin{bmatrix}2\beta_1A_1^{\mathsf{T}}A_1+Q_1 & \cdots & \mathbf{0} & A_1^{\mathsf{T}}\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \cdots & 2\beta_pA_p^{\mathsf{T}}A_p+Q_p & A_p^{\mathsf{T}}\\ A_1 & \cdots & A_p & \sum\limits_{i=1}^p\frac{1}{\beta_i}\mathbf{I}\end{bmatrix}$$

is symmetric positive definite for any $\beta_i>0$ and $Q_i\succ-\beta_iA_i^{\mathsf{T}}A_i(i=1,\cdots,p)$. Moreover, the inequalities (12) and (13) hold with $Q=diag(Q_1,\cdots,Q_p)$.

Proof. We first prove the positive definiteness of H. For any $\beta_i>0$ and $Q_i\succ\mathbf{0}\succ-\beta_iA_i^{\mathsf{T}}A_i$, by the structure of H given by Lemma A.1 we have

$$\begin{aligned}H &\succ \begin{bmatrix}\beta_1A_1^{\mathsf{T}}A_1 & \cdots & \mathbf{0} & A_1^{\mathsf{T}}\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \cdots & \beta_pA_p^{\mathsf{T}}A_p & A_p^{\mathsf{T}}\\ A_1 & \cdots & A_p & \sum\limits_{i=1}^p\frac{1}{\beta_i}\mathbf{I}\end{bmatrix}\\ &= \begin{bmatrix}\beta_1A_1^{\mathsf{T}}A_1 & \cdots & \mathbf{0} & A_1^{\mathsf{T}}\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\ A_1 & \cdots & \mathbf{0} & \frac{1}{\beta_1}\mathbf{I}\end{bmatrix}+\cdots+\begin{bmatrix}\mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\ \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \cdots & \beta_pA_p^{\mathsf{T}}A_p & A_p^{\mathsf{T}}\\ \mathbf{0} & \cdots & A_p & \frac{1}{\beta_p}\mathbf{I}\end{bmatrix}\end{aligned}$$

$$
= \begin{pmatrix} \sqrt{\beta_1} A_1^{\mathsf{T}} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \frac{1}{\sqrt{\beta_1}} \mathbf{I} \end{pmatrix} \begin{pmatrix} \sqrt{\beta_1} A_1^{\mathsf{T}} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \frac{1}{\sqrt{\beta_1}} \mathbf{I} \end{pmatrix}^{\mathsf{T}} + \ldots + \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \sqrt{\beta_p} A_p^{\mathsf{T}} \\ \frac{1}{\sqrt{\beta_p}} \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \sqrt{\beta_p} A_p^{\mathsf{T}} \\ \frac{1}{\sqrt{\beta_p}} \mathbf{I} \end{pmatrix}^{\mathsf{T}}.
$$

Clearly, the term in the right-hand-side of the above final equality is nonnegative and hence the matrix H is symmetric positive definite.

By the fourth step of Algorithm 2, we have for any $i = 1, 2, \cdots, p$, that

$$
\theta_i(\mathbf{x}_i) - \theta_i(\widetilde{\mathbf{x}}_i^k) + \left\langle \mathbf{x}_i - \widetilde{\mathbf{x}}_i^k, -A_i^{\mathsf{T}} \lambda^k + \left(2\beta_i A_i^{\mathsf{T}} A_i + Q_i\right)(\widetilde{\mathbf{x}}_i^k - \mathbf{x}_i^k) - d_i^k \right\rangle \geq 0, \ \forall \mathbf{x}_i \in \mathcal{R}^{n_i},
$$

that is,

$$
\begin{aligned}
&\theta_i(\mathbf{x}_i) - \theta(\widetilde{\mathbf{x}}_i^k) + \left\langle \mathbf{x}_i - \widetilde{\mathbf{x}}_i^k, -A_i^{\mathsf{T}} \widetilde{\lambda}^k \right\rangle - \left\langle \mathbf{x}_i - \widetilde{\mathbf{x}}_i^k, d_i^k \right\rangle \\
\geq &\left\langle \mathbf{x}_i - \widetilde{\mathbf{x}}_i^k, \left(2\beta_i A_i^{\mathsf{T}} A_i + Q_i\right)(\mathbf{x}_i^k - \widetilde{\mathbf{x}}_i^k) + A_i^{\mathsf{T}}(\lambda^k - \widetilde{\lambda}^k) \right\rangle.
\end{aligned}
$$

Besides, it follows from the update of $\widetilde{\lambda}^k$ that $\widetilde{\lambda}^k \in \Lambda$ and

$$
\left\langle \lambda - \widetilde{\lambda}^k, \sum_{i=1}^p A_i \widetilde{\mathbf{x}}_i^k - b \right\rangle = \left\langle \lambda - \widetilde{\lambda}^k, \sum_{i=1}^p A_i(\mathbf{x}_i^k - \widetilde{\mathbf{x}}_i^k) + \sum_{i=1}^p \frac{1}{\beta_i}(\lambda^k - \widetilde{\lambda}^k) \right\rangle, \ \forall \lambda \in \Lambda.
$$

Combine the last inequality and equality together with the block structure of H to obtain the inequality (4). Then, analogous to the rest proof of Corollary 3.1, the inequality (6) holds too.

The inequalities (12) and (13) can be proved similar to Corollary 3.1 with Q being a diagonal matrix, that is, $Q = \mathrm{diag}(Q_1, \cdots, Q_p)$. ■

Based on Lemma A.1, the global convergence and sublinear convergence rate of eAI-ALM can be established as the rest parts of Section 3.2. Notice that both Algorithm 1 and Algorithm 2 obey primal-dual updates. In practice, we can first update the dual variable and then update the primal variables to obtain a dual-primal version which is omitted here. Interested readers can find the similar discussions in [4, Remark 3.1].

References

[1] J. Bai, J. Li, F. Xu, H. Zhang, Generalized symmetric ADMM for separable convex optimization, Comput. Optim. Appl. 70 (2018), 129-170.

[2] J. Bai, W. Hager, H. Zhang, An inexact accelerated stochastic ADMM for separable convex optimization, Comput. Optim. Appl. 81 (2022), 479-518.

[3] J. Bai, D. Han, H. Sun, H. Zhang, Convergence analysis of an inexact accelerated stochastic ADMM with larger stepsizes, CSIAM Trans. Appl. Math. (2022), DOI: 10.4208/csiam-am.SO-2021-0021.

[4] J. Bai, H. Zhang, An new insight on augmented Lagrangian method with applications in machine learning, Optimization Online, (2021), http://www.optimization-online.org/DB_FILE/2021/08/8559.pdf.

[5] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging Vis. 20 (2004), 89-97.

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

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

[8] B. He, X Yuan, On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, SIAM J. Numer. Anal. 50 (2012), 700-709.

[9] B. He, X. Yuan, W. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl. 56 (2013), 559-572.

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

[11] M. Kang, M. Kang, M. Jung, Inexact accelerated augmented Lagrangian methods, Comput Optim. Appl. 62 (2015), 373-404.

[12] M. Li, L. Liao, X. Yuan, Inexact alternating direction methods of multipliers with logarithmic-quadratic proximal regularization, J. Optim. Theory Appl. 159 (2013), 412-436.

[13] Y. Ma, J. Bai, H. Sun, An inexact ADMM with proximal-indifinite term and larger stepsize, Optimization Online, (2022), http://www.optimization-online.org/DB_FILE/2022/02/8793.pdf.

[14] Z. Peng, D. Wu, An inexact parallel splitting augmented Lagrangian method for large system of linear equations, Appl. Math. Comput. 216 (2010), 1624-1636.

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

[16] P. Weiss, G. Aubert, L. Blanc-Fraud, Some applications of ℓ_∞-constraints in image processing, INRIA Research Report (2006).

[17] Z. Wu, M. Li, General inexact primal-dual hybrid gradient methods for saddle-point problems and convergence analysis, Asia Pac. J. Oper. Res. 8 (2021), 1-27.

[18] J. Xie, On inexact ADMMs with relative error criteria, Comput. Optim. Appl. 71 (2018), 743-765.

[19] Y. Zu, Accelerated first-order primal-dual proximal mehtods for linearly constrained composite convex programming, SIAM J. Optim. 27 (2017), 1459-1484.