Inertial primal-dual methods for linear equality constrained convex optimization problems

Xin He · Rong Hu · Ya-Ping Fang

Received: date / Accepted: date

Abstract  Inspired by a second-order primal-dual dynamical system [Zeng X, Lei J, Chen J. Dynamical primal-dual accelerated method with applications to network optimization. 2019; arXiv:1912.03690], we propose an inertial primal-dual method for the linear equality constrained convex optimization problem. When the objective function has a "nonsmooth + smooth" composite structure, we further propose an inexact inertial primal-dual method by linearizing the smooth individual function and solving the subproblem inexactly. Assuming merely convexity, we prove that the proposed methods enjoy $O(1/k^2)$ convergence rate on $\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)$ and $O(1/k)$ convergence rate on primal feasibility, where $\mathcal{L}$ is the Lagrangian function and $(x^*, \lambda^*)$ is a saddle point of $\mathcal{L}$. Numerical results are reported to demonstrate the validity of the proposed methods.

Keywords  Inertial primal-dual method · Linear equality constrained convex optimization problem · Convergence rate · Inexactness.

Mathematics Subject Classification (2010) 90C06 · 90C25 · 68W40 · 49M27

This work was partially supported by National Natural Science Foundation of China (11471230).

Xin He
Department of Mathematics, Sichuan University, Chengdu, Sichuan, P.R. China
E-mail: hexinuser@163.com

Rong Hu
Department of Applied Mathematics, Chengdu University of Information Technology, Chengdu, Sichuan, P.R. China
E-mail: ronghumath@aliyun.com

Ya-Ping Fang
Department of Mathematics, Sichuan University, Chengdu, Sichuan, P.R. China
E-mail: ypfang@scu.edu.cn
1 Introduction

Consider the linear equality constrained convex optimization problem:

\[
\min_x \quad F(x), \quad \text{s.t.} \ Ax = b, \tag{1}
\]

where \( F : \mathbb{R}^n \to \mathbb{R} \) is a closed convex but possibly nonsmooth function, \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \). The problem (1) captures a number of important applications arising in various areas, and the following are three concrete examples.

**Example 1.1** The basis pursuit problem (see e.g. [8, 9]):

\[
\min_x \quad \|x\|_1, \quad \text{s.t.} \ Ax = b, \tag{2}
\]

where \( A \in \mathbb{R}^{m \times n} \) with \( m \ll n \), and \( \| \cdot \|_1 \) is the \( \ell_1 \)-norm of \( \mathbb{R}^n \) defined by \( \|x\|_1 = \sum_{i=1}^n |x_i| \).

Algorithms for the basis pursuit problem can be found in [27] and [29].

**Example 1.2** The linearly constrained \( \ell_1 - \ell_2 \) minimization problem [15]:

\[
\min_x \quad \|x\|_1 + \frac{\beta}{2} \|x\|_2^2, \quad \text{s.t.} \ Ax = b, \tag{3}
\]

where \( \beta > 0 \) and \( \| \cdot \|_2 \) is the \( \ell_2 \)-norm of \( \mathbb{R}^n \) defined by \( \|x\|_2^2 = \sum_{i=1}^n x_i^2 \). When \( \beta \) is small enough, a solution of the problem (3) is also a solution of the basis pursuit problem (2). Since the problem (3) has the regularization term \( \frac{\beta}{2} \|x\|_2^2 \), it is less sensitive to noise than the basis pursuit problem (2).

**Example 1.3** The global consensus problem [7]:

\[
\min_{X \in \mathbb{R}^{n \times N}} \quad F(X) = \sum_{i=1}^N f_i(X_i), \quad \text{s.t.} \ X_i = X_j, \quad \forall i, j \in \{1, 2, \cdots, N\},
\]

where \( f_i : \mathbb{R}^n \to \mathbb{R} \) is convex, \( i = 1, 2, \cdots, N \). The global consensus problem is a widely investigated model that has important applications in signal processing [19], routing of wireless sensor networks [18] and optimal consensus of agents [24].

Recall that \((x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m\) is a KKT point of the problem (1) if

\[
\begin{aligned}
-A^T \lambda^* &\in \partial F(x^*), \\
Ax^* - b &= 0,
\end{aligned}
\tag{4}
\]

where \( \partial F \) is the classical subdifferential of \( F \) defined by

\[
\partial F(x) = \{ v \in \mathbb{R}^n | F(y) \geq F(x) + \langle v, y - x \rangle, \quad \forall y \in \mathbb{R}^n \}.
\]

Let \( \Omega \) be the KKT point set of the problem (1). It is well-known that \( x^* \) is a solution of the problem (1) if and only if there exists \( \lambda^* \in \mathbb{R}^m \) such that \((x^*, \lambda^*) \in \Omega \) if and only if

\[
\mathcal{L}(x^*, \lambda) \leq \mathcal{L}(x^*, \lambda^*) \leq \mathcal{L}(x, \lambda^*), \quad \forall (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m,
\]

where \( \mathcal{L}(x, \lambda) = F(x) + \langle \lambda, Ax - b \rangle \).
where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is the Lagrangian function associated with the problem (1) defined by

$$\mathcal{L}(x, \lambda) = F(x) + \langle \lambda, Ax - b \rangle.$$ 

A classical method for solving the problem (1) is the augmented Lagrangian method (ALM) [6]:

$$\begin{align*}
x_{k+1} & \in \arg \min_x \mathcal{L}_\sigma(x, \lambda_k), \\
\lambda_{k+1} & = \lambda_k + \sigma (Ax_{k+1} - b),
\end{align*}$$

(5)

where $\mathcal{L}_\sigma : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ with $\sigma > 0$ is the augment Lagrangian function defined by

$$\mathcal{L}_\sigma(x, \lambda) = F(x) + \langle \lambda, Ax - b \rangle + \frac{\sigma}{2} \|Ax - b\|^2.$$

Obviously, $\mathcal{L}_\sigma(x, \lambda) = \mathcal{L}(x, \lambda) + \frac{\sigma}{2} \|Ax - b\|^2$. In general, since $\mathcal{L}_\sigma$ is not strictly convex, the subproblem may have more than one solutions and be difficult to solve. To overcome this disadvantage, the proximal ALM [10] has been proposed:

$$\begin{align*}
x_{k+1} & = \arg \min_x F(x) + \langle A^T \lambda_k, x \rangle + \frac{\sigma}{2} \|Ax - b\|^2 + \frac{1}{2} \|x - x_k\|^2_P, \\
\lambda_{k+1} & = \lambda_k + \sigma (Ax_{k+1} - b),
\end{align*}$$

(6)

where $\|x\|^2_P = x^T P x$ with a positive semidefinite matrix $P$ and $P + \sigma A^T A$ is positive definite.

In some practical situations, the objective function $F$ has the composite structure: $F(x) = f(x) + g(x)$, where $f$ is a convex but possibly nonsmooth function and $g$ is a convex smooth function. Then the problem (1) becomes the linearly constrained composite convex optimization problem:

$$\min_x f(x) + g(x), \quad \text{s.t. } Ax = b.$$  

(7)

An application of the method (6) to the problem (7) with linearizing the smooth function $g$ leads to the linearized ALM [28]:

$$\begin{align*}
x_{k+1} & \in \arg \min_x f(x) + \langle \nabla g(x_k), x \rangle + \frac{\sigma}{2} \|Ax - b\|^2 + \frac{1}{2} \|x - x_k\|^2_P, \\
\lambda_{k+1} & = \lambda_k + \sigma (Ax_{k+1} - b).
\end{align*}$$

(8)

1.1 Related works

Under the assumption that $F$ is smooth, He and Yuan [13] showed that the iteration-complexity of the method (5) is $O(1/k)$ in terms of the objective residual of the associated $\mathcal{L}(x, \lambda)$. When $F$ is nonsmooth, Gu et al. [12] proved that the method (5) enjoys a worst-case $O(1/k)$ convergence rate in the ergodic sense. A worst-case $O(1/k)$ convergence rate in the non-ergodic sense of the method (6) was shown in [17]. When $g$ has a Lipschitz continuous gradient with constant $L_g$ and $P \succ L_g Id$, Xu [28] proved that the method (8) achieves $O(1/k)$ convergence rate in the ergodic sense. Tran-Dinh and Zhu [26] proposed a modified version of the method (8) and proved that the
objective residual and feasibility violation sequences generated by the method both enjoy $O(1/k)$ non-ergodic convergence rate. Liu et al. [20] investigated the nonergodic convergence rate of inexact augmented Lagrangian method for problem (7).

Generally, naive first-order methods converge slowly. Much effort has been made to accelerate the existing first-order methods in past decades. Nesterov [21] first proposed an accelerated version of the classical gradient method for a smooth convex optimization problem, and proved that the accelerated inertial gradient method enjoys $O(1/k^2)$ convergence rate. Beck and Teboulle [5] proposed an iterative shrinkage-thresholding algorithm for solving the linear inverse problem, which achieves $O(1/k^2)$ convergence rate. The acceleration idea of [21] was further applied in Nesterov [22] to design the accelerated methods for unconstrained convex composite optimization problems. Su et al. [25] first studied accelerated methods from a continuous-time perspective. Since then, some new accelerated inertial methods based on the second-order dynamical system have been proposed for unconstrained optimization problems (see e.g. [13, 4]). For more results on inertial methods for unconstrained optimization problems, we refer the reader to [2, 11, 23].

Meanwhile, inertial methods for linearly constrained optimization problems have also been well-developed. He and Yuan [13] proposed an accelerated inertial ALM for the problem (1) and proved that its convergence rate is $O(1/k^2)$ by using an extrapolation technique similar to [5]. Kang et al. [15] presented an inexact version of the accelerated ALM with inexact calculations of subproblems and showed that the convergence rate remains $O(1/k^2)$ under the assumption that $F$ is strongly convex. Kang et al. [15] further presented an accelerated Bregman method for the linearly constrained $\ell_1-\ell_2$ minimization problem, a convergence rate of $O(1/k^2)$ was proved when the accelerated Bregman method is applied to solve the problem (1). To linearize the augmented term of Bregman method, Huang et al. [14] raised an accelerated linearized Bregman algorithm with $O(1/k^2)$ convergence rate. For the problem (7), Tran-Dinh and Zhu [26] proposed an inertial primal-dual method which enjoys $O(1/k\sqrt{\log k})$ convergence rate. Xu [28] proposed an accelerated version of the linearized ALM (8), named the accelerated linearized augmented Lagrangian method, and it enjoys $O(1/k^2)$ convergence rate under specific parameter settings.

1.2 Inertial primal-dual methods

We first propose Algorithm 1, an inertial version of the proximal ALM (6), for solving the problem (1). Algorithm 1 is also inspired by the following second-order primal-dual dynamical system:

$$
\begin{aligned}
\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) &= -\nabla F(x(t)) - A^T(\lambda(t) + \beta t\dot{\lambda}(t)) - \sigma A^T(Ax(t) - b), \\
\ddot{\lambda}(t) + \frac{\alpha}{t}\dot{\lambda}(t) &= A(x(t) + \beta t\dot{x}(t)) - b,
\end{aligned}
$$

(9)

which was proposed by Zeng et al. [30] for the problem (1). Indeed, taking a fixed step size $h > 0$, setting $t_k = kh, x_k = x(t_k), \lambda_k = \lambda(t_k)$, and discretizing implicitly the right of (9), we get
Algorithm 1: Inertial proximal primal-dual method for the problem (1)

 Initialization: choose $x_0 \in \mathbb{R}^n$, $\lambda_0 \in \mathbb{R}^m$. Set $x_1 = x_0$, $\lambda_1 = \lambda_0$, $M_0 \in S_+(n)$. Choose parameters $s > 0$, $\sigma > 0$, $\alpha \geq 3$.

 for $k = 1, 2, \cdots$ do

 Step 1: Compute $\bar{x}_k = x_k + \frac{\sigma}{\alpha - 1}(x_k - x_{k-1})$, $\bar{\lambda}_k = \lambda_k + \frac{\sigma}{\alpha - 1}(\lambda_k - \lambda_{k-1})$.

 Step 2: Compute $\check{\lambda}_k = \frac{k + \alpha - 1}{\alpha - 1}\bar{\lambda}_k - \frac{k}{\alpha - 1}\lambda_k$, $\rho_k = \sigma + \frac{\beta(k + \alpha - 1)s}{(\alpha - 1)^2}$.

 Set

 $\eta_k = \frac{1}{\rho_k} \left( \frac{(k + \alpha - 1)ks}{(\alpha - 1)^2} Ax_k + \left( \frac{(k + \alpha - 1)s}{\alpha - 1} \right) b \right)$.

 Choose $M_k \in S_+(n)$ and update

 $x_{k+1} \in \arg\min_x F(x) + \frac{1}{2s} ||x - \bar{x}_k||_M^2 + \frac{\rho_k}{2} ||Ax - \eta_k||^2 + (A^T \bar{\lambda}_k, x)$. (10)

 if A stopping condition is satisfied then

 Return $(x_{k+1}, \lambda_{k+1})$

 end

 end

Algorithm 2: Inexact inertial linearized proximal primal-dual method for the problem (1)

 Initialization: choose $x_0 \in \mathbb{R}^n$, $\lambda_0 \in \mathbb{R}^m$. Set $x_1 = x_0$, $\lambda_1 = \lambda_0$, $M_0 \in S_+(n)$, $\epsilon_0 = 0$. Choose parameters $s > 0$, $\sigma > 0$, $\alpha \geq 3$.

 for $k = 1, 2, \cdots$ do

 Step 1: Compute $\bar{x}_k = x_k + \frac{\sigma}{\alpha - 1}(x_k - x_{k-1})$, $\bar{\lambda}_k = \lambda_k + \frac{\sigma}{\alpha - 1}(\lambda_k - \lambda_{k-1})$.

 Step 2: Compute $\check{\lambda}_k = \frac{k + \alpha - 1}{\alpha - 1}\bar{\lambda}_k - \frac{k}{\alpha - 1}\lambda_k$, $\rho_k = \sigma + \frac{\beta(k + \alpha - 1)s}{(\alpha - 1)^2}$.

 Set

 $\eta_k = \frac{1}{\rho_k} \left( \frac{(k + \alpha - 1)ks}{(\alpha - 1)^2} Ax_k + \left( \frac{(k + \alpha - 1)s}{\alpha - 1} \right) b \right)$.

 Choose $M_k \in S_+(n)$, $\epsilon_k \in \mathbb{R}^n$, and update

 $x_{k+1} \in \arg\min_x f(x) + \frac{1}{2s} ||x - \bar{x}_k||^2_M + \frac{\rho_k}{2} ||Ax - \eta_k||^2 + (\nabla g(\tilde{\epsilon}_k) + AT \tilde{\lambda}_k - \epsilon_k, x)$. (10)

 if A stopping condition is satisfied then

 Return $(x_{k+1}, \lambda_{k+1})$

 end

end
inexact inertial proximal primal-dual method (Algorithm 2) for the problem (7). As a comparison to Algorithm 1, we solve the subproblem inexactly by finding an approximate solution instead of an exact solution.

1.3 Outline

The rest of the paper is organized as follows. In Section 2, we investigate the convergence analysis of the proposed methods. In Section 3, we performed numerical experiments. Finally, we give a concluding remark in Section 4.

2 Convergence analysis

In this section we analyze the convergence rate of Algorithm 1 and Algorithm 2. Assuming merely convexity, we show that both of them enjoy $O(1/k^2)$ convergence rates in terms of Lagrangian function. To do so, we first recall some standard notations and results which will be used in the paper. In what follows, we always use $\| \cdot \|$ to denote the $\ell_2$-norm. Let $S_+(n)$ denote the set of all positive semidefinite matrices in $\mathbb{R}^{n \times n}$. For $M \in S_+(n)$, we introduce the semi-norm on $\mathbb{R}^n$: $\| x \|_M = \sqrt{x^T M x}$ for any $x \in \mathbb{R}^n$. This introduces on $S_+(n)$ the following partial ordering: for any $M_1, M_2 \in S_+(n)$,

$$M_1 \succeq M_2 \iff \| x \|_{M_1} \geq \| x \|_{M_2}, \quad \forall x \in \mathbb{R}^n.$$ 

Furthermore, we define for $\kappa > 0$ the set $P_\kappa(n) := \{ M \in S_+(n) : M \succeq \kappa Id \}$, where $Id$ is the identity matrix. For any $x, y \in \mathbb{R}^n$, the following equality hold:

$$\frac{1}{2} \| x \|_M^2 - \frac{1}{2} \| y \|_M^2 = \langle x, M(x - y) \rangle - \frac{1}{2} \| x - y \|_M^2, \quad \forall M \in S_+(n). \quad (11)$$

Now we start to analyze Algorithm 1.

**Lemma 1** Let $\{(x_k, \lambda_k)\}_{k \geq 1}$ be the sequence generated by Algorithm 1. Then $\{(x_k, \lambda_k)\}_{k \geq 1}$ satisfies

$$M_k(x_{k+1} - \bar{x}_k) \in -s(\partial F(x_{k+1}) + \sigma A^T(Ax_{k+1} - b) + A^T(\lambda_{k+1} + \frac{k}{\alpha - 1}(\lambda_{k+1} - \lambda_k))), \quad (12)$$

**Proof** From step 2, we have

$$0 \in \partial F(x_{k+1}) + \frac{1}{s} M_k(x_{k+1} - \bar{x}_k) + \rho_k A^T(Ax_{k+1} - \eta_k) + A^T \hat{\lambda}_k.$$ 

This yields

$$M_k(x_{k+1} - \bar{x}_k) \in -s(\partial F(x_{k+1}) + A^T(\rho_k(Ax_{k+1} - \eta_k) + \hat{\lambda}_k)). \quad (13)$$
It follows from Step 2 and Step 3 that

\[ \rho_k (Ax_{k+1} - \eta_k) + \hat{\lambda}_k = (\sigma + \frac{(k + \alpha - 1)^2}{(\alpha - 1)^2})Ax_{k+1} - \frac{(k + \alpha - 1)k}{\alpha - 1}Ax_k - (\sigma + \frac{(k + \alpha - 1)^2}{\alpha - 1})b + \hat{\lambda}_k \]

\[ = (\sigma + \frac{(k + \alpha - 1)^2}{(\alpha - 1)^2})Ax_{k+1} - \frac{(k + \alpha - 1)s}{\alpha - 1}(Ax_{k+1} - b + \frac{k}{\alpha - 1}A(x_{k+1} - x_k)) + \hat{\lambda}_k \]

\[ = (\sigma + \frac{(k + \alpha - 1)^2}{(\alpha - 1)^2})Ax_{k+1} - \frac{(k + \alpha - 1)}{\alpha - 1}(\lambda_{k+1} - \lambda_k) + \frac{k + \alpha - 1}{\alpha - 1}\hat{\lambda}_k - \frac{k}{\alpha - 1}\lambda_k \]

This together with (13) yields (12).

**Remark 1** By Step 3 of Algorithm 1 and Lemma 1, Algorithm 1 can be viewed as a modified version of the discrete formula (10) with \( F \) being nonsmooth, \( \beta = \frac{1}{\alpha - 1} \), and \( s = h^2 \).

With Lemma 1 in hands, we discuss the convergence rate of Algorithm 1.

**Theorem 1** Suppose that \( F \) is a closed convex function, \( \Omega \neq \emptyset \) and \( M_{k-1} \supseteq M_k \). Let \( \{(x_k, \lambda_k)\}_{k \geq 1} \) be the sequence generated by Algorithm 1. Then for any \( (x^*, \lambda^*) \in \Omega \), the following conclusions hold:

(i) \[ L(x_k, \lambda^*) - L(x^*, \lambda^*) \leq \frac{C(\alpha - 1)^2}{(k + \alpha - 1)ks}, \]

\[ \|Ax_k - b\| \leq \frac{\sqrt{2C(\alpha - 1)}}{\sqrt{(k + \alpha - 1)\sigma}ks}, \]

where

\[ C = \frac{\alpha s}{(\alpha - 1)^2}(\mathcal{L}^*(x_1, \lambda^*) - \mathcal{L}^*(x^*, \lambda^*)) + \frac{1}{2}\|x_1 - x^*\|_{M_0}^2 + \frac{1}{2}\|\lambda_1 - \lambda^*\|^2. \]

(ii) \[ \sum_{k=1}^{+\infty} k^2(\|x_{k+1} - \tilde{x}_k\|_{M_k}^2 + \|\lambda_{k+1} - \tilde{\lambda}_k\|^2) < +\infty. \]

(iii) When \( \alpha = 3 \), we have

\[ \sum_{k=1}^{+\infty} (L(x_k, \lambda^*) - L(x^*, \lambda^*)) < +\infty, \quad \sum_{k=1}^{+\infty} \|Ax_k - b\|^2 < +\infty. \]

When \( \alpha > 3 \), we have

\[ \sum_{k=1}^{+\infty} k(L(x_k, \lambda^*) - L(x^*, \lambda^*)) < +\infty, \quad \sum_{k=1}^{+\infty} k\|Ax_k - b\|^2 < +\infty. \]
Proof Fix \((x^*,\lambda^*)\in\Omega\). Define the energy sequence \(\{E_k\}_{k\geq 1}\) by
\[
E_k = \frac{(k + \alpha - 1)k}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_k,\lambda^*) - \mathcal{L}^\sigma(x^*,\lambda^*)) + \frac{1}{2}\|\hat{x}_k - x^*\|_{M_{k-1}}^2 + \frac{1}{2}\|\hat{\lambda}_k - \lambda^*\|^2,
\]
where
\[
\hat{x}_k = \frac{k + \alpha - 1}{\alpha - 1}x_k - \frac{k}{\alpha - 1}x_k.
\]
By computation,
\[
\hat{x}_{k+1} = \frac{k + \alpha - 1}{\alpha - 1}(x_{k+1} + \frac{k}{k + \alpha} (x_{k+1} - x_k)) - \frac{k + 1}{\alpha - 1}x_{k+1} = \frac{k + \alpha - 1}{\alpha - 1}x_{k+1} - \frac{k}{\alpha - 1}x_k
\]
\[
= \frac{k + \alpha - 1}{\alpha - 1}(x_{k+1} - x_k) + \frac{k + \alpha - 1}{\alpha - 1}x_k - \frac{k}{\alpha - 1}x_k = \hat{x}_k + \frac{k + \alpha - 1}{\alpha - 1}(x_{k+1} - x_k)
\]
and
\[
\hat{x}_{k+1} - x^* = x_{k+1} - x^* + \frac{k}{\alpha - 1}(x_{k+1} - x_k).
\]
Similarly, we have
\[
\hat{\lambda}_{k+1} = \hat{\lambda}_k + \frac{k + \alpha - 1}{\alpha - 1} (\lambda_{k+1} - \hat{\lambda}_k)
\]
and
\[
\hat{\lambda}_{k+1} - \lambda^* = \lambda_{k+1} - \lambda^* + \frac{k}{\alpha - 1} (\lambda_{k+1} - \lambda_k).
\]
By the definition of \(\mathcal{L}^\sigma(x,\lambda)\), we get \(\partial_s \mathcal{L}^\sigma(x,\lambda) = \partial F(x) + AT\lambda + \sigma A^T (Ax - b)\). Combining this and equality (19), we can rewrite (12) as
\[
M_k(x_{k+1} - \bar{x}_k)
\]
\[
\in - s(\partial F(x_{k+1}) + \sigma A^T (Ax_{k+1} - b) + AT\lambda^* + AT(\lambda_{k+1} - \lambda^*) + \frac{k}{\alpha - 1} (\lambda_{k+1} - \lambda_k))
\]
\[
= - s \partial_s \mathcal{L}^\sigma(x_{k+1},\lambda^*) - sA^T(\hat{\lambda}_{k+1} - \lambda^*),
\]
which implies
\[
\xi_k := - \frac{1}{s} M_k(x_{k+1} - \bar{x}_k) - A^T(\hat{\lambda}_{k+1} - \lambda^*) \in \partial_{\lambda} \mathcal{L}^\sigma(x_{k+1},\lambda^*).
\]
Since \(M_{k-1} \succ M_k\), it follows from (11) and (10) that
\[
\frac{1}{2}\|\hat{x}_{k+1} - x^*\|_{M_k}^2 - \frac{1}{2}\|\hat{x}_k - x^*\|_{M_{k-1}}^2
\]
\[
= \frac{1}{2}\|\hat{x}_{k+1} - x^*\|_{M_k}^2 - \frac{1}{2}\|\hat{x}_k - x^*\|_{M_k}^2 - \frac{1}{2}\|\hat{x}_k - x^*\|_{M_{k-1}}^2 - \frac{1}{2}\|\hat{x}_k - x^*\|_{M_{k-1}-M_k}
\]
\[
\leq \langle \hat{x}_{k+1} - x^*, M_k(\hat{x}_{k+1} - \bar{x}_k) \rangle - \frac{1}{2}\|\hat{x}_{k+1} - \bar{x}_k\|_{M_k}^2
\]
\[
= \frac{k + \alpha - 1}{\alpha - 1} \langle \hat{x}_{k+1} - x^*, M_k(x_{k+1} - \bar{x}_k) \rangle - \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2} \|x_{k+1} - \bar{x}_k\|_{M_k}^2
\]
\[
= - \frac{(k + \alpha - 1)s}{\alpha - 1} (\langle \hat{x}_{k+1} - x^*, \xi_k \rangle + \langle \hat{x}_{k+1} - x^*, A^T(\hat{\lambda}_{k+1} - \lambda^*) \rangle)
\]
\[
- \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2} \|x_{k+1} - \bar{x}_k\|_{M_k}^2.
\]
Since $\mathcal{L}^\sigma(x, \lambda^*)$ is a convex function with respect to $x$, from (17) and (20) we get
\[
\langle \hat{x}_{k+1} - x^*, \xi_k \rangle = \langle x_{k+1} - x^*, \xi_k \rangle + \frac{k}{\alpha - 1} \langle x_{k+1} - \bar{x}_k, \xi_k \rangle \\
\geq \mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*) + \frac{k}{\alpha - 1} (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x_k, \lambda^*)).
\]
(22)
Combining (21) and (22) together, we have
\[
\frac{1}{2} \| \hat{x}_{k+1} - x^* \|^2_{M_k} - \frac{1}{2} \| \hat{x}_k - x^* \|^2_{M_{k-1}} \\
\leq - \frac{(k + \alpha - 1)s}{\alpha - 1} (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) - \frac{1}{2} \| L_{x_{k+1}} - L_{x_k} \|_{\lambda_k} \\
- \frac{(k + \alpha - 1)s}{\alpha - 1} \langle \hat{x}_{k+1} - x^*, A^T (\bar{\lambda}_k - \lambda^*) \rangle - \frac{1}{2} \| x_{k+1} - \bar{x}_k \|^2_{M_k}.
\]
(23)
Since $Ax^* = b$, it follows from Step 3 of Algorithm I and (17) that
\[
\lambda_{k+1} - \bar{\lambda}_k = s(Ax_{k+1} - Ax^* + \frac{k}{\alpha - 1} (Ax_{k+1} - x_k)) = sA(\hat{x}_{k+1} - x^*).
\]
This together with (11) and (18) yields
\[
\frac{1}{2} \| \hat{\lambda}_{k+1} - \lambda^* \|^2 - \frac{1}{2} \| \hat{\lambda}_k - \lambda^* \|^2 = (\hat{\lambda}_{k+1} - \lambda^*, \hat{\lambda}_k - \lambda^*) - \frac{1}{2} \| \hat{\lambda}_{k+1} - \hat{\lambda}_k \|^2 \\
= \frac{k + \alpha - 1}{\alpha - 1} (\hat{\lambda}_{k+1} - \lambda^*, \hat{\lambda}_k - \lambda^*) - \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2} \| \hat{\lambda}_{k+1} - \hat{\lambda}_k \|^2
\]
(24)
By computation, we get
\[
\frac{(k + \alpha)(k + 1)s}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) - \frac{(k + \alpha - 1)ks}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_k, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) \\
= \frac{(2k + \alpha)s}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) + \frac{(k + \alpha - 1)ks}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x_k, \lambda^*)).
\]
(25)
It follows (23)-(24) that
\[
\mathcal{E}_{k+1} - \mathcal{E}_k \\
= \frac{(k + \alpha)(k + 1)s}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) - \frac{(k + \alpha - 1)ks}{(\alpha - 1)^2} (\mathcal{L}^\sigma(x_k, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) \\
+ \frac{1}{2} \| \hat{x}_{k+1} - x^* \|^2_{M_k} - \frac{1}{2} \| \hat{x}_k - x^* \|^2_{M_{k-1}} + \frac{1}{2} \| \hat{\lambda}_{k+1} - \lambda^* \|^2 - \frac{1}{2} \| \hat{\lambda}_k - \lambda^* \|^2 \\
\leq \frac{(3 - \alpha)k + \alpha - (\alpha - 1)^2}{(\alpha - 1)^2} s(\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) \\
- \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2} (\| x_{k+1} - \bar{x}_k \|^2_{M_k} + \| \lambda_{k+1} - \bar{\lambda}_k \|^2).
\]
(26)
As $\alpha \geq 3$ and $(x^*, \lambda^*) \in \Omega$, we have $(3 - \alpha)k + \alpha - (\alpha - 1)^2 < 0$ and
\[
\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*) = \mathcal{L}(\hat{x}_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*) + \frac{\sigma}{2} \| Ax_{k+1} - b \|^2 \geq 0.
\]
(27)
This together with (26) implies
\[ E_{k+1} \leq E_k. \] (28)

Then from (14) and (27), we have
\[ \frac{1}{2}(k+\alpha-3) \leq \frac{1}{2} + \frac{1}{2} \left( \alpha - 1 \right)^2 \left( \left\| x_{k+1} - x^* \right\|^2_{M_0} + \frac{1}{2} \left\| \lambda_1 - \lambda^* \right\|^2 \right) \leq \frac{1}{2}(k+\alpha-1) \]
which yields
\[ \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \leq \frac{(\alpha - 1)^2}{(k+\alpha - 1)k} \left( \frac{\alpha s}{(\alpha - 1)^2} \left( \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \right) + \frac{1}{2} \left\| x_1 - x^* \right\|^2_{M_0} + \frac{1}{2} \left\| \lambda_1 - \lambda^* \right\|^2 \right), \]
and
\[ \left\| Ax_k - b \right\| \leq \sqrt{\frac{2(\alpha - 1)^2}{(k+\alpha - 1)\sigma k} \left( \frac{\alpha s}{(\alpha - 1)^2} \left( \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \right) + \frac{1}{2} \left\| x_1 - x^* \right\|^2_{M_0} + \frac{1}{2} \left\| \lambda_1 - \lambda^* \right\|^2 \right). \]

The proof of (i) is complete.

Next we prove (ii) and (iii). From (28), we get
\[ \sum_{k=1}^{n} \left( \frac{3-\alpha}{\alpha - 1} \right) \leq \sum_{k=1}^{n} \frac{(k+\alpha - 3)k}{(\alpha - 1)^2} \left( \left\| x_{k+1} - x^* \right\|^2_{M_k} + \left\| \lambda_{k+1} - \lambda^* \right\|^2 \right) \]
which is just (ii), and (iii) follows directly from (31).

Clearly, (71) is just (ii), and (iii) follows directly from (31).

Remark 2 In the case that \( \alpha \geq 3 \), it was shown in [30, Theorem 3.1] that \( \mathcal{L}(x(t), \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O(1/t^2) \) and \( \left\| Ax(t) - b \right\| = O(1/t) \), where \( x(t) \) is the trajectory of the system (9). Theorem 1 shows that Algorithm 1 preserves the fast convergence properties of the system (9).
Remark 3 From (i), we obtain that \( L(x_k, \lambda^*) - L(x^*, \lambda^*) = O(1/k) \) and \( \|Ax_k - b\| = O(1/k) \). Since 
\[
|f(x_k) - f(x^*)| \leq L(x_k, \lambda^*) - L(x^*, \lambda^*) + \|\lambda^*\|\|Ax_k - b\|,
\]
we obtain the \( O(1/k) \) convergence rate of the objective function. From (ii), we also get that 
\[
\|\lambda_{k+1} - \bar{\lambda}_k\| + \|x_{k+1} - \bar{x}_k\|M_k = o\left(\frac{1}{k}\right).
\]

To investigate the convergence of Algorithm 2, we need the following assumption.

**Assumption (H):** \( \Omega \neq \emptyset, f \) is a closed convex function, and \( g \) is a convex smooth function and has a Lipschitz continuous gradient with constant \( L_g > 0 \), i.e.,
\[
\|\nabla g(x) - \nabla g(y)\| \leq L_g\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,
\]
equivalently,
\[
g(x) \leq g(y) + \langle \nabla g(y), x - y \rangle + \frac{L_g}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \tag{32}
\]

**Lemma 2** Let \( \{(x_k, \lambda_k)\}_{k \geq 1} \) the sequence generated by Algorithm 2. Then \( \{(x_k, \lambda_k)\}_{k \geq 1} \) satisfies
\[
M_k(x_{k+1} - \bar{x}_k) \in -s(\partial f(x_{k+1}) + \nabla g(\bar{x}_k) + \sigma A^T(Ax_{k+1} - b) + A^T(\lambda_{k+1} + \frac{k}{n-1}(\lambda_{k+1} - \lambda_k)) - \epsilon_k).
\]

**Proof** From Step 2 of Algorithm 2, we have
\[
0 \in \partial f(x_{k+1}) + \nabla g(\bar{x}_k) + \frac{1}{s}M_k(x_{k+1} - \bar{x}_k) + \rho_k A^T(Ax_{k+1} - \eta_k) + A^T\lambda_k - \epsilon_k,
\]
which yields
\[
M_k(x_{k+1} - \bar{x}_k) \in -s(\partial f(x_{k+1}) + \nabla g(\bar{x}_k) + A^T(\rho_k (Ax_{k+1} - \eta_k) + \lambda_k) - \epsilon_k).
\]
The rest of the proof is similar as the one of Lemma 1 and so we omit it.

To analyze the convergence of Algorithm 2 we need the following discrete version of the Gronwall-Bellman lemma.

**Lemma 3** [2, Lemma 5.14] Let \( \{a_k\}_{k \geq 1} \) and \( \{b_k\}_{k \geq 1} \) be two nonnegative sequences such that 
\[
\sum_{k=1}^{+\infty} b_k < +\infty \text{ and } a_k \leq c^2 + \sum_{j=1}^{k} b_j a_j
\]
for all \( k \in \mathbb{N} \) where \( c \geq 0 \). Then 
\[
a_k \leq c + \sum_{j=1}^{+\infty} b_j
\]
for all \( k \in \mathbb{N} \).
Theorem 2 Assume that Assumption (H) holds, $M_k \in \mathcal{P}_{sL_g}(n)$ with $M_{k-1} \succ M_k$ for all $k \geq 0$, and
\[ \sum_{k=1}^{+\infty} k\|e_k\| < +\infty. \]

Let $\{(x_k, \lambda_k)\}_{k \geq 1}$ be the sequence generated by Algorithm 2. Then for any $(x^*, \lambda^*) \in \Omega$, the following conclusions hold:

(i) \[
\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \leq \frac{C(\alpha - 1)^2}{(k + \alpha - 1)ks},
\]
where \[
\|Ax_k - b\| \leq \frac{\sqrt{2C(\alpha - 1)}}{(k + \alpha - 1)\sigma ks},
\]
with \[
C = \mathcal{E}_1 + \frac{s}{\alpha - 1} \left( \frac{2\mathcal{E}_1}{sL_g} + \frac{2}{(\alpha - 1)L_g} \sum_{j=1}^{+\infty} (j + \alpha - 1)\|e_j\| \right) \times \sum_{j=1}^{+\infty} (j + \alpha - 1)\|e_j\|,
\]
and \[
\mathcal{E}_1 = \frac{\alpha s}{(\alpha - 1)^2}(\mathcal{L}^*(x_1, \lambda^*) - \mathcal{L}^*(x^*, \lambda^*)) + \frac{1}{2}\|x_1 - x^*\|_M^2 + \frac{1}{2}\|\lambda_1 - \lambda^*\|^2.
\]

(ii) When $\alpha = 3$, we have \[
\sum_{k=1}^{+\infty} (\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty, \quad \sum_{k=1}^{+\infty} \|Ax_k - b\|^2 < +\infty.
\]

When $\alpha > 3$, we have \[
\sum_{k=1}^{+\infty} k(\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty, \quad \sum_{k=1}^{+\infty} k\|Ax_k - b\|^2 < +\infty.
\]

Proof Fix $(x^*, \lambda^*) \in \Omega$. Define the energy sequence $\{\mathcal{E}_k^j\}_{k \geq 1}$ by
\[
\mathcal{E}_k^j = \mathcal{E}_k - \sum_{j=1}^{k} \frac{(j + \alpha - 2)s}{\alpha - 1}(\bar{x}_j - x^*, \epsilon_{j-1}),
\]
where $\mathcal{E}_k$ is defined in (14) and $\bar{x}_k$ is defined in (15).

By same arguments as in the proof of Theorem 1 we get
\[
\bar{x}_{k+1} - x_k = \frac{k + \alpha - 1}{\alpha - 1}(x_{k+1} - x_k), \tag{35}
\]
\[
\bar{x}_{k+1} - x^* = x_{k+1} - x^* + \frac{k}{\alpha - 1}(x_{k+1} - x_k), \tag{36}
\]
\[
\bar{\lambda}_{k+1} - \bar{\lambda}_k = \frac{k + \alpha - 1}{\alpha - 1}(\lambda_{k+1} - \lambda_k), \tag{37}
\]
\[
\bar{\lambda}_{k+1} - \lambda^* = \lambda_{k+1} - \lambda^* + \frac{k}{\alpha - 1}(\lambda_{k+1} - \lambda_k). \tag{38}
\]
For notation simplicity, we denote
\[ \mathcal{L}'(x) = f(x) + \langle \lambda^*, Ax - b \rangle + \frac{\sigma}{2} \|Ax - b\|^2. \] (39)

Then \( \mathcal{L}' \) is a convex function, \( \partial \mathcal{L}'(x) = \partial f(x) + A^T \lambda^* + \sigma A^T (Ax - b) \), and
\[ \mathcal{L}'(x, \lambda^*) = \mathcal{L}'(x) + g(x). \] (40)

Rewriting (33), we get
\[ M_k(x_{k+1} - \bar{x}_k) \in -s(\partial f(x_{k+1}) + \sigma A^T (Ax_{k+1} - b) + A^T \lambda^*) - s \nabla g(\bar{x}_k) \]
\[ -sA^T (\lambda_{k+1} - \lambda^* + \frac{k}{\alpha - 1} (\lambda_{k+1} - \lambda_k)) + s \epsilon_k \]
\[ = -s \partial \mathcal{L}'(x_{k+1}) - s \nabla g(\bar{x}_k) - sA^T (\lambda_{k+1} - \lambda^*) + s \epsilon_k, \]

which yields
\[ \xi_k := -\frac{1}{s} M_k(x_{k+1} - \bar{x}_k) - \nabla g(\bar{x}_k) - A^T (\lambda_{k+1} - \lambda^*) + \epsilon_k \in \partial \mathcal{L}'(x_{k+1}). \] (41)

Since \( M_{k-1} \succ M_k \), it follows from (11), (35) and (36) that
\[
\frac{1}{2} \|\hat{x}_{k+1} - x^*\|^2_{M_k} - \frac{1}{2} \|\bar{x}_k - x^*\|^2_{M_{k-1}} \\
\leq - \frac{(k + \alpha - 1)s}{\alpha - 1} \langle \hat{x}_{k+1} - x^*, \xi_k \rangle + \langle \bar{x}_{k+1} - x^*, A^T (\lambda_{k+1} - \lambda^*) \rangle \\
+ \langle \hat{x}_{k+1} - x^*, \nabla g(\bar{x}_k) \rangle - \langle \hat{x}_{k+1} - x^*, \epsilon_k \rangle \rangle - \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2} \|x_{k+1} - \bar{x}_k\|^2_{M_k}. \] (42)

From (36) and (11), we have
\[ \langle \hat{x}_{k+1} - x^*, \xi_k \rangle = \langle x_{k+1} - x^*, \xi_k \rangle + \frac{k}{\alpha - 1} \langle x_{k+1} - x_k, \xi_k \rangle \]
\[ \geq \mathcal{L}'(x_{k+1}) - \mathcal{L}'(x^*) + \frac{k}{\alpha - 1} (\mathcal{L}'(x_{k+1}) - \mathcal{L}'(x_k)), \] (43)

where the inequality follows from the convexity of \( \mathcal{L}' \). From (42) we have
\[ g(x_{k+1}) \leq g(\bar{x}_k) + \langle \nabla g(\bar{x}_k), x_{k+1} - \bar{x}_k \rangle + \frac{L_g}{2} \|x_{k+1} - \bar{x}_k\|^2. \] (44)

By the convexity of \( g \), we have
\[ \langle \nabla g(\bar{x}_k), x_{k+1} - \bar{x}_k \rangle = \langle \nabla g(\bar{x}_k), x_{k+1} - x^* \rangle + \langle \nabla g(\bar{x}_k), x^* - \bar{x}_k \rangle \]
\[ \leq \langle \nabla g(\bar{x}_k), x_{k+1} - x^* \rangle + g(x^*) - g(\bar{x}_k) \] (45)

and
\[ \langle \nabla g(\bar{x}_k), x_{k+1} - \bar{x}_k \rangle = \langle \nabla g(\bar{x}_k), x_{k+1} - x_k \rangle + \langle \nabla g(\bar{x}_k), x_k - \bar{x}_k \rangle \]
\[ \leq \langle \nabla g(\bar{x}_k), x_{k+1} - x_k \rangle + g(x_k) - g(\bar{x}_k). \] (46)
It follows from (43)–(46) that
\[\langle \nabla g(\bar{x}_k), x_{k+1} - x^* \rangle \geq g(x_{k+1}) - g(x^*) - \frac{L_g}{2}\|x_{k+1} - \bar{x}_k\|^2,\]
and
\[\langle \nabla g(\bar{x}_k), x_{k+1} - x_k \rangle \geq g(x_{k+1}) - g(x_k) - \frac{L_g}{2}\|x_{k+1} - \bar{x}_k\|^2.\]
This together with (36) yields
\[\langle \bar{x}_{k+1} - x^*, \nabla g(\bar{x}_k) \rangle = \langle \nabla g(\bar{x}_k), x_{k+1} - x^* \rangle + \frac{k}{\alpha - 1}\langle \nabla g(\bar{x}_k), x_{k+1} - x_k \rangle \]
\[\geq g(x_{k+1}) - g(x^*) + \frac{k}{\alpha - 1}(g(x_{k+1}) - g(x_k)) \]
\[- \frac{(k + \alpha - 1)L_g}{2(\alpha - 1)}\|x_{k+1} - \bar{x}_k\|^2. \tag{47}\]

It follows from (42), (43) and (44) that
\[
\frac{1}{2}\|\bar{x}_{k+1} - x^*\|^2_{M_k} - \frac{1}{2}\|\bar{x}_k - x^*\|^2_{M_k-1} \\
\leq -\frac{(k + \alpha - 1)s}{\alpha - 1}(\mathcal{L}^f(x_{k+1}) + g(x_{k+1}) - (\mathcal{L}^f(x^*) + g(x^*))) \\
- \frac{(k + \alpha - 1)k_s}{(\alpha - 1)^2}(\mathcal{L}^f(x_{k+1}) + g(x_{k+1}) - (\mathcal{L}^f(x_k) + g(x_k))) \\
- \frac{(k + \alpha - 1)s}{\alpha - 1}(\bar{x}_{k+1} - x^*, A^T(\tilde{\lambda}_{k+1} - \lambda^*)) + \frac{(k + \alpha - 1)s}{\alpha - 1}(\bar{x}_{k+1} - x^*, \epsilon_k) \tag{48}\]
\[- \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2}\|x_{k+1} - \bar{x}_k\|^2_{M_k-\mathcal{P}_{sL_g}L_{\mathcal{P}_{sL_g}I}d} \\
\leq -\frac{(k + \alpha - 1)s}{\alpha - 1}(\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) \\
- \frac{(k + \alpha - 1)k_s}{(\alpha - 1)^2}(\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x_k, \lambda^*)) \\
- \frac{(k + \alpha - 1)s}{\alpha - 1}(\bar{x}_{k+1} - x^*, A^T(\tilde{\lambda}_{k+1} - \lambda^*)) + \frac{(k + \alpha - 1)s}{\alpha - 1}(\bar{x}_{k+1} - x^*, \epsilon_k),
\]
where the second inequality follows from the assumption $M_k \in \mathcal{P}_{sL_g}L_{\mathcal{P}_{sL_g}I}d$.

It follows from (48) and (24)–(25) that
\[
\mathcal{E}_{k+1} - \mathcal{E}_k = \mathcal{E}_{k+1} - \mathcal{E}_k = \frac{(k + \alpha - 1)s}{\alpha - 1}\langle \bar{x}_{k+1} - x^*, \epsilon_k \rangle \\
\leq \frac{(3 - \alpha)(k + \alpha - (\alpha - 1)^2s(\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) \\
- \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2}\|\tilde{\lambda}_{k+1} - \bar{\lambda}_k\|^2. \tag{49}\]
Since $\alpha \geq 3$, and $(x^*, \lambda^*) \in \Omega$, $\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*) \geq 0$, we get
\[\mathcal{E}_{k+1} \leq \mathcal{E}_k,\]
which implies
\[ E_k \leq E_1 + \sum_{j=1}^{k} \frac{(j + \alpha - 2)s}{\alpha - 1} \langle \hat{x}_j - x^*, \epsilon_{j-1} \rangle. \] (50)

This together with the definition of $E_k$ implies
\[ \| \hat{x}_k - x^* \|^2_{M_{k-1}} \leq 2E_1 + \frac{2s}{\alpha - 1} \sum_{j=1}^{k} (j + \alpha - 2) \| \hat{x}_j - x^* \| \cdot \| \epsilon_{j-1} \|. \]

Then it follows from $\| \hat{x}_k - x^* \|^2_{M_{k-1}} \geq sL_g \| \hat{x}_k - x^* \|^2$ that
\[ \| \hat{x}_k - x^* \|^2 \leq \frac{2E_1}{sL_g} + \frac{2}{L_g(\alpha - 1)} \sum_{j=1}^{k} (j + \alpha - 2) \| \hat{x}_j - x^* \| \cdot \| \epsilon_{j-1} \|. \] (51)

Since $\sum_{j=1}^{+\infty} j \| \epsilon_j \| < +\infty$, we have
\[ \sum_{j=1}^{+\infty} (j + \alpha - 1) \| \epsilon_j \| < +\infty. \]

Applying Lemma 3 with $a_k = \| \hat{x}_k - x^* \|$ to (51), we obtain
\[ \| \hat{x}_k - x^* \| \leq \sqrt{\frac{2E_1}{sL_g} + \frac{2}{(\alpha - 1)L_g} \sum_{j=1}^{+\infty} (j + \alpha - 1) \| \epsilon_j \|} < +\infty, \quad \forall k \geq 1. \] (52)

This together with (50) yields
\[ E_k \leq E_1 + \frac{s}{\alpha - 1} \left( \sqrt{\frac{2E_1}{sL_g} + \frac{2}{(\alpha - 1)L_g} \sum_{j=1}^{+\infty} (j + \alpha - 1) \| \epsilon_j \|} \right) \times \sum_{j=1}^{+\infty} (j + \alpha - 1) \| \epsilon_j \|, \] (53)

for any $k \geq 1$. By the definition of $E_k$, we have
\[ \frac{(k + \alpha - 1)k}{(\alpha - 1)^2} \left( \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) + \frac{\sigma}{2} \| Ax_k - b \|^2 \right) \leq E_k, \]

which together with (53) yields (i).

Summing the inequality (49) over $k = 1, 2, \cdots$, and using the Cauchy-Schwarz inequality, we obtain
\[
\sum_{k=1}^{+\infty} \frac{(3 - \alpha)k + \alpha - (\alpha - 1)^2}{(\alpha - 1)^2} s(\mathcal{L}^2(x_{k+1}, \lambda^*) - \mathcal{L}^2(x^*, \lambda^*))
+ \sum_{k=1}^{+\infty} \frac{(k + \alpha - 1)^2}{2(\alpha - 1)^2} \| \lambda_{k+1} - \lambda_k \|^2
\leq E_1 + \frac{s}{\alpha - 1} \sup_{k \geq 1} \| \hat{x}_k - x^* \| \sum_{j=1}^{+\infty} (j + \alpha - 1) \| \epsilon_j \|
\leq +\infty,
\]

where the second inequality follows from the definition of $E_k^*$ and (52). The rest of the proof is similar as the one of Theorem 1 and so omit it.
Remark 4 With the augmented term $\|Ax - \eta_k\|^2$, it is difficult to find an exact solution of the subproblem of Algorithms 2. From Theorem 2 when the perturbed sequence $\{\epsilon_k\}_{k \geq 1}$ satisfying the condition $\sum_{k=1}^{\infty} k\|\epsilon_k\| < +\infty$ used in [2,3], Algorithms 2 can achieve the fast convergence rate. So we can find an inexact approximate solution to solve subproblem. From the proof process of Theorem 2 we can obtain that Algorithm 1 also converges with small perturbation under assumption $M_k \in \mathcal{P}_\kappa(n)$ for some $\kappa > 0$. The numerical experiments in section 3 show the effectiveness of the inexact algorithms.

3 Numerical experiments

3.1 The quadratic programming problem

In this subsection, we test the algorithms on the nonnegative linearly constrained quadratic programming problem (NLCP):

$$
\min \frac{1}{2} x^T Q x + q^T x, \quad \text{s.t.} \ Ax = b, x \geq 0,
$$

where $q \in \mathbb{R}^n$, $Q \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Here, we compare Algorithm 2 with the accelerated linearized augmented Lagrangian method (AALM [28, Algorithm 1]), which enjoys $O(1/k^2)$ convergence rate with adaptive parameters.

Set $m = 100$ and $n = 500$. Let $q$ be generated by standard Gaussian distribution, $b$ be generated by uniform distribution, $Q = 2H^T H$ with $H \in \mathbb{R}^{n \times n}$ generated by standard Gaussian distribution. Then $Q$ may not be positive definite. The optimal value $F(x^*)$ is obtained by Matlab function `quadprog` with tolerance $10^{-15}$. In this case, $F(x) = f(x) + g(x)$ with $f(x) = \mathcal{I}_{y \geq 0}(x)$, $g(x) = \frac{1}{2} x^T Q x + q^T x$, where $\mathcal{I}_{y \geq 0}$ is the indicator function of the set $\{y | y \geq 0\}$, i.e.,

$$
\mathcal{I}_{y \geq 0}(x) = \begin{cases} 0, & x \geq 0, \\ +\infty, & \text{otherwise}. \end{cases}
$$

Set the parameters of Algorithm 2 as: $\alpha = 20$, $s = \|Q\|$, $\sigma = 1$, $M_k = s \|Q\| Id$. Set the parameters of AALM [28, Algorithm 1]) with adaptive parameters, in which $\alpha_k = \frac{\alpha}{k^2}$, $\beta_k = \gamma_k = \|Q\|$, $P_k = \frac{2\|Q\|}{k} Id$. Subproblems for both algorithms are solved by interior-point algorithms to a tolerance $\text{subtol}$. Figure 1 describes the distance of optimal value $|F(x_k) - F(x^*)|$ and violation of feasibility $\|Ax_k - b\|$ given different methods for the first 500 iterations. As shown in Figure 1, Algorithm 2 performs better and more stable than AALM under different $\text{subtol}$. 


3.2 The basis pursuit problem

Consider the following basis pursuit problem:

\[
\min_x \| x \|_1, \quad \text{s.t. } Ax = b,
\]

where \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \) and \( m \leq n \). Let \( A \) be generated by standard Gaussian distribution. The number of nonzero elements of the original solution \( x^* \) is fixed at \( 0.1 \times n \), and the nonzero elements are selected randomly in \([-2, 2]\). Set \( b = Ax^* \). We compare Algorithm 1 with the inexact augmented Lagrangian method (IAL [20, Algorithm 1]). Here, subproblems for both algorithms are solved by fast iterative shrinkage-thresholding algorithm (FISTA [5, 20, Algorithm 2]), and the stopping condition of the FISTA is when

\[
\frac{\| x_k - x_{k-1} \|^2}{\max\{\|x_{k-1}\|, 1\}} \leq 10^{-4}
\]

is satisfied or the number of iterations exceeds 100. In each test, we calculate the residual error \( \| Ax_k - b \| \) (Res) and the relative error of the solution \( \frac{\| x_k - x^* \|}{\| x^* \|} \) (Rel) with the stopping condition \( \text{Res} + \text{Rel} \leq 10^{-5} \). Set the parameters of Algorithm 1 as \( \alpha = 20, s = 0.1, \sigma = 0.05, M_k = 0.01Id \), and the parameter of IAL as \( \beta = 0.1 \). Let \text{Init} and \text{Time} denote the number of iterations, and the CPU time in seconds, respectively. In Table 1 we report the results for the basis pursuit problem with different dimensions. From Table 1 we observe that when the subproblem is solved with low accuracy, Algorithm 1 is faster than IAL in terms of the number of iterations and the cpu time.
3.3 The linearly constrained $\ell_1 - \ell_2$ minimization problem

Consider the following problem:

$$
\min_x \|x\|_1 + \frac{\beta}{2}\|x\|_2^2 \quad \text{s.t.} \ Ax = b,
$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Let $m = 2500, n = 5000$, and $A$ be generated by standard Gaussian distribution. Suppose that the original solution (signal) $x^* \in \mathbb{R}^n$ has only 250 non-zero elements which are generated by the Gaussian distribution $\mathcal{N}(0, 4)$ in the interval $[-2, 2]$ and that the noise $\omega$ is selected randomly with $\|\omega\| = 10^{-4}$.

$$
b = Ax^* + \omega.
$$

Set parameters for Algorithm 2 (AL2) with $\alpha = 200, s = 1, \sigma = 0.05, M_k = s\beta$, and the parameters of IAALM ([16, Algorithm 1]) with $\gamma = 1$. Subproblems are solved by FISTA and the stopping condition is when

$$
\frac{\|x_k - x_{k-1}\|^2}{\max\{\|x_k - x_{k-1}\|, 1\}} \leq 10^{-8}
$$

is satisfied or the number of iterations exceeds 100. We terminate all the methods when $\|Ax_k - b\| \leq 5 \times 10^{-4}$. In each test, we calculate the residual error $\text{res} = \|Ax - b\|$, the relative error $\text{rel} = \frac{\|x - x^*\|}{\|x^*\|}$ and the signal-to-noise ratio

$$
\text{SNR} = \log_{10} \frac{\|x^* - \text{mean}(x^*)\|^2}{\|x - x^*\|^2},
$$

where $x$ is the recovery signal.

In Table 2 we present the numerical results of Algorithm 2 and IAALM for various $\beta$. Based on the $\text{Rel}$ and $\text{SNR}$, it is seen that the sparse original signal is well restored when $\beta \leq 1$. This is also shown in Figure 2.
Table 2: Comparison of the Algorithm 2 with IAALM for various $\beta$

| $\beta$ | 0.01 | 0.05 | 0.1 | 0.5 | 1   | 1.5 |
|---------|------|------|-----|-----|-----|-----|
| Init    | AL2  | 27   | 27  | 26  | 26  | 33  | 58  |
|         | IAALM| 54   | 53  | 50  | 50  | 52  | 100+|
| Time    | AL2  | 1.61e+2 | 1.74e+2 | 1.60e+2 | 1.58e+2 | 2.08e+2 | 3.76e+2 |
|         | IAALM| 2.54e+2 | 2.59e+2 | 2.45e+2 | 2.44e+2 | 2.66e+2 | 6.92e+2 |
| Res     | AL2  | 1.5e-4  | 1.3e-4  | 1.7e-4  | 3.2e-4  | 1.5e-4  | 4.6e-4  |
|         | IAALM| 4.3e-4  | 3.2e-4  | 4.8e-4  | 4.0e-4  | 4.6e-4  | 1.0e-2  |
| Rel     | AL2  | 1.5e-7  | 1.2e-7  | 1.6e-7  | 3.8e-7  | 1.3e-7  | 1.2e-1  |
|         | IAALM| 4.5e-7  | 3.2e-7  | 5.0e-7  | 4.4e-7  | 6.3e-7  | 1.2e-1  |
| SNR     | AL2  | 1.37e+2 | 1.38e+2 | 1.36e+2 | 1.28e+2 | 1.37e+2 | 1.85e+1 |
|         | IAALM| 1.27e+2 | 1.30e+2 | 1.26e+2 | 1.27e+2 | 1.23e+2 | 1.86e+1 |

Fig. 2: Original sparse signal and the final estimated solution of Algorithm 2 with $\beta = 0.5, 1.5$.

4 Conclusion

In this paper, we propose two inertial primal-dual methods for solving linear equality constrained convex optimization problems. Assuming merely convexity, we show the inertial primal-dual methods own fast convergence rates even if the subproblem is solved inexactly. The numerical results
demonstrate the validity and superior performance of our methods over some some existing methods.

**Declarations**

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

**References**

1. Apidopoulos, V., Aujol, J.F., Dossal, C.: Convergence rate of inertial Forward-Backward algorithm beyond Nesterov’s rule. Math. Program. **180**(1), 137-156 (2020).
2. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**(1-2), 123-175 (2018).
3. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**(3), 1824-1834 (2016).
4. Attouch, H., Peypouquet, J., Redont, P.: A dynamical approach to an inertial forward-backward algorithm for convex minimization. SIAM J. Optim. **24**(1), 232-256 (2014).
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183-202 (2009).
6. Bertsekas, D.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, (1982).
7. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1-122 (2010).
8. Candes, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal. Proc. Mag. **25**(2), 21-30 (2008).
9. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129-159 (2001).
10. Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Program. **64**(1-3), 81-101 (1994).
11. Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Math. Program. **141**(1-2), 349-382 (2013).
12. Gu, G., He, B., Yuan, X.: Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach. Comput. Optim. Appl. **59**(1-2), 135-161 (2014).
13. He, B., Yuan, X.: On the acceleration of augmented Lagrangian method for linearly constrained optimization. Optimization online. [http://www.optimization-online.org/DB_FILE/2010/10/2760.pdf](http://www.optimization-online.org/DB_FILE/2010/10/2760.pdf) (2010).
14. Huang, B., Ma, S., Goldfarb, D.: Accelerated linearized Bregman method. J. Sci. Comput. **54**(2), 428-453 (2013).
15. Kang, M., Yun, S., Woo, H., Kang, M.: Accelerated Bregman method for linearly constrained $\ell_1-\ell_2$ minimization. J. Sci. Comput. **56**(3), 515-534 (2013).
16. Kang, M., Kang, M., Jung, M.: Inexact accelerated augmented Lagrangian methods. Comput. Optim. Appl. **62**(2), 373-404 (2015).
17. Ma, F., Ni, M.: A class of customized proximal point algorithms for linearly constrained convex optimization. Comput. Appl. Math. **37**(2), 896-911 (2018).
18. Madan, R., Lall, S.: Distributed algorithms for maximum lifetime routing in wireless sensor networks. IEEE Trans Wirel. Commun. **5**(8), 2185-2193 (2006).
19. Nedic, A., Ozdaglar, A.: Cooperative Distributed Multi-agent. Convex optimization in signal processing and communications. Cambridge University Press, (2010).
20. Liu, Y.F., Liu, X., Ma, S.: On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. Math. Oper. Res. 44(2), 632-650 (2019).
21. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Insov. Math. Dokl. 27, 372-376 (1983).
22. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. 140(1), 125-161 (2013).
23. Nocedal, J., Wright, S.: Numerical Optimization. Springer Science and Business Media, (2006).
24. Shi, G., Johansson, K.H.: Randomized optimal consensus of multi-agent systems. Automatica. 48(12), 3018-3030 (2012).
25. Su, W., Boyd, S., Candes, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. J. Mach. Learn. Res. 17(1), 5312-5354 (2016).
26. Tran-Dinh, Q., Zhu, Y.: Non-stationary first-Order primal-Dual algorithms with fast convergence rates. SIAM J. Optim. 30(4), 2866-2896 (2020).
27. Van Den Berg, E., Friedlander, M.P.: Probing the Pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. 31(2), 890-912 (2009).
28. Xu, Y.: Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. SIAM J. Optim. 27(3), 1459-1484 (2017).
29. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for $\ell_1$-minimization with applications to compressed sensing. SIAM J. Imaging Sci. 1(1), 143-168 (2008).
30. Zeng, X., Lei, J., Chen, J.: Dynamical primal-dual accelerated method with applications to network optimization. https://arxiv.org/abs/1912.03690 (2019).