ORIGINAL PAPER

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

Xin He[1] · Rong Hu[2] · Ya-Ping Fang[1] (ORCID)

Abstract

In this paper, we propose an inertial accelerated 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 $\mathcal{O}(1/k^2)$ convergence rate on the objective residual and the feasibility violation in the primal model. Numerical results are reported to demonstrate the validity of the proposed methods.

Keywords Inertial accelerated primal-dual method · Linear equality constrained convex optimization problem · $\mathcal{O}(1/k^2)$ convergence rate · Inexactness

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

1 Introduction

Consider the linear equality constrained convex optimization problem:

$$\min_{x} \quad F(x), \quad s.t. \ Ax = b, \tag{1}$$

✉ Ya-Ping Fang
ypfang@scu.edu.cn

Xin He
hexinuser@163.com

Rong Hu
ronghumath@aliyun.com

[1] Department of Mathematics, Sichuan University, Chengdu, Sichuan, People's Republic of China

[2] Department of Applied Mathematics, Chengdu University of Information Technology, Chengdu, Sichuan, People's Republic of China

where $F : \mathbb{R}^n \to \mathbb{R}$ is a 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., [9, 10]):

$$\min_{x} \quad \|x\|_1, \quad s.t. \ Ax = b, \tag{2}$$

where $A \in \mathbb{R}^{m \times n}$ with $m \ll n$, and $\| \cdot \|_1$ is the ℓ_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 [31] and [33].

Example 1.2 The linearly constrained $\ell_1 - \ell_2$ minimization problem [17]:

$$\min_{x} \quad \|x\|_1 + \frac{\beta}{2}\|x\|_2^2, \quad s.t. \ Ax = b, \tag{3}$$

where $\beta > 0$ and $\| \cdot \|_2$ is the ℓ_2-norm of $\mathbb{R}^n$ defined by $\|x\|_2^2 = \sum_{i=1}^{n} x_i^2$. When β 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 [8]:

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

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 [23], routing of wireless sensor networks [22] and optimal consensus of agents [28].

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

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

where ∂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 Ω be the KKT point set of the problem (1). Then, for any $(x^*, \lambda^*) \in \Omega$, from (4) we have

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

where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 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]:

$$\begin{cases} x_{k+1} \in \arg\min_x \mathcal{L}(x, \lambda_k) + \frac{\sigma}{2}\|Ax - b\|^2, \\ \lambda_{k+1} = \lambda_k + \sigma(Ax_{k+1} - b). \end{cases} \tag{5}$$

In general, since $\mathcal{L}(x, \lambda_k) + \frac{\sigma}{2}\|Ax - b\|^2$ is not strictly convex, the subproblem may have more than one solutions and be difficult to solve. To overcome this disadvantage, the proximal ALM [11] has been proposed:

$$\begin{cases} x_{k+1} = \arg\min_x F(x) + \langle A^T\lambda_k, x\rangle + \frac{\sigma}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x_k\|_P^2, \\ \lambda_{k+1} = \lambda_k + \sigma(Ax_{k+1} - b), \end{cases} \tag{6}$$

where $\|x\|_P^2 = x^T P x$ 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:

$$\min_x \quad f(x) + g(x), \quad s.t. \ Ax = b. \tag{7}$$

An application of the method (6) to the problem (7) with linearizing the smooth function g leads to the linearized ALM [32]:

$$\begin{cases} x_{k+1} \in \arg\min_x f(x) + \langle\nabla g(x_k) + A^T\lambda_k, x\rangle + \frac{\sigma}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x_k\|_P^2, \\ \lambda_{k+1} = \lambda_k + \sigma(Ax_{k+1} - b). \end{cases} \tag{8}$$

1.1 Related works

Under the assumption that F is smooth, He and Yuan [15] showed that the iteration-complexity of the method (5) is $\mathcal{O}(1/k)$ in terms of the objective residual of the associated $\mathcal{L}(x, \lambda)$. When F is nonsmooth, Gu et al. [13] proved that the method (5) enjoys a worst-case $\mathcal{O}(1/k)$ convergence rate in the ergodic sense. A worst-case $\mathcal{O}(1/k)$ convergence rate in the non-ergodic sense of the method (6) was shown in [21]. When g has a Lipschitz continuous gradient with constant L_g and $P \succ L_g Id$, Xu [32] proved that the method (8) achieves $\mathcal{O}(1/k)$ convergence rate in the ergodic sense. Tran-Dinh and Zhu [30] proposed a modified version of the method (8) and proved that the objective residual and feasibility violation sequences generated by the method both enjoy $\mathcal{O}(1/k)$ non-ergodic convergence rate. Liu et al. [24] investigated the non-ergodic convergence rate of an inexact augmented Lagrangian method for the 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 [25] 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 $\mathcal{O}(1/k^2)$ convergence rate. Beck and Teboulle [5] proposed an iterative shrinkage-thresholding algorithm for solving the linear inverse problem, which

achieves $\mathcal{O}(1/k^2)$ convergence rate. The acceleration idea of [25] was further applied in Nesterov [26] to design the accelerated methods for unconstrained convex composite optimization problems. Su et al. [29] 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., [1, 3, 4]). For more results on inertial methods for unconstrained optimization problems, we refer the reader to [2, 12, 27].

Meanwhile, inertial accelerated methods for linearly constrained optimization problems have also been well-developed. When f is differentiable, He and Yuan [15] proposed an accelerated inertial ALM for the problem (1) and proved that its convergence rate is $\mathcal{O}(1/k^2)$ by using an extrapolation technique similar to [5]. When f is a differentiable function with Lipschitz continuous gradient, by time discretization of dynamical system, Boţ et al. [7] proposed an inertial ALM with $\mathcal{O}(1/k^2)$ convergence rate and provided the convergence of the sequence of iterates. Kang et al. [18] presented an inexact version of the accelerated ALM with inexact calculations of subproblems and showed that the convergence rate remains $\mathcal{O}(1/k^2)$ under the assumption that F is strongly convex. Kang et al. [17] further presented an accelerated Bregman method for the linearly constrained $\ell_1 - \ell_2$ minimization problem, and a convergence rate of $\mathcal{O}(1/k^2)$ was proved when the accelerated Bregman method is applied to solve the problem (1). To linearize the augmented term of the Bregman method, Huang et al. [16] raised an accelerated linearized Bregman algorithm with $\mathcal{O}(1/k^2)$ convergence rate. For the problem (7), Tran-Dinh and Zhu [30] proposed an inertial primal-dual method which enjoys $o(1/k\sqrt{\log k})$ convergence rate. Xu [32] proposed an accelerated version of the linearized ALM (8), named the accelerated linearized augmented Lagrangian method, which is formulated as follows:

$$
\begin{cases}
\hat{x}_k = (1 - \alpha_k)\bar{x}_k + \alpha_k x_k, \\
x_{k+1} \in \arg\min_x f(x) + \frac{\beta_k}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x_k\|_{P_k}^2 + \langle \nabla g(\hat{x}_k) + A^T \lambda_k, x \rangle, \\
\bar{x}_{k+1} = (1 - \alpha_k)\bar{x}_k + \alpha_k x_{k+1}, \\
\lambda_{k+1} = \lambda_k + \gamma_k (Ax_{k+1} - b).
\end{cases}
\tag{9}
$$

It was shown in Xu [32] that the algorithm (9) enjoys $\mathcal{O}(1/k^2)$ convergence rate under specific parameter settings. It is worth mentioning that to achieve the $\mathcal{O}(1/k^2)$ rate, linearization to the augmented term is not allowed in the algorithm (9) since it may cause great difficulty on solving subproblems. Xu [32] did not discuss the convergence analysis of the method when the subproblem is solved inexactly.

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 inspired by the second-order primal-dual dynamical system in [14, 34] and the Nesterov accelerated methods for unconstrained optimization problem [2, 5, 25]. When the objective has the composite structure: $F(x) = f(x) + g(x)$, by linearizing the smooth function g and introducing the perturbed sequence $\{\epsilon_k\}_{k \geq 1}$ in Step 2 of Algorithm 1, we propose an inexact inertial

accelerated 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.

Algorithm 1 Inertial accelerated primal-dual method for problem (1).

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

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

    Step 1: Compute
$$\bar{x}_k = x_k + \frac{k-2}{k+\alpha-2}(x_k - x_{k-1}), \quad \bar{\lambda}_k = \lambda_k + \frac{k-2}{k+\alpha-2}(\lambda_k - \lambda_{k-1}).$$

    Step 2: Set
$$\hat{\lambda}_k = \frac{k+\alpha-2}{\alpha-1}\bar{\lambda}_k - \frac{k-1}{\alpha-1}\lambda_k, \quad \eta_k = \frac{k-1}{k+\alpha-2}Ax_k + \frac{\alpha-1}{k+\alpha-2}b.$$

    Choose $M_k \in \mathbb{S}_+(n)$ and update
$$x_{k+1} \in \arg\min_x F(x) + \frac{k+\alpha-2}{2sk}\|x - \bar{x}_k\|_{M_k}^2 + \frac{sk(k+\alpha-2)}{2(\alpha-1)^2}\|Ax - \eta_k\|^2$$
$$+ \langle A^T\hat{\lambda}_k, x \rangle.$$

    Step 3: $\lambda_{k+1} = \bar{\lambda}_k + \frac{sk}{k+\alpha-2}(Ax_{k+1} - b + \frac{k-1}{\alpha-1}A(x_{k+1} - x_k))$,

    if *A stopping condition is satisfied* **then**
       | Return (x_{k+1}, λ_{k+1})
    end

end

Algorithm 2 Inexact inertial accelerated linearized primal-dual method for problem (7).

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

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

    Step 1: Compute
$$\bar{x}_k = x_k + \frac{k-2}{k+\alpha-2}(x_k - x_{k-1}), \quad \bar{\lambda}_k = \lambda_k + \frac{k-2}{k+\alpha-2}(\lambda_k - \lambda_{k-1}).$$

    Step 2: Set
$$\hat{\lambda}_k = \frac{k+\alpha-2}{\alpha-1}\bar{\lambda}_k - \frac{k-1}{\alpha-1}\lambda_k, \quad \eta_k = \frac{k-1}{k+\alpha-2}Ax_k + \frac{\alpha-1}{k+\alpha-2}b.$$

    Choose $M_k \in \mathbb{S}_+(n)$, $\epsilon_k \in \mathbb{R}^n$ and update
$$x_{k+1} \in \arg\min_x f(x) + \frac{k+\alpha-2}{2sk}\|x - \bar{x}_k\|_{M_k}^2 + \frac{sk(k+\alpha-2)}{2(\alpha-1)^2}\|Ax - \eta_k\|^2$$
$$+ \langle \nabla g(\bar{x}_k) + A^T\hat{\lambda}_k - \epsilon_k, x \rangle.$$

    Step 3: $\lambda_{k+1} = \bar{\lambda}_k + \frac{sk}{k+\alpha-2}(Ax_{k+1} - b + \frac{k-1}{\alpha-1}A(x_{k+1} - x_k))$,

    if *A stopping condition is satisfied* **then**
       | Return (x_{k+1}, λ_{k+1})
    end

end

1.3 Outline

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

2 Convergence analysis

In this section we analyze the convergence rates of Algorithm 1 and Algorithm 2. Assuming merely convexity, we show that both of them enjoy $\mathcal{O}(1/k^2)$ convergence rates in terms of the objective function and the primal feasibility.

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 ℓ_2-norm.

Let $\mathbb{S}_+(n)$ denote the set of all positive symmetric semidefinite matrixes in $\mathbb{R}^{n\times n}$ and Id be the identity matrix. For $M \in \mathbb{S}_+(n)$, 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 $\mathbb{S}_+(n)$ the following partial ordering: for any $M_1, M_2 \in \mathbb{S}_+(n)$,

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

For any $x, y \in \mathbb{R}^n$, the following equality holds:

$$\frac{1}{2}\|x\|_M^2 - \frac{1}{2}\|y\|_M^2 = \langle x, M(x-y)\rangle - \frac{1}{2}\|x-y\|_M^2, \quad \forall M \in \mathbb{S}_+(n). \quad (10)$$

Now, we start to analyze Algorithm 1.

Lemma 1 *Let $\{(x_k, \lambda_k, \bar{x}_k)\}_{k\ge 1}$ be the sequence generated by Algorithm 1. Then,*

$$\frac{k+\alpha-2}{k}M_k(x_{k+1} - \bar{x}_k) \in -s\left(\partial F(x_{k+1}) + A^T(\lambda_{k+1} + \frac{k-1}{\alpha-1}(\lambda_{k+1} - \lambda_k))\right). \quad (11)$$

Proof From step 2, we have

$$0 \in \partial F(x_{k+1}) + \frac{k+\alpha-2}{sk}M_k(x_{k+1} - \bar{x}_k) + \frac{sk(k+\alpha-2)}{(\alpha-1)^2}A^T(Ax_{k+1} - \eta_k) + A^T\hat{\lambda}_k.$$

This yields

$$\frac{k+\alpha-2}{k}M_k(x_{k+1} - \bar{x}_k) \in -s\left(\partial F(x_{k+1}) + A^T\left(\frac{sk(k+\alpha-2)}{(\alpha-1)^2}(Ax_{k+1} - \eta_k) + \hat{\lambda}_k\right)\right). \quad (12)$$

It follows from Step 2 and Step 3 that

$$
\frac{sk(k+\alpha-2)}{(\alpha-1)^2}(Ax_{k+1}-\eta_k)+\hat{\lambda}_k
$$
$$
=\frac{sk(k+\alpha-2)}{(\alpha-1)^2}Ax_{k+1}-\frac{sk(k-1)}{(\alpha-1)^2}Ax_k-\frac{sk}{\alpha-1}b+\hat{\lambda}_k
$$
$$
=\frac{sk}{\alpha-1}(Ax_{k+1}-b+\frac{k-1}{\alpha-1}A(x_{k+1}-x_k))+\hat{\lambda}_k
$$
$$
=\frac{k+\alpha-2}{\alpha-1}(\lambda_{k+1}-\bar{\lambda}_k)+\frac{k+\alpha-2}{\alpha-1}\bar{\lambda}_k-\frac{k-1}{\alpha-1}\lambda_k
$$
$$
=\lambda_{k+1}+\frac{k-1}{\alpha-1}(\lambda_{k+1}-\lambda_k).
$$

This together with (12) yields (11).   □

Lemma 2 *Suppose that F is a convex function, $\Omega \neq \emptyset$ and $M_{k-1} \succcurlyeq M_k$. Let $\{(x_k, \lambda_k, \bar{x}_k, \hat{\lambda}_k)\}_{k\geq 1}$ be the sequence generated by Algorithm 1 and $(x^*, \lambda^*) \in \Omega$. Define*

$$
\mathcal{E}_k=\frac{s(k^2-k)}{(\alpha-1)^2}(\mathcal{L}(x_k,\lambda^*)-\mathcal{L}(x^*,\lambda^*))+\frac{1}{2}\|\hat{x}_k-x^*\|_{M_{k-1}}^2+\frac{1}{2}\|\hat{\lambda}_k-\lambda^*\|^2 \tag{13}
$$

with

$$
\hat{x}_k=\frac{k+\alpha-2}{\alpha-1}\bar{x}_k-\frac{k-1}{\alpha-1}x_k. \tag{14}
$$

Then, for any $k \geq 1$, we have

$$
\mathcal{E}_{k+1}\leq\mathcal{E}_k-\frac{(k+\alpha-2)^2}{2(\alpha-1)^2}(\|x_{k+1}-\bar{x}_k\|_{M_k}^2+\|\lambda_{k+1}-\bar{\lambda}_k\|^2).
$$

Proof By computation,

$$
\hat{x}_{k+1}=\frac{k+\alpha-1}{\alpha-1}(x_{k+1}+\frac{k-1}{k+\alpha-1}(x_{k+1}-x_k))-\frac{k}{\alpha-1}x_{k+1}
$$
$$
=\frac{k+\alpha-2}{\alpha-1}x_{k+1}-\frac{k-1}{\alpha-1}x_k \tag{15}
$$
$$
=\frac{k+\alpha-2}{\alpha-1}(x_{k+1}-\bar{x}_k)+\frac{k+\alpha-2}{\alpha-1}\bar{x}_k-\frac{k-1}{\alpha-1}x_k
$$
$$
=\hat{x}_k+\frac{k+\alpha-2}{\alpha-1}(x_{k+1}-\bar{x}_k)
$$

and

$$
\hat{x}_{k+1}-x^*=x_{k+1}-x^*+\frac{k-1}{\alpha-1}(x_{k+1}-x_k). \tag{16}
$$

Similarly, we have

$$
\hat{\lambda}_{k+1}=\hat{\lambda}_k+\frac{k+\alpha-2}{\alpha-1}(\lambda_{k+1}-\bar{\lambda}_k) \tag{17}
$$

and

$$
\hat{\lambda}_{k+1}-\lambda^*=\lambda_{k+1}-\lambda^*+\frac{k-1}{\alpha-1}(\lambda_{k+1}-\lambda_k). \tag{18}
$$

By the definition of $\mathcal{L}(x, \lambda)$, we get $\partial_x \mathcal{L}(x, \lambda) = \partial F(x) + A^T \lambda$. Combining this and equality (18), we can rewrite (11) as

$$\frac{k + \alpha - 2}{k} M_k(x_{k+1} - \bar{x}_k) \in -s(\partial F(x_{k+1}) + A^T \lambda^* + A^T (\lambda_{k+1}$$

$$-\lambda^* + \frac{k-1}{\alpha - 1}(\lambda_{k+1} - \lambda_k)))$$

$$= -s\partial_x \mathcal{L}(x_{k+1}, \lambda^*) - sA^T (\hat{\lambda}_{k+1} - \lambda^*),$$

which implies

$$\xi_k := -\frac{k + \alpha - 2}{sk} M_k(x_{k+1} - \bar{x}_k) - A^T (\hat{\lambda}_{k+1} - \lambda^*) \in \partial_x \mathcal{L}(x_{k+1}, \lambda^*). \quad (19)$$

Since $M_{k-1} \succcurlyeq M_k \succcurlyeq 0$, it follows from (10) and (15) that

$$\frac{1}{2}\|\hat{x}_{k+1} - x^*\|^2_{M_k} - \frac{1}{2}\|\hat{x}_k - x^*\|^2_{M_{k-1}}$$

$$= \frac{1}{2}\|\hat{x}_{k+1} - x^*\|^2_{M_k} - \frac{1}{2}\|\hat{x}_k - x^*\|^2_{M_k} - \frac{1}{2}\|\hat{x}_k - x^*\|^2_{M_{k-1} - M_k}$$

$$\leq \langle \hat{x}_{k+1} - x^*, M_k(\hat{x}_{k+1} - \hat{x}_k)\rangle - \frac{1}{2}\|\hat{x}_{k+1} - \hat{x}_k\|^2_{M_k}$$

$$= \frac{k + \alpha - 2}{\alpha - 1} \langle \hat{x}_{k+1} - x^*, M_k(x_{k+1} - \bar{x}_k)\rangle$$

$$- \frac{(k + \alpha - 2)^2}{2(\alpha - 1)^2} \|x_{k+1} - \bar{x}_k\|^2_{M_k} \quad (20)$$

$$= -\frac{sk}{\alpha - 1}(\langle \hat{x}_{k+1} - x^*, \xi_k\rangle + \langle \hat{x}_{k+1} - x^*, A^T (\hat{\lambda}_{k+1} - \lambda^*)\rangle)$$

$$- \frac{(k + \alpha - 2)^2}{2(\alpha - 1)^2} \|x_{k+1} - \bar{x}_k\|^2_{M_k}.$$

Since $\mathcal{L}(x, \lambda^*)$ is a convex function with respect to x, from (16) and (19) we get

$$\langle \hat{x}_{k+1} - x^*, \xi_k\rangle = \langle x_{k+1} - x^*, \xi_k\rangle + \frac{k-1}{\alpha - 1} \langle x_{k+1} - x_k, \xi_k\rangle$$

$$\geq \mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*) + \frac{k-1}{\alpha - 1}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x_k, \lambda^*)). \quad (21)$$

Combining (20) and (21) together, we have

$$\frac{1}{2}\|\hat{x}_{k+1} - x^*\|^2_{M_k} - \frac{1}{2}\|\hat{x}_k - x^*\|^2_{M_{k-1}}$$

$$\leq -\frac{sk}{\alpha - 1}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) - \frac{s(k^2 - k)}{(\alpha - 1)^2}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x_k, \lambda^*))$$

$$- \frac{sk}{\alpha - 1} \langle \hat{x}_{k+1} - x^*, A^T (\hat{\lambda}_{k+1} - \lambda^*)\rangle - \frac{(k + \alpha - 2)^2}{2(\alpha - 1)^2} \|x_{k+1} - \bar{x}_k\|^2_{M_k}. \quad (22)$$

Since $Ax^* = b$, it follows from Step 3 of Algorithm 1 and (16) that

$$\lambda_{k+1} - \bar{\lambda}_k = \frac{sk}{k + \alpha - 2}(Ax_{k+1} - Ax^* + \frac{k-1}{\alpha - 1} A(x_{k+1} - x_k)) = \frac{sk}{k + \alpha - 2} A(\hat{x}_{k+1} - x^*).$$

This together with (10) and (17) yields

$$\frac{1}{2}\|\hat{\lambda}_{k+1} - \lambda^*\|^2 - \frac{1}{2}\|\hat{\lambda}_k - \lambda^*\|^2 = \langle\hat{\lambda}_{k+1} - \lambda^*, \hat{\lambda}_{k+1} - \hat{\lambda}_k\rangle - \frac{1}{2}\|\hat{\lambda}_{k+1} - \hat{\lambda}_k\|^2$$

$$= \frac{k+\alpha-2}{\alpha-1}\langle\hat{\lambda}_{k+1} - \lambda^*, \lambda_{k+1} - \bar{\lambda}_k\rangle - \frac{(k+\alpha-2)^2}{2(\alpha-1)^2}\|\lambda_{k+1} - \bar{\lambda}_k\|^2 \qquad (23)$$

$$= \frac{sk}{\alpha-1}\langle\hat{\lambda}_{k+1} - \lambda^*, A(\hat{x}_{k+1} - x^*)\rangle - \frac{(k+\alpha-2)^2}{2(\alpha-1)^2}\|\lambda_{k+1} - \bar{\lambda}_k\|^2.$$

It follows from (22) and (23) that

$$\mathcal{E}_{k+1} - \mathcal{E}_k$$
$$= \frac{s(k^2+k)}{(\alpha-1)^2}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) - \frac{s(k^2-k)}{(\alpha-1)^2}(\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*))$$
$$+ \frac{1}{2}\|\hat{x}_{k+1} - x^*\|^2_{M_k} - \frac{1}{2}\|\hat{x}_k - x^*\|^2_{M_{k-1}} + \frac{1}{2}\|\hat{\lambda}_{k+1} - \lambda^*\|^2 - \frac{1}{2}\|\hat{\lambda}_k - \lambda^*\|^2$$
$$\leq \frac{(3-\alpha)sk}{(\alpha-1)^2}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) - \frac{(k+\alpha-2)^2}{2(\alpha-1)^2}(\|x_{k+1} - \bar{x}_k\|^2_{M_k} + \|\lambda_{k+1} - \bar{\lambda}_k\|^2)$$
$$\leq -\frac{(k+\alpha-2)^2}{2(\alpha-1)^2}(\|x_{k+1} - \bar{x}_k\|^2_{M_k} + \|\lambda_{k+1} - \bar{\lambda}_k\|^2),$$

where the last inequality follows from $\alpha \geq 3$ and $(x^*, \lambda^*) \in \Omega$. This yields the desire result. $\qquad\square$

To obtain the fast convergence rates, we need the following lemma.

Lemma 3 *([20, Lemma 2], [19, Lemma 3.18]) Let $\{a_k\}_{k=1}^{+\infty}$ be a sequence of vectors in $\mathbb{R}^n$ such that*

$$\|(\tau + (\tau - 1)K)a_{K+1} + \sum_{k=1}^{K} a_k\| \leq C, \qquad \forall K \geq 1,$$

where $\tau > 1$ and $C \geq 0$. Then, $\|\sum_{k=1}^{K} a_k\| \leq C$ for all $K \geq 1$.

Now, we discuss the $\mathcal{O}(1/k^2)$ convergence rate of Algorithms 1.

Theorem 1 *Suppose that F is a convex function, $\Omega \neq \emptyset$ and $M_{k-1} \succcurlyeq M_k$. Let $\{(x_k, \lambda_k, \bar{x}_k, \bar{\lambda}_k)\}_{k\geq1}$ be the sequence generated by Algorithm 1 and $(x^*, \lambda^*) \in \Omega$. The following conclusions hold:*

(i) $\sum_{k=1}^{+\infty} k^2(\|x_{k+1} - \bar{x}_k\|^2_{M_k} + \|\lambda_{k+1} - \bar{\lambda}_k\|^2) < +\infty.$
(ii) *For all $k > 1$,*

$$\|Ax_k - b\| \leq \frac{4(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s(k-1)(k+\alpha-3)},$$

$$|F(x_k) - F(x^*)| \leq \frac{(\alpha-1)^2\mathcal{E}_1}{s(k^2-k)} + \frac{4(\alpha-1)^2\sqrt{2\mathcal{E}_1}\|\lambda^*\|}{s(k-1)(k+\alpha-3)},$$

where $\mathcal{E}_1 = \frac{1}{2}\|x_1 - x^*\|_{M_0}^2 + \frac{1}{2}\|\lambda_1 - \lambda^*\|^2.$

Proof From Lemma 2, we have

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq -\frac{(k+\alpha-2)^2}{2(\alpha-1)^2}(\|x_{k+1} - \bar{x}_k\|_{M_k}^2 + \|\lambda_{k+1} - \bar{\lambda}_k\|^2) \leq 0, \quad \forall k \geq 1. \quad (24)$$

By the definition of $\mathcal{E}_k$ and (24), $\{\mathcal{E}_k\}_{k\geq 1}$ is a nonincreasing and positive sequence. As a consequence, $\mathcal{E}_k$ converges to some point. It follows from (24) that

$$\begin{aligned}
\sum_{k=1}^{+\infty} & \frac{(k+\alpha-2)^2}{2(\alpha-1)^2}(\|x_{k+1} - \bar{x}_k\|_{M_k}^2 + \|\lambda_{k+1} - \bar{\lambda}_k\|^2)) \\
& \leq \lim_{K\to+\infty} \sum_{k=1}^{K}(\mathcal{E}_k - \mathcal{E}_{k+1}) \\
& = \mathcal{E}_1 - \lim_{K\to+\infty} \mathcal{E}_{K+1} \\
& < +\infty,
\end{aligned} \quad (25)$$

which is (i).

Combining (13) and (24), we have

$$\|\hat{\lambda}_k - \lambda^*\| \leq \sqrt{2\mathcal{E}_k} \leq \sqrt{2\mathcal{E}_1}, \qquad \forall k \geq 1.$$

This yields

$$\begin{aligned}
\left\| \sum_{k=1}^{K}(\hat{\lambda}_{k+1} - \hat{\lambda}_k) \right\| & = \left\| \hat{\lambda}_{K+1} - \hat{\lambda}_1 \right\| \\
& \leq \left\| \hat{\lambda}_{K+1} - \lambda^* \right\| + \left\| \hat{\lambda}_1 - \lambda^* \right\| \\
& \leq 2\sqrt{2\mathcal{E}_1}
\end{aligned} \quad (26)$$

for all $K \geq 1$. It follows form (17) and Step 3 of Algorithm 1 that

$$\begin{aligned}
\left\| \sum_{k=1}^{K}(\hat{\lambda}_{k+1} - \hat{\lambda}_k) \right\| & = \left\| \sum_{k=1}^{K} \frac{k+\alpha-2}{\alpha-1}(\lambda_{k+1} - \bar{\lambda}_k) \right\| \\
& = \frac{s}{\alpha-1} \left\| \sum_{k=1}^{K} k(Ax_{k+1} - b + \frac{k-1}{\alpha-1}A(x_{k+1} - x_k)) \right\| \\
& = \frac{s}{(\alpha-1)^2} \left\| \sum_{k=1}^{K}(k(k+\alpha-2)(Ax_{k+1} - b) - k(k-1)(Ax_k - b)) \right\| \\
& = \frac{s}{(\alpha-1)^2} \left\| K(K+\alpha-2)(Ax_{K+1} - b) + \sum_{k=1}^{K}(\alpha-3)(k-1)(Ax_k - b) \right\|.
\end{aligned}$$

This together with (26) implies

$$\left\| K(K+\alpha-2)(Ax_{K+1}-b)+\sum_{k=1}^{K}((\alpha-3)(k-1)(Ax_k-b))\right\| \leq \frac{2(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s}. \quad (27)$$

When $\alpha = 3$, it follows from (27) that

$$\|K(K+\alpha-2)(Ax_{K+1}-b)\| \leq \frac{2(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s}.$$

When $\alpha > 3$: applying Lemma 3 with $a_k = (\alpha-3)(k-1)(Ax_k-b)$, $\tau = \frac{\alpha-2}{\alpha-3}$ and $C = \frac{2(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s}$, from (27), we obtain

$$\left\| \sum_{k=1}^{K}((\alpha-3)(k-1)(Ax_k-b))\right\| \leq \frac{2(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s},$$

which together with (27) yields

$$\|K(K+\alpha-2)(Ax_{K+1}-b)\| \leq \frac{4(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s}.$$

From above discussion, when $\alpha \geq 3$, we have

$$\|Ax_k - b\| \leq \frac{4(\alpha-1)^2\sqrt{2\mathcal{E}_1}}{s(k-1)(k+\alpha-3)}, \quad \forall k > 1. \quad (28)$$

It follows form the definition of $\mathcal{E}_k$ and (24) that

$$\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \leq \frac{(\alpha-1)^2\mathcal{E}_k}{s(k^2-k)} \leq \frac{(\alpha-1)^2\mathcal{E}_1}{s(k^2-k)}$$

for all $k > 1$. This together with (28) implies that

$$\begin{aligned}
|F(x_k) - F(x^*)| &= |\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) - \langle \lambda^*, Ax_k - b \rangle| \\
&\leq \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) + \|\lambda^*\|\|Ax_k - b\| \\
&\leq \frac{(\alpha-1)^2\mathcal{E}_1}{s(k^2-k)} + \frac{4(\alpha-1)^2\sqrt{2\mathcal{E}_1}\|\lambda^*\|}{s(k-1)(k+\alpha-3)}
\end{aligned}$$

for all $k > 1$. The proof is complete. □

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

Assumption (H) $\Omega \neq \emptyset$, f is a convex function, 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,

$$g(x) \leq g(y) + \langle \nabla g(y), x-y \rangle + \frac{L_g}{2}\|x-y\|^2, \quad \forall x, y \in \mathbb{R}^n. \quad (29)$$

Lemma 4 *Let* $\{(x_k, \lambda_k, \bar{x}_k)\}_{k \geq 1}$ *the sequence generated by Algorithm 2. Then,*

$$\frac{k + \alpha - 2}{k} M_k(x_{k+1} - \bar{x}_k)$$

$$\in -s\left(\partial f(x_{k+1}) + \nabla g(\bar{x}_k) + A^T(\lambda_{k+1} + \frac{k-1}{\alpha-1}(\lambda_{k+1} - \lambda_k)) - \epsilon_k\right). \quad (30)$$

Proof From Step 2 of Algorithm 2 , we have

$$0 \in \partial f(x_{k+1}) + \nabla g(\bar{x}_k) + \frac{k + \alpha - 2}{sk} M_k(x_{k+1} - \bar{x}_k)$$

$$+ \frac{sk(k + \alpha - 2)}{(\alpha - 1)^2} A^T(Ax_{k+1} - \eta_k) + A^T \hat{\lambda}_k - \epsilon_k,$$

which yields

$$\frac{k + \alpha - 2}{k} M_k(x_{k+1} - \bar{x}_k)$$

$$\in -s\left(\partial f(x_{k+1}) + \nabla g(\bar{x}_k) + A^T\left(\frac{sk(k + \alpha - 2)}{(\alpha - 1)^2}(Ax_{k+1} - \eta_k) + \hat{\lambda}_k\right) - \epsilon_k\right).$$

The rest of the proof is similar as the one of Lemma 1, and so we omit it. □

Lemma 5 *Assume that Assumption (H) holds, and* $M_{k-1} \succcurlyeq M_k \succcurlyeq sL_g Id$. *Let* $\{(x_k, \lambda_k, \hat{\lambda}_k)\}_{k \geq 1}$ *be the sequence generated by Algorithm 2 and* $(x^*, \lambda^*) \in \Omega$. *Define*

$$\mathcal{E}_k^\epsilon = \mathcal{E}_k - \sum_{j=1}^k \frac{s(j-1)}{\alpha - 1} \langle \hat{x}_j - x^*, \epsilon_{j-1} \rangle, \quad (31)$$

where $\mathcal{E}_k$ *is defined in* (13) *and* $\hat{x}_k$ *is defined in* (14). *Then, for any* $k \geq 1$,

$$\mathcal{E}_{k+1}^\epsilon \leq \mathcal{E}_k^\epsilon.$$

Proof By same arguments as in the proof of Lemma 2, we get

$$\hat{x}_{k+1} - \hat{x}_k = \frac{k + \alpha - 2}{\alpha - 1}(x_{k+1} - \bar{x}_k), \quad (32)$$

$$\hat{x}_{k+1} - x^* = x_{k+1} - x^* + \frac{k-1}{\alpha-1}(x_{k+1} - x_k), \quad (33)$$

$$\hat{\lambda}_{k+1} - \hat{\lambda}_k = \frac{k + \alpha - 2}{\alpha - 1}(\lambda_{k+1} - \bar{\lambda}_k), \quad (34)$$

$$\hat{\lambda}_{k+1} - \lambda^* = \lambda_{k+1} - \lambda^* + \frac{k-1}{\alpha-1}(\lambda_{k+1} - \lambda_k). \quad (35)$$

For notation simplicity, we denote

$$\mathcal{L}^f(x) = f(x) + \langle \lambda^*, Ax - b \rangle. \quad (36)$$

Then, $\mathcal{L}^f$ is a convex function, $\partial \mathcal{L}^f(x) = \partial f(x) + A^T \lambda^*$, and

$$\mathcal{L}(x, \lambda^*) = \mathcal{L}^f(x) + g(x). \quad (37)$$

It follows from (30) and (35) that

$$\frac{k+\alpha-2}{k}M_k(x_{k+1}-\bar{x}_k) \in -s(\partial f(x_{k+1})+A^T\lambda^*)-s\nabla g(\bar{x}_k)$$

$$-sA^T(\lambda_{k+1}-\lambda^*+\frac{k-1}{\alpha-1}(\lambda_{k+1}-\lambda_k))+s\epsilon_k$$

$$= -s\partial\mathcal{L}^f(x_{k+1})-s\nabla g(\bar{x}_k)-sA^T(\hat{\lambda}_{k+1}-\lambda^*)+s\epsilon_k,$$

which yields

$$\xi_k := -\frac{k+\alpha-2}{ks}M_k(x_{k+1}-\bar{x}_k)-\nabla g(\bar{x}_k)-A^T(\hat{\lambda}_{k+1}-\lambda^*)+\epsilon_k \in \partial\mathcal{L}^f(x_{k+1}). \tag{38}$$

Since $M_{k-1} \succcurlyeq M_k$, it follows from (10), (32) and (38) that

$$\frac{1}{2}\|\hat{x}_{k+1}-x^*\|_{M_k}^2 - \frac{1}{2}\|\hat{x}_k-x^*\|_{M_{k-1}}^2$$

$$\leq \langle\hat{x}_{k+1}-x^*, M_k(\hat{x}_{k+1}-\hat{x}_k)\rangle - \frac{1}{2}\|\hat{x}_{k+1}-\hat{x}_k\|_{M_k}^2$$

$$= -\frac{sk}{\alpha-1}(\langle\hat{x}_{k+1}-x^*,\xi_k\rangle + \langle\hat{x}_{k+1}-x^*, A^T(\hat{\lambda}_{k+1}-\lambda^*)\rangle \tag{39}$$

$$+\langle\hat{x}_{k+1}-x^*,\nabla g(\bar{x}_k)\rangle - \langle\hat{x}_{k+1}-x^*,\epsilon_k\rangle) - \frac{(k+\alpha-2)^2}{2(\alpha-1)^2}\|x_{k+1}-\bar{x}_k\|_{M_k}^2.$$

From (33) and (38), we have

$$\langle\hat{x}_{k+1}-x^*,\xi_k\rangle = \langle x_{k+1}-x^*,\xi_k\rangle + \frac{k-1}{\alpha-1}\langle x_{k+1}-x_k,\xi_k\rangle$$

$$\geq \mathcal{L}^f(x_{k+1})-\mathcal{L}^f(x^*) + \frac{k-1}{\alpha-1}(\mathcal{L}^f(x_{k+1})-\mathcal{L}^f(x_k)), \quad (40)$$

where the inequality follows from the convexity of $\mathcal{L}^f$. Since g has a Lipschitz continuous gradient, from (29) we get

$$g(x_{k+1}) \leq g(\bar{x}_k) + \langle\nabla g(\bar{x}_k), x_{k+1}-\bar{x}_k\rangle + \frac{L_g}{2}\|x_{k+1}-\bar{x}_k\|^2. \tag{41}$$

By the convexity of g, we have

$$\langle\nabla g(\bar{x}_k), x_{k+1}-\bar{x}_k\rangle = \langle\nabla g(\bar{x}_k), x_{k+1}-x^*\rangle + \langle\nabla g(\bar{x}_k), x^*-\bar{x}_k\rangle$$

$$\leq \langle\nabla g(\bar{x}_k), x_{k+1}-x^*\rangle + g(x^*) - g(\bar{x}_k) \tag{42}$$

and

$$\langle\nabla g(\bar{x}_k), x_{k+1}-\bar{x}_k\rangle = \langle\nabla g(\bar{x}_k), x_{k+1}-x_k\rangle + \langle\nabla g(\bar{x}_k), x_k-\bar{x}_k\rangle$$

$$\leq \langle\nabla g(\bar{x}_k), x_{k+1}-x_k\rangle + g(x_k) - g(\bar{x}_k). \tag{43}$$

It follows from (41)–(43) 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 (33) yields

$$\langle \hat{x}_{k+1} - x^*, \nabla g(\bar{x}_k) \rangle = \langle \nabla g(\bar{x}_k), x_{k+1} - x^* \rangle + \frac{k-1}{\alpha-1} \langle \nabla g(\bar{x}_k), x_{k+1} - x_k \rangle$$

$$\geq g(x_{k+1}) - g(x^*) + \frac{k-1}{\alpha-1}(g(x_{k+1}) - g(x_k)) \qquad (44)$$

$$- \frac{(k+\alpha-2)L_g}{2(\alpha-1)} \|x_{k+1} - \bar{x}_k\|^2.$$

It follows from (39), (40) and (44) that

$$\frac{1}{2}\|\hat{x}_{k+1} - x^*\|_{M_k}^2 - \frac{1}{2}\|\hat{x}_k - x^*\|_{M_{k-1}}^2$$

$$\leq -\frac{sk}{\alpha-1}(\mathcal{L}^f(x_{k+1}) + g(x_{k+1}) - (\mathcal{L}^f(x^*) + g(x^*)))$$

$$- \frac{sk(k-1)}{(\alpha-1)^2}(\mathcal{L}^f(x_{k+1}) + g(x_{k+1}) - (\mathcal{L}^f(x_k) + g(x_k)))$$

$$- \frac{sk}{\alpha-1}\langle \hat{x}_{k+1} - x^*, A^T(\hat{\lambda}_{k+1} - \lambda^*) \rangle + \frac{sk}{\alpha-1}\langle \hat{x}_{k+1} - x^*, \epsilon_k \rangle \qquad (45)$$

$$- \frac{k+\alpha-2}{2(\alpha-1)^2}\|x_{k+1} - \bar{x}_k\|_{(k+\alpha-2)M_k - sL_g kId}^2$$

$$\leq -\frac{sk}{\alpha-1}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) - \frac{sk(k-1)}{(\alpha-1)^2}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x_k, \lambda^*))$$

$$- \frac{sk}{\alpha-1}\langle \hat{x}_{k+1} - x^*, A^T(\hat{\lambda}_{k+1} - \lambda^*) \rangle + \frac{sk}{\alpha-1}\langle \hat{x}_{k+1} - x^*, \epsilon_k \rangle,$$

where the second inequality follows from the assumption $M_k \succcurlyeq sL_gId \succcurlyeq \frac{sL_gk}{k+\alpha-2}Id$.

It follows from (23), (31) and (45) that

$$\mathcal{E}_{k+1}^{\epsilon} - \mathcal{E}_k^{\epsilon} = \mathcal{E}_{k+1} - \mathcal{E}_k - \frac{sk}{\alpha-1}\langle \hat{x}_{k+1} - x^*, \epsilon_k \rangle$$

$$\leq \frac{(3-\alpha)sk}{(\alpha-1)^2}(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) - \frac{(k+\alpha-2)^2}{2(\alpha-1)^2}\|\lambda_{k+1} - \bar{\lambda}_k\|^2$$

$$\leq 0, \qquad (46)$$

where the last inequality follows from $\alpha \geq 3$ and $(x^*, \lambda^*) \in \Omega$. $\qquad \square$

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

Lemma 6 [2, Lemma 5.14] *Let $\{a_k\}_{k\geq 1}$ and $\{b_k\}_{k\geq 1}$ be two nonnegative sequences such that $\sum_k^{+\infty} b_k < +\infty$ and*

$$a_k^2 \leq c^2 + \sum_{j=1}^k b_j a_j$$

for all $k \geq 1$, where $c \geq 0$. Then,

$$a_k \leq c + \sum_{j=1}^{+\infty} b_j$$

for all $k \geq 1$.

Theorem 2 *Assume that Assumption (H) holds, $M_{k-1} \succcurlyeq M_k \succcurlyeq s L_g Id$ for all $k \geq 1$, and*

$$\sum_{k=1}^{+\infty} k \|\epsilon_k\| < +\infty.$$

Let $\{(x_k, \lambda_k)\}_{k \geq 1}$ be the sequence generated by Algorithm 2 and $(x^*, \lambda^*) \in \Omega$. Then, for all $k > 1$,

$$\|Ax_k - b\| \leq \frac{4(\alpha - 1)^2 \sqrt{2C}}{s(k-1)(k+\alpha-3)},$$

$$|f(x_k) + g(x_k) - f(x^*) - g(x^*)| \leq \frac{(\alpha-1)^2 C}{s(k^2 - k)} + \frac{4(\alpha-1)^2 \sqrt{2C} \|\lambda^*\|}{s(k-1)(k+\alpha-3)},$$

where

$$C := \frac{1}{2} \|x_1 - x^*\|_{M_0}^2 + \frac{1}{2} \|\lambda_1 - \lambda^*\|^2 + \frac{s}{\alpha-1} \left(\sqrt{\frac{2\mathcal{E}_1}{s L_g}} + \frac{2}{(\alpha-1)L_g} \sum_{j=1}^{+\infty} j \|\epsilon_j\| \right) \times \sum_{j=1}^{+\infty} j \|\epsilon_j\|.$$

Proof From Lemma 5 we have

$$\mathcal{E}_{k+1}^{\epsilon} \leq \mathcal{E}_k^{\epsilon} \leq \mathcal{E}_1^{\epsilon},$$

and it yields

$$\mathcal{E}_k \leq \mathcal{E}_1 + \sum_{j=1}^{k} \frac{s(j-1)}{\alpha-1} \langle \hat{x}_j - x^*, \epsilon_{j-1} \rangle. \tag{47}$$

This together with (13) and Cauchy-Schwarz inequality implies

$$\|\hat{x}_k - x^*\|_{M_{k-1}}^2 \leq 2\mathcal{E}_1 + \frac{2s}{\alpha-1} \sum_{j=1}^{k} (j-1) \|\hat{x}_j - x^*\| \cdot \|\epsilon_{j-1}\|.$$

Since $M_{k-1} \succcurlyeq s L_g Id$,

$$\|\hat{x}_k - x^*\|^2 \leq \frac{1}{s L_g} \|\hat{x}_k - x^*\|_{M_{k-1}}^2 \leq \frac{2\mathcal{E}_1}{s L_g} + \frac{2}{(\alpha-1)L_g} \sum_{j=1}^{k} (j-1) \|\hat{x}_j - x^*\| \cdot \|\epsilon_{j-1}\|. \tag{48}$$

Since $\sum_{j=1}^{+\infty} j \|\epsilon_j\| < +\infty$, applying Lemma 6 with $a_k = \|\hat{x}_k - x^*\|$ to (48), we obtain

$$\|\hat{x}_k - x^*\| \leq \sqrt{\frac{2\mathcal{E}_1}{s L_g}} + \frac{2}{(\alpha-1)L_g} \sum_{j=1}^{+\infty} j \|\epsilon_j\| < +\infty, \quad \forall k \geq 1. \tag{49}$$

This together with (47) yields

$$
\mathcal{E}_k \leq \mathcal{E}_1 + \frac{s}{\alpha - 1} \sup_{k \geq 1} \|\hat{x}_k - x^*\| \times \sum_{j=1}^{+\infty} j \|\epsilon_j\|
$$

$$
\leq \mathcal{E}_1 + \frac{s}{\alpha - 1} \left(\sqrt{\frac{2\mathcal{E}_1}{sL_g}} + \frac{2}{(\alpha - 1)L_g} \sum_{j=1}^{+\infty} j \|\epsilon_j\| \right) \times \sum_{j=1}^{+\infty} j \|\epsilon_j\| \qquad (50)
$$

for any $k \geq 1$. Denote

$$
C := \mathcal{E}_1 + \frac{s}{\alpha - 1} \left(\sqrt{\frac{2\mathcal{E}_1}{sL_g}} + \frac{2}{(\alpha - 1)L_g} \sum_{j=1}^{+\infty} j \|\epsilon_j\| \right) \times \sum_{j=1}^{+\infty} j \|\epsilon_j\|
$$

$$
= \frac{1}{2} \|x_1 - x^*\|_{M_0}^2 + \frac{1}{2} \|\lambda_1 - \lambda^*\|^2
$$

$$
+ \frac{s}{\alpha - 1} \left(\sqrt{\frac{2\mathcal{E}_1}{sL_g}} + \frac{2}{(\alpha - 1)L_g} \sum_{j=1}^{+\infty} j \|\epsilon_j\| \right) \times \sum_{j=1}^{+\infty} j \|\epsilon_j\|.
$$

By the definition of $\mathcal{E}_k$ and (50), we have

$$
\frac{s(k^2 - k)}{(\alpha - 1)^2} (\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) \leq \mathcal{E}_k \leq C
$$

and

$$
\|\hat{\lambda}_k - \lambda^*\| \leq \sqrt{2\mathcal{E}_k} \leq \sqrt{2C}, \qquad \forall k \geq 1.
$$

By similar arguments in Theorem 1, we obtain

$$
\|Ax_k - b\| \leq \frac{4(\alpha - 1)^2 \sqrt{2C}}{s(k - 1)(k + \alpha - 3)}
$$

and

$$
|f(x_k) + g(x_k) - f(x^*) - g(x^*)|
$$

$$
\leq \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) + \|\lambda^*\| \|Ax_k - b\|
$$

$$
\leq \frac{(\alpha - 1)^2 C}{s(k^2 - k)} + \frac{4(\alpha - 1)^2 \sqrt{2C} \|\lambda^*\|}{s(k - 1)(k + \alpha - 3)}
$$

for all $k > 1$.   □

Remark 1 It was shown in [32, Theorem 2.9] that the algorithm (9) with adaptive parameters enjoys $\mathcal{O}(1/k^2)$ rate. Xu [32] did not discuss whether the convergence rate of algorithm (9) is preserved when the subproblem is solved inexactly. By Theorem 2, the $\mathcal{O}(1/k^2)$ convergence rate of Algorithm 2 is preserved even if the subproblem is solved inexactly, provided the errors are sufficiently small. The numerical experiments in section 3 also show the effectiveness of the inexact algorithm.

3 Numerical experiments

In this section, we present numerical experiments to illustrate the efficiency of the proposed algorithms. All codes are run on a PC (with 2.3 GHz Quad-Core Intel Core i5 and 8 GB memory) under MATLAB Version 9.4.0.813654 (R2018a).

3.1 The quadratic programming problem

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

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

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 (Al2) with the accelerated linearized augmented Lagrangian method (AALM [32, Algorithm 1]), which enjoys $\mathcal{O}(1/k^2)$ 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, $A = [B, Id]$ with $B \in \mathbb{R}^{m \times (n-m)}$ generated by standard Gaussian 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.,

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

Set the parameters of Algorithm 2 as: $s = \|Q\|$, $M_k = s * \|Q\| Id$. Set the parameters of AALM ([32, Algorithm 1]) with adaptive parameters, in which $\alpha_k = \frac{2}{k+1}$, $\beta_k = \gamma_k = \|Q\|k$, $P_k = \frac{2\|Q\|}{k} Id$. Subproblems for both algorithms are solved by interior-point algorithms to a tolerance $subtol$. Figure 1 describes the distance of optimal value $|F(x_k) - F(x^*)|$ and violation of feasibility $\|Ax_k - b\|$ given Al2 with $\alpha = 10, 20, 30$ and AALM for the first 500 iterations. As shown in Fig. 1, Algorithm 2 performs better and more stable than AALM under different $subtol$.

3.2 The basis pursuit problem

Consider the following basis pursuit problem:

$$\min_{x} \quad \|x\|_1, \quad s.t. \ Ax = b,$$

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 $0.1 * n$, and the nonzero elements are selected randomly in $[-2, 2]$. Set $b = Ax^*$. We compare Algorithm 1 with the inexact augmented Lagrangian method (IAL [24, Algorithm 1]). Here, subproblems for both algorithms are solved by fast iterative
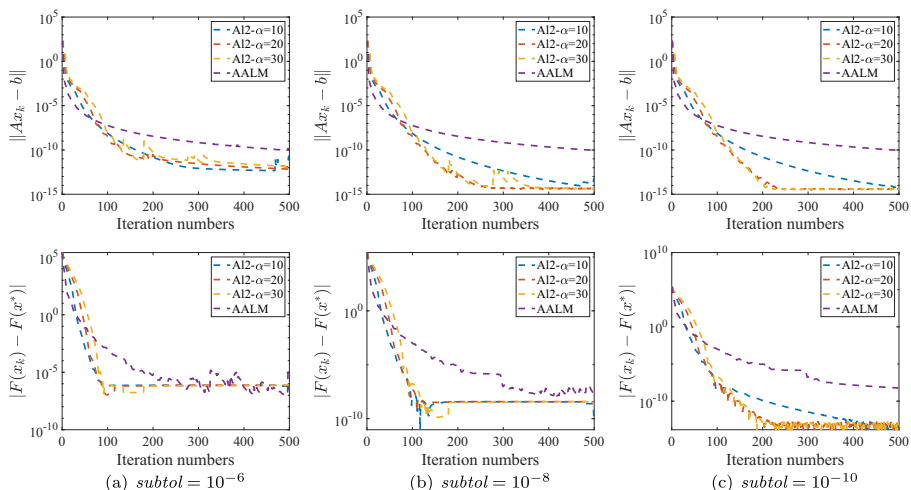
Fig. 1 Error of objective function and constraint of Al2 and AALM with different *subtol*

shrinkage-thresholding algorithm (FISTA [5], [24, Algorithm 2]), and the stopping condition of the FISTA is when

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

is satisfied or the number of iterations exceeds 100, where accuracy $subtol = 1e - 4, \; 1e - 6, \; 1e - 8$. 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 $Res + Rel \leq 1e - 8$. Set the parameters of Algorithm 1 as $\alpha = n$, $s = 100$, $M_k = 0$, and the parameter of IAL as $\beta = 1$. Let *Init* and *Time* denote the number of iterations, and the CPU time in seconds, respectively. Under different tolerance *subtol* of subproblem, Tables 1, 2 and 3 report the results for the basis pursuit problem with

Table 1 Numerical results of Algorithm 1 and IAL with $subtol = 1e - 4$

ID	Algorithm 1				IAL			
	Res	*Rel*	*Init*	*Time*	*Res*	*Rel*	*Init*	*Time*
$m = 60, n = 100$	8.0e-9	3.5e-10	158	0.07	7.4e-9	3.2e-10	186	0.09
$m = 200, n = 300$	7.4e-9	1.1e-10	231	0.78	9.3e-9	1.4e-10	281	0.92
$m = 300, n = 500$	8.3e-9	7.8e-11	278	2.26	9.1e-9	8.5e-11	322	2.63
$m = 600, n = 1000$	7.5e-9	3.3e-11	300	10.73	7.8e-9	3.5e-11	374	13.78
$m = 1000, n = 1500$	8.7e-9	2.6e-11	284	35.12	7.6e-9	2.3e-11	327	46.28

Table 2 Numerical results of Algorithm 1 and IAL with $subtol = 1e - 6$

ID	Algorithm 1				IAL			
	Res	Rel	Init	Time	Res	Rel	Init	Time
$m = 60, n = 100$	9.0e-9	3.9e-10	86	0.05	6.2e-9	2.6e-10	130	0.08
$m = 200, n = 300$	6.0e-9	8.6e-10	140	0.48	9.2e-9	1.4e-10	174	0.58
$m = 300, n = 500$	8.8e-9	8.0e-11	185	1.61	9.5e-9	8.9e-11	215	1.78
$m = 600, n = 1000$	7.8e-9	3.6e-11	232	8.34	8.4e-9	3.7e-11	262	9.68
$m = 1000, n = 1500$	9.6e-9	2.7e-11	193	24.25	8.6e-9	2.6e-11	277	34.67

different dimensions. We observe that when the subproblem is solved with different 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:

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

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Let $m = 1500, n = 3000$, and A be generated by standard Gaussian distribution. Suppose that the original solution (signal) $x^* \in \mathbb{R}^n$ has only 150 non-zero elements which are generated by the Gaussian distribution $\mathcal{N}(0, 4)$ in the interval $[-2, 2]$ and that the noise ω is selected randomly with $\|\omega\| = 10^{-4}$,

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

Table 3 Numerical results of Algorithm 1 and IAL with $subtol = 1e - 8$

ID	Algorithm 1				IAL			
	Res	Rel	Init	Time	Res	Rel	Init	Time
$m = 60, n = 100$	8.5e-9	3.7e-10	32	0.03	8.7e-9	3.0e-10	37	0.03
$m = 200, n = 300$	8.3e-9	1.2e-10	85	0.28	8.4e-9	1.2e-10	100	0.34
$m = 300, n = 500$	9.8e-9	9.0e-11	95	0.84	9.5e-9	8.5e-11	121	1.00
$m = 600, n = 1000$	9.0e-9	4.2e-11	108	4.32	8.1e-9	3.3e-11	136	5.47
$m = 1000, n = 1500$	9.8e-9	2.8e-11	108	13.63	8.6e-9	2.4e-11	144	18.67

Table 4 Numerical results of Algorithm 2 and IAALM with $subtol = 1e - 8$

	β	0.01	0.05	0.1	0.5	1	1.5
Init	Al2	10	10	10	13	15	42
	IAALM	35	37	34	33	35	100+
Time	Al2	13.15	11.80	10.47	12.38	14.99	46.75
	IAALM	38.12	36.77	35.34	36.50	34.54	106.45
Res	Al2	4.18e-4	3.37e-4	3.18e-4	2.47e-4	4.74e-4	4.34e-4
	IAALM	4.91e-4	4.82e-4	4.75e-4	4.22e-4	4.21e-4	6.50e-3
Rel	Al2	1.31e-6	8.01e-7	6.35e-7	4.18e-7	9.09e-7	7.58e-2
	IAALM	9.05e-7	8.47e-7	9.19e-7	8.15e-7	9.62e-7	7.34e-2
SNR	Al2	1.17e+2	1.22e+2	1.24e+2	1.28e+2	1.21e+2	2.24e+1
	IAALM	1.20e+2	1.21e+2	1.21e+2	1.22e+2	1.20e+2	2.27e+1

Set parameters for Algorithm 2 (Al2) with $\alpha = 20$, $s = 1$, $M_k = s\beta$, and the parameters of IAALM ([18, Algorithm 1]) with $\gamma = 1$. Subproblems are solved by FISTA and the stopping condition is when

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

is satisfied or the number of iterations exceeds 100, where accuracy $subtol = 1e - 6$, $1e - 8$. We terminate all the methods when $\|Ax_k - b\| \leq 5 * 10^{-4}$. In each test, we calculate the residual error $res = \|Ax - b\|$, the relative error $rel = \frac{\|x - x^*\|}{\|x^*\|}$ and the signal-to-noise ratio

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

where x is the recovery signal.

In Table 4, we present the numerical results of Algorithm 2 and IAALM for various β. When $subtol = 1e - 6$, IAALM does not work well, we list the numerical results of Algorithm 2 in Table 5. Based on the *Rel* and *SNR*, it is seen that the sparse original signal is well restored when $\beta \leq 1$. This is also shown in Fig. 2.

Table 5 Numerical results of Algorithm 2 with $subtol = 1e - 6$

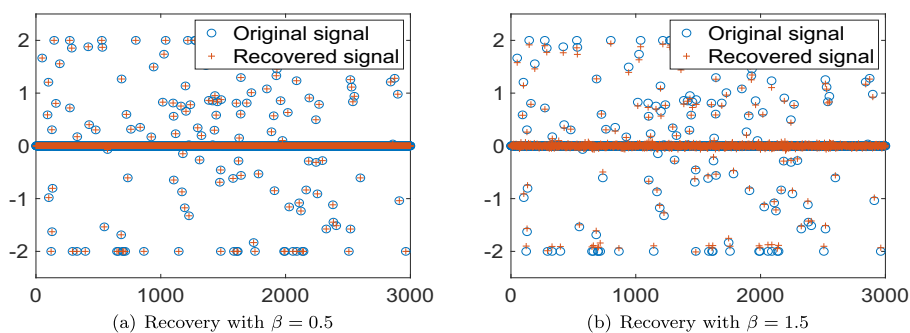| β	0.01	0.05	0.1	0.5	1	1.5
Init	25	22	17	13	18	100+
Time	8.29	8.11	7.52	8.65	12.74	36.48
Res	4.62e-4	4.89e-4	4.93e-4	4.36e-4	4.36e-4	1.50e-3
Rel	1.52e-6	1.68e-6	1.30e-6	7.17e-7	9.12e-7	8.10e-2
SNR	1.16e+2	1.15e+2	1.18e+2	1.22e+2	1.21e+2	2.18e+1

Fig. 2 Original sparse signal and the final estimated solution of Algorithm 2 with $subtol = 1e - 8$

4 Conclusion

In this paper, we propose two inertial accelerated primal-dual methods for solving linear equality constrained convex optimization problems. Assuming merely convexity, we show the inertial primal-dual methods own $\mathcal{O}(1/k^2)$ convergence rates even if the subproblem is solved inexactly. The numerical results demonstrate the validity and superior performance of our methods over some existing methods.

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. Boţ, R.I., Csetnek, E.R., Nguyen, D.K.: Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates. arXiv:2111.09370 (2021)
8. 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)
9. Candes, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal. Proc. Mag. **25**(2), 21–30 (2008)
10. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129–159 (2001)
11. Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Program. **64**(1-3), 81–101 (1994)

12. 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)
13. 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)
14. He, X., Hu, R., Fang, Y.P.: Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems. SIAM J. Control Optim. **59**(5), 3278–3301 (2021)
15. 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 (2010)
16. Huang, B., Ma, S., Goldfarb, D.: Accelerated linearized Bregman method. J. Sci. Comput. **54**(2), 428–453 (2013)
17. 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)
18. Kang, M., Kang, M., Jung, M.: Inexact accelerated augmented Lagrangian methods. Comput. Optim. Appl. **62**(2), 373–404 (2015)
19. Lin, Z., Li, H., Fang, C.: Accelerated optimization for machine learning. Springer, Singapore (2019)
20. Li, H., Lin, Z.: Accelerated alternating direction method of multipliers: An optimal $\mathcal{O}(1/k)$ nonergodic analysis. J. Sci. Comput. **79**(2), 671–699 (2019)
21. Ma, F., Ni, M.: A class of customized proximal point algorithms for linearly constrained convex optimization. Comput. Appl. Math. **37**(2), 896–911 (2018)
22. Madan, R., Lall, S.: Distributed algorithms for maximum lifetime routing in wireless sensor networks. IEEE Trans Wirel. Commun. **5**(8), 2185–2193 (2006)
23. Nedic, A., Ozdaglar, A.: Cooperative Distributed Multi-agent. Convex Optimization in Signal Processing and Communications. Cambridge University Press (2010)
24. 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)
25. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. Insov. Math. Dokl. **27**, 372–376 (1983)
26. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125–161 (2013)
27. Nocedal, J., Wright, S.: Numerical optimization. Springer Science and Business Media (2006)
28. Shi, G., Johansson, K.H.: Randomized optimal consensus of multi-agent systems. Automatica **48**(12), 3018–3030 (2012)
29. 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)
30. Tran-Dinh, Q., Zhu, Y.: Non-stationary first-Order primal-Dual algorithms with fast convergence rates. SIAM J. Optim. **30**(4), 2866–2896 (2020)
31. Van Den Berg, E., Friedlander, M.P.: Probing the Pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. **31**(2), 890–912 (2009)
32. Xu, Y.: Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. SIAM J. Optim. **27**(3), 1459–1484 (2017)
33. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. SIAM J. Imaging Sci. **1**(1), 143–168 (2008)
34. Zeng, X., Lei, J., Chen, J.: Dynamical primal-dual accelerated method with applications to network optimization. arXiv:1912.03690 (2019)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.