Fast primal-dual algorithms via dynamics for linearly constrained convex optimization problems

Xin He\textsuperscript{a}, Rong Hu\textsuperscript{b}, Ya-Ping Fang\textsuperscript{a}

\textsuperscript{a}Department of Mathematics, Sichuan University, Chengdu, Sichuan, P.R. China
\textsuperscript{b}Department of Applied Mathematics, Chengdu University of Information Technology, Chengdu, Sichuan, P.R. China

Abstract

By time discretization of a primal-dual dynamical system, we propose an inexact primal-dual algorithm, linked to the Nesterov’s acceleration scheme, for the linear equality constrained convex optimization problem. We also consider an inexact linearized primal-dual algorithm for the composite problem with linear constraints. Under suitable conditions, we show that these algorithms enjoy fast convergence properties. Finally, we study the convergence properties of the primal-dual dynamical system to better understand the accelerated schemes of the proposed algorithms. We also report numerical experiments to demonstrate the effectiveness of the proposed algorithms.

Key words: Linearly constrained convex optimization problem; primal-dual algorithm; inertial primal-dual dynamic; convergence rate; Nesterov’s acceleration.

1 Introduction

Let \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) be a proper convex function, \( A \in \mathbb{R}^{m \times n} \), and \( b \in \mathbb{R}^m \). Consider the linearly constrained convex optimization problem:

\[
\min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \ Ax = b.
\]

In some practical situations, the objective function \( f \) has the composite structure: \( f(x) = f_1(x) + f_2(x) \), where \( f_1 \) is a convex but possibly nondifferentiable function and \( f_2 \) is a convex differentiable function with Lipschitz gradient. In this case, the problem (1) can be rewritten as

\[
\min_{x \in \mathbb{R}^n} f_1(x) + f_2(x), \quad \text{s.t.} \ Ax = b.
\]

The problems (1) and (2) are basic models for many important applications arising in various areas, such as compressive sensing (Candès & Wakin, 2008), image processing (Zhang, Burger, Bresson, & Osher, 2010) and machine learning (Boyd et al., 2011; Lin, Li, & Fang, 2020).

Denote the KKT point set of the problem (1) by \( \Omega \).

For any \((x^\ast, \lambda^\ast) \in \Omega\), we have

\[
\begin{cases}
-A^T \lambda^\ast \in \partial f(x^\ast), \\
Ax^\ast = b,
\end{cases}
\]

where

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

Recall the Lagrangian function of the problem (1):

\[
\mathcal{L}(x, \lambda) = f(x) + \langle \lambda, Ax - b \rangle,
\]

where \( \lambda \in \mathbb{R}^m \) is the Lagrange multiplier. From (3) we have

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

A benchmark algorithm for the problem (1) is the augmented Lagrangian method (ALM):

\[
\begin{cases}
x_{k+1} \in \arg \min_x \ f(x) + \langle \lambda_k, Ax - b \rangle + \frac{\sigma}{2} \|Ax - b\|^2, \\
\lambda_{k+1} = \lambda_k + \sigma (Ax_{k+1} - b).
\end{cases}
\]

ALM plays a fundamental theoretical and algorithmic role in solving the problem (1) as well as the problem (2). Here, we mention some of nice works concerning fast convergence properties of ALM and its variants. By applying the Nesterov’s acceleration technique (Beck, & Teboulle, 2009; Nesterov, 1983) to
ALM, He, & Yuan (2010) developed an accelerated augmented Lagrangian method (AALM) for the problem (1), and proved that AALM enjoys an $O(1/k^2)$ convergence rate when $f$ is differentiable. When $f$ is non-differentiable, the $O(1/k^2)$ convergence rate of AALM was established in (Kang, Yun, Woo, & Kang, 2013). Kang, Kang, & Jung (2015) further proposed an inexact linearized primal-dual algorithm for the problems (1) and (2) with the $O(1/k^2)$ convergence rate under the strong convexity assumption of $f$. Huang, Ma, & Goldfarb (2013) considered an accelerated linearized Bregman method for solving the basis pursuit and related sparse optimization problems, and proved that it owns the $O(1/k^2)$ convergence rate. He, Hu, & Fang (2021c) proposed two inertial accelerated primal-dual algorithms for the problems (1) and (2) with the $O(1/k^2)$ convergence rate on the objective residual and the feasibility violation. By applying the linearization technique to ALM, Xu (2017) presented the linearized ALM for the problem (2):

$$\begin{align*}
\min_{x_{k+1}} & \quad f(x) + \langle \nabla f(x_k), x \rangle + \frac{\alpha}{2}\|Ax - b\|^2 + \frac{\beta}{2}\|x - x_k\|^2, \\
\lambda_{k+1} & = \lambda_k + \sigma(Ax_{k+1} - b),
\end{align*}$$

(4)
and showed that it enjoys the $O(1/k)$ ergodic convergence rate, where $\|x\|_P = \sqrt{x^TPx}$ with a positive semidefinite matrix $P$. By adapting the parameters in (4) during the iterations, Xu (2017) further proposed the accelerated linearized ALM method with the $O(1/k^2)$ convergence rates when $f_2$ has a Lipschitz gradient. For more results on fast ALM-type algorithms, we refer the reader to (He, Hu, & Fang, 2021b; Lin, Li, & Fang, 2020; Luo, 2021a,b).

In this paper, by time discretization of the following inertial primal-dual dynamical system

$$\begin{align*}
\dot{x}(t) + \bar{\sigma}\dot{z}(t) &= -\beta(t)\langle \nabla f(x(t)), x \rangle + \sigma A\lambda(t) + \epsilon(t), \\
\dot{\lambda}(t) &= \alpha \beta(t) A\lambda(t) + \delta\dot{z}(t) - b, \\
\end{align*}$$

(5)
where $t_0 > 0$, $\alpha > 0$, $\beta > 0$, $\beta: [t_0, +\infty) \to (0, +\infty)$ is a scaling coefficient, and $\epsilon: [t_0, +\infty) \to \mathbb{R}^n$ is a perturbation, we attempt to propose primal-dual algorithms with fast convergence properties for the problem (1) and the problem (2). We also study the convergence properties of the dynamic (5) which helps us to understand the acceleration schemes of the proposed algorithms. Throughout this paper, we always assume that the saddle point set $\Omega$ is nonempty.

The paper is organized as follows: In Section 2, by the time discretization of the dynamic (5), we propose an inexact fast primal-dual algorithm for the problem (1) and show its fast convergence properties. In Section 3, by applying the linearization technique to the algorithm in Section 2, we propose and study an inexact linearized primal-dual algorithm for the composite problem (2). Section 4 is devoted to the study of convergence properties of the inertial primal-dual dynamic (5). The numerical experiments are given in Section 5. Finally, we end the paper with a conclusion.

2 Fast primal-dual algorithm

In this section, we propose a new fast primal-dual algorithm from the time discretization of the inertial primal-dual dynamic (5). Set $\epsilon_k = k, x_k = x(t_k)$, $\lambda_k = \lambda(k)$, $\beta_k = \beta(k)$, $\epsilon_k = \epsilon(k)$, and $\bar{\sigma}x(k) = \frac{\alpha}{\kappa}\langle \nabla f(x_k+1) - Ax_k + \epsilon(k), x_k+1 - x_k \rangle$. Consider the following discretization scheme of (5) with a nondifferentiable function $f$:

$$\begin{align*}
x_{k+1} &= 2x_k + x_k - 1 + \frac{\alpha - \theta}{k}\langle x_{k+1} - x_k, x_k \rangle + \frac{\theta}{k}\langle x_k - x_{k-1} \rangle \\
\lambda_{k+1} &= \lambda_k + k\beta_k \langle Ax_{k+1} - b + \delta(k + 1 - \theta)A(x_{k+1} - x_k) \rangle.
\end{align*}$$

(6a)

(6b)

By computation, from (6) we obtain the following fast inexact primal-dual algorithm (Algorithm 1), where $\epsilon_k$ can be treated as a small perturbation. Now, we show that Algorithm 1 is equivalent to the time discretization scheme (6).

Algorithm 1 Fast inexact primal-dual algorithm for the problem (1)

Initialization: Choose $x_0 \in \mathbb{R}^n$, $\lambda_0 \in \mathbb{R}^m$, $\epsilon_0 = 0$. Set $x_1 = x_0$, $\lambda_1 = \lambda_0$. Choose parameters $\delta > 0$, $\alpha > 0$, $\theta \in \mathbb{R}$.

For $k = 1, 2, \ldots$

Step 1: Compute $\bar{x}_k = x_k + \frac{k - \alpha}{k + \alpha - \theta}\langle x_k - x_{k-1} \rangle$.

Step 2: Choose $\beta_k > 0$ and $\epsilon_k \in \mathbb{R}^n$. Set $\theta_k = k\beta_k(1 + \delta(k + 1 - \theta))$.

$$\eta_k = \frac{1}{1 + \delta(k + 1 - \theta)}(\delta(k + 1 - \theta)Ax_k + b).$$

Update

$$\begin{align*}
x_{k+1} &= \min_{x \in \mathbb{R}^n} f(x) + \frac{k + \alpha - \theta}{2k\beta_k}\|x - \bar{x}_k\|^2 \\
&\quad + \frac{\theta_k}{2}\|Ax - \eta_k\|^2 + \langle A^T\lambda_k - \epsilon_k, x \rangle.
\end{align*}$$

(6a)

Step 3:

$$\lambda_{k+1} = \lambda_k + k\beta_k \langle Ax_{k+1} - b + \delta(k + 1 - \theta)A(x_{k+1} - x_k) \rangle.$$

If a stopping condition is satisfied then return $(x_{k+1}, \lambda_{k+1})$

end

Proposition 1 Algorithm 1 is equivalent to the time discretization scheme (6).

Proof. The equation (6b) is directly derived from Step 3 of Algorithm 1. By using optimality condition, from Step 2 of Algorithm 1, we get

$$0 \in \partial f(x_{k+1}) + \frac{k + \alpha - \theta}{k\beta_k}(x_{k+1} - \bar{x}_k) + A^T(\theta_k(\lambda_{k+1} - \eta_k) + \lambda_k) - \frac{\epsilon_k}{\beta_k}.$$
which can be rewritten as
\[
\frac{k + \alpha - \theta}{k} (x_{k+1} - x_k) \leq -\beta_k (\partial f(x_{k+1}) + A^T (\partial_k (A x_{k+1} - \eta_k) + \lambda_k)) + \epsilon_k. \tag{7}
\]

It follows from Step 2 and Step 3 of Algorithm 1 that
\[
\begin{aligned}
\vartheta_k (A x_{k+1} - \eta_k) + \lambda_k &= (k \beta_k + \delta k (k + 1 - \theta) \beta_k) A x_{k+1} - \delta k (k + 1 - \theta) \beta_k A x_k - k \beta_k b + \lambda_k \\
&= \lambda_k + k \beta_k (A x_{k+1} - b + \delta (k + 1 - \theta) A (x_{k+1} - x_k)) \\
&= \lambda_{k+1},
\end{aligned}
\tag{8}
\]
and
\[
\begin{aligned}
\frac{k + \alpha - \theta}{k} (x_{k+1} - \bar{x}_k) &= \frac{k + \alpha - \theta}{k} (x_{k+1} - x_k) - \frac{k - \theta}{k} (x_k - x_{k-1}) \\
&= x_{k+1} - 2x_k + x_{k-1} + \frac{\alpha - \theta}{k} (x_{k+1} - x_k) + \frac{\theta}{k} (x_k - x_{k-1}). \tag{9}
\end{aligned}
\]
As a consequence of (7)-(9), the equation (6a) holds.

Then the sequence \(\{x_k, \lambda_k\}\) generated by Algorithm 1 satisfies the equation (6). Since the calculation process from above is reversible, from (6), we also can obtain Algorithm 1.

**Remark 1** The term \(\beta_k\) in (6) (Algorithm 1) and the \(\beta(t)\) in the dynamic (5) can be treated as scaling parameters, and it plays a key role in deriving the fast convergence of algorithms and dynamics. The importance of scaling parameters have been recognized in the design of accelerated algorithms for convex optimization problems (See e.g. Attouch, Chbani, & Riahi, 2018; Fazlyab, Koppel, Pucciado, & Ribeiro, 2017; Wibisono, Wilson, & Jordan, 2016).

**2.1 Convergence rate analysis**

Before discussing the convergence properties of Algorithm 1, we first recall the following two equalities:

\[
\frac{1}{2} \|x\|^2 - \frac{1}{2} \|y\|^2 = \langle x, x - y \rangle - \frac{1}{2} \|x - y\|^2, \tag{10}
\]
\[
2 \langle x - y, z \rangle = \|x - y\|^2 + \|x - z\|^2 - \|y - z\|^2, \tag{11}
\]
for any \(x, y, z \in \mathbb{R}^n\).

**Lemma 1** Let \(\{(x_k, \lambda_k)\}_{k \geq 1}\) be the sequence generated by Algorithm 1 and \((x^*, \lambda^*) \in \Omega\). Denote
\[
u_k := \frac{1}{k} (x_k - x^*) + (k - \theta)(x_k - x_{k-1}). \tag{12}
\]
Define the energy sequence
\[
E_k := E_k - \sum_{j=1}^k \langle u_j, (j - 1) \epsilon_{j-1} \rangle, \tag{13}
\]
with
\[
E_k := k (k + 1 - \theta) \beta_k (\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) + \frac{1}{2} \|u_k\|^2 + \frac{\alpha \delta - \delta - 1}{25^2} \|x_k - x^*\|^2 + \frac{1}{2 \delta} \|\lambda_k - \lambda^*\|^2. \tag{14}
\]

Then for any \(k \geq \max\{\theta - 1, 1\}\):
\[
E_k^* - E_k^* \leq \left( (k + 1)(k + 2 - \theta) \beta_k - k \left( k + 1 - \theta + \frac{1}{\delta} \right) \beta_k \right) \times (\mathcal{L}(x_k, \lambda) - \mathcal{L}(x^*, \lambda)) - \frac{1}{2 \delta} \|\lambda_k - \lambda^*\|^2 + \frac{(2\delta + 1 - \theta + 1)(1 + \delta - \alpha \delta)}{2 \delta^2} \|x_k - x^*\|^2. \tag{15}
\]

**Proof.** By the definition of \(\mathcal{L}\), we have
\[
\partial_x \mathcal{L}(x, \lambda) = \partial f(x) + A^T \lambda.
\]
This together with (12) and (6a) implies
\[
u_{k+1} - u_k = \left( \frac{1}{k} - \alpha + 1 \right) (x_{k+1} - x_k) + (\alpha - 1) (x_{k+1} - x_k) + (k + 1 - \theta) (x_{k+1} - x_k) - (k - \theta) (x_k - x_{k-1})
\]
\[
- \left( \frac{1}{k} - \alpha + 1 \right) (x_{k+1} - x_k) + k (x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha - \theta}{k} (x_{k+1} - x_k) + \frac{\theta}{k} (x_k - x_{k-1})
\]
\[
\in \left( \frac{1}{k} - \alpha + 1 \right) (x_{k+1} - x_k) - k \beta_k A^T (\lambda_{k+1} - \lambda^*)
\]
\[
- k \beta_k (\partial f(x_{k+1}) + A^T \lambda^*) + k \beta_k
\]
\[
= -k \beta_k \partial_x \mathcal{L}(x_{k+1}, \lambda^*) - k \beta_k A^T (\lambda_{k+1} - \lambda^*)
\]
\[
+ \frac{1 + \delta - \alpha \delta}{\delta} \|x_{k+1} - x_k\| + k \beta_k.
\]
Denote
\[
\xi_k := -\frac{1}{k \beta_k} \nu_{k+1} - u_k - A^T \lambda_{k+1} - \lambda^*
\]
\[
+ \frac{1 + \delta - \alpha \delta}{\delta} \|x_{k+1} - x_k\| \in \partial_x \mathcal{L}(x_{k+1}, \lambda^*). \tag{17}
\]
Then, combining (10) and (16), we have
\[
\frac{1}{2} \|u_{k+1}\|^2 - \frac{1}{2} \|u_k\|^2 = \langle u_{k+1}, u_k - u_k \rangle - \frac{1}{2} \|u_{k+1} - u_k\|^2
\]
\[
\leq -k \beta_k \langle \xi_{k+1}, \xi_k \rangle - k \beta_k \langle u_{k+1}, A^T (\lambda_{k+1} - \lambda^*) \rangle
\]
\[
+ \frac{1 + \delta - \alpha \delta}{\delta} \|u_{k+1} - x_{k+1} - x_k\| + \|u_{k+1}, k \epsilon_k\|. \tag{18}
\]
By (12), (17) and \(k + 1 - \theta \geq 0\), we get
\[
\langle u_{k+1}, \xi_k \rangle = \frac{1}{k} \langle x_{k+1} - x^*, \xi_k \rangle + (k + 1 - \theta) (x_{k+1} - x_k, \xi_k)
\]
\[
\geq \frac{1}{\delta} \|\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)\|
\]
\[
+ (k + 1 - \theta) \|\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x_k, \lambda^*)\|. \tag{19}
\]
It follows from (12) and (11) that
\[
\langle u_{k+1}, x_{k+1} - x_k \rangle = \frac{1}{\delta} \langle x_{k+1} - x^*, x_{k+1} - x_k \rangle \\
+ (k + 1 - \theta) \| x_{k+1} - x_k \|^2 \\
= \frac{1}{2\delta} (\| x_{k+1} - x^* \|^2 + \| x_{k+1} - x_k \|^2 - \| x_k - x^* \|^2) \\
+ (k + 1 - \theta) \| x_{k+1} - x_k \|^2
\]
This yields the desire results.

This together with (18) and (19) yields
\[
\frac{1}{2} \| u_{k+1} \|^2 - \frac{1}{2} \| u_k \|^2 \\
+ \frac{\alpha \delta - 1}{2\delta^2} (\| x_{k+1} - x^* \|^2 - \| x_k - x^* \|^2) \\
\leq -\frac{k \beta \delta}{\delta} (\mathcal{L}(x_{k+1}, x^*) - \mathcal{L}(x^*, x^*)) \\
- k(k + 1 - \theta) \beta k \| x_{k+1} - x_k \|^2 \\
- k \beta k \langle u_{k+1}, A^T (x_{k+1} - x^*) \rangle + \langle u_{k+1}, k \epsilon_k \rangle \\
+ \frac{(2\delta(k + 1 - \theta) + 1)(1 + \delta - \alpha \delta)}{2\delta} \| x_{k+1} - x_k \|^2.
\]
By Step 3 of Algorithm 1, \(Ax^* = b\), and (12), we get
\[
\frac{1}{2\delta} (\| x_{k+1} - x^* \|^2 - \| x_k - x^* \|^2) \\
= \langle x_{k+1} - x_k, k \epsilon_k \rangle \\
= \langle x_{k+1} - x_k, \frac{1}{\delta} (\lambda_{k+1} - \lambda_k) \rangle \\
= \frac{1}{2\delta} \| x_{k+1} - x_k \|^2.
\]
where the first equality is deduced from (10). It follows from (14) and (21) - (22) that
\[
\mathcal{E}^e_{k+1} = \mathcal{E}^e_{k+1} - \mathcal{E}^e_k - \langle u_{k+1}, k \epsilon_k \rangle \\
= (k + 1)(k + 2 - \theta) \beta k (\mathcal{L}(x_{k+1}, \lambda) - \mathcal{L}(x^*, \lambda)) \\
- k(k + 1 - \theta) \beta k \mathcal{L}(x_{k+1} - x_k) \\
+ \frac{1}{2\delta} \| u_{k+1} \|^2 - \frac{1}{2\delta} \| u_k \|^2 \\
+ \frac{\alpha \delta - 1}{2\delta^2} (\| x_{k+1} - x_k \|^2 - \| x_k - x_k \|^2) \\
+ \frac{1}{2\delta} (\| x_{k+1} - \lambda \|^2 - \| x_k - \lambda \|^2) - \langle u_{k+1}, k \epsilon_k \rangle \\
\leq (k + 1)(k + 2 - \theta) \beta k (1 - \theta + \frac{1}{\delta}) \beta k \\
\times (\mathcal{L}(x_{k+1}, \lambda) - \mathcal{L}(x^*, \lambda)) \\
+ \frac{(2\delta(k + 1 - \theta) + 1)(1 + \delta - \alpha \delta)}{2\delta^2} \| x_{k+1} - x_k \|^2 \\
- \frac{1}{2\delta} \| x_{k+1} - x_k \|^2.
\]
This yields the desire results.

To derive the convergence rates, we need the following scaling condition: there exists \(k_1 \in \mathbb{N}\) and \(k_1 \geq \theta - 1\) such that
\[
\beta_{k+1} \leq \frac{k(k + 1 - \theta + \frac{1}{\delta})}{k + 1}(1 + 2 - \theta) \beta_k, \quad \forall k \geq k_1.
\]
Now, we start to discuss the fast convergence properties of Algorithm 1 by the Lyapunov analysis approach.

**Theorem 1** Let \(\{\lambda_k\}_{k \geq 1}\) be the sequence generated by Algorithm 1 and \((x^*, \lambda^*) \in \Omega\). Assume that \(\frac{1}{\delta} \leq \alpha - 1\),
\[
\sum_{k=1}^{+\infty} k \| \epsilon_k \| < +\infty
\]
and the condition (23) holds. Then the dual sequence \(\{\lambda_k\}_{k \geq 1}\) is bounded and the following statements are true:
(i) \(\sum_{k=1}^{+\infty} (k(k + 1 - \theta + \frac{1}{\delta}) \beta_k - (k + 1)(k + 2 - \theta) \beta_{k+1}) (\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty\);
(ii) \(\sum_{k=1}^{+\infty} \| \lambda_{k+1} - \lambda_k \|^2 < +\infty\);
(iii) When \(\frac{1}{\delta} < \alpha - 1\), the primal sequence \(\{x_k\}_{k \geq 1}\) is bounded and
\[
\| x_{k+1} - x_k \| = O \left( \frac{1}{k} \right), \quad \sum_{k=1}^{+\infty} k \| x_{k+1} - x_k \|^2 < +\infty;
\]
(iv) When \(\lim_{k \to +\infty} k^2 \beta_k = +\infty\):
\[
\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O \left( \frac{1}{k^2 \beta_k} \right).
\]

**Proof.** From assumptions, we can get \(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \leq 0, 1 + \delta - \alpha \delta \leq 0\) and \((k + 1)(k + 2 - \theta) \beta_{k+1} - k(k + 1 - \theta + \frac{1}{\delta}) \beta_k \leq 0\) for any \(k \geq k_1\). Then, it follows from Lemma 1 that
\[
\mathcal{E}^e_k \leq \mathcal{E}^e_{k_1}, \quad \forall k \geq k_1.
\]
By the definitions of \(\mathcal{E}_k\) and \(\mathcal{E}_k^e\), we get
\[
\frac{1}{2} \| u_k \|^2 \leq \mathcal{E}_k = \mathcal{E}_k^e + \sum_{j=1}^{k} (u_j, (j - 1) \epsilon_{j-1}) \\
\leq |\mathcal{E}_k^e| + \sum_{j=1}^{k} (j - 1) \| u_j \| \| \epsilon_{j-1} \|
\]
for any \(k \geq k_1\). It follows from Lemma 4 that
\[
\sup_{k \geq k_1} \| u_k \| \leq \sqrt{2 |\mathcal{E}_k^e|} + \sum_{j=1}^{+\infty} j \| \epsilon_j \| < +\infty.
\]
This yields
\[
\sup_{k \geq k_1} \mathcal{E}_k \leq \mathcal{E}_k^e + \sup_{k \geq k_1} \| u_k \| \sum_{j=1}^{+\infty} j \| \epsilon_j \| < +\infty.
\]
and
\[ \inf_{k \geq k_1} \mathcal{E}'_k \geq - \sup_{k \geq k_1} \| u_k \| \sum_{j=1}^{+\infty} j \| \epsilon_j \| > -\infty. \]

So \( \{ \mathcal{E}'_k \}_{k \geq 1} \) and \( \{ \mathcal{E}'_k \}_{k \geq 1} \) are both bounded. Summing the inequality (15) over \( k = 1, 2, \cdots \), the boundedness of \( \{ \mathcal{E}'_k \}_{k \geq 1} \) implies (i), (ii) and
\[ (1 + \delta - \alpha \delta) \sum_{k=1}^{+\infty} k \| x_{k+1} - x_k \|^2 < +\infty. \]

Since \( \{ \mathcal{E}'_k \}_{k \geq 1} \) is bounded, it follows from (14) and assumptions that \( \{ \| u_k \| \}_{k \geq 1} \), \( \{ (\alpha \delta - \delta - 1) \| x_k - x^* \| \}_{k \geq 1} \) and \( \{ \| \lambda_k - \lambda^* \| \}_{k \geq 1} \) are bounded. So the dual sequence \( \{ \lambda_k \}_{k \geq 1} \) is bounded. And when \( \frac{1}{\delta} < \alpha - 1 \): the primal sequence \( \{ x_k \}_{k \geq 1} \) is bounded,
\[ \sup_{k \geq 1} (k + 1 - \theta) \| x_{k+1} - x_k \| \leq \sup_{k \geq 1} \| u_{k+1} \| + \sup_{k \geq 1} \frac{1}{\delta} \| x_{k+1} - x^* \| \]
\[ < +\infty, \]
which means
\[ \| x_{k+1} - x_k \| = \mathcal{O} \left( \frac{1}{k} \right). \]

Since \( \{ \mathcal{E}'_k \}_{k \geq 1} \) is bounded, from (14) we get
\[ \sup_{k \geq 1} k (k + 1 - \theta) \beta_k (\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty, \]
which yields (iv).

**Remark 2** If the condition (23) is replaced by the strengthened one: there exists \( \kappa \in (0, \frac{1}{\delta}) \) such that
\[ (k + 1)(k + 2 - \theta) \beta_{k+1} \leq k (k + 1 - \theta + \kappa) \beta_k \quad \forall k \geq k_1, \]
then by Theorem 1 (i) we have
\[ \sum_{k=1}^{+\infty} k \beta_k (\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty. \]

**Remark 3** Combining (23) with \( \frac{1}{\delta} \leq \alpha - 1 \), we get
\[ \beta_{k+1} \leq \frac{k(k + \alpha - \theta)}{(k + 1)(k + 2 - \theta)} \beta_k, \]
which is just the assumption \( (H_{\beta, \theta}) \) used by (Attouch, Chbani, & Riahi, 2018) for convergence rate analysis of inertial proximal algorithms for the unconstrained optimization problems. Attouch, Chbani, & Riahi. (2018) proposed inertial proximal algorithms with scaling \( \beta_k \) for the unconstrained optimization problem, and proved the \( \mathcal{O}(k^2) \) convergence rate under the above condition. See (Attouch, Chbani, & Riahi., 2018, Theorem 3.1 and Theorem 7.1). Theorem 1 can be viewed an extension of (Attouch, Chbani, & Riahi., 2018, Theorem 3.1 and Theorem 7.1) to the linear equality constrained optimization problem.

**Remark 4** When the problem dimension \( n \) is very large, finding an exact solution of the subproblem in Step 2 of Algorithm 1 may be computationally expensive. The assumption \( \sum_{k=1}^{+\infty} k \| \epsilon_k \| < +\infty \) on perturbation sequence \( \{ \epsilon_k \}_{k \geq 1} \) shows that we can solve the subproblem inexactly by finding an approximate solution.

### 2.2 Improved convergence results

Under the condition (23) and \( \lim_{k \to +\infty} k^2 \beta_k = +\infty \), by Theorem 1 we only obtain \( \mathcal{O}(1/k^2) \) convergence rate of \( \mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) \). Can we choose appropriate parameters so that the objective residual and the feasibility violation have the same convergence rate as the \( \mathcal{O}(1/k^2) \)? The answer is yes. In the next, by choosing a specific scaling \( \beta_k \), we will show the improved convergence results, and the parameter \( \theta \) plays the important roles in convergence analysis. Assume that the scaling \( \beta_k \) satisfies
\[ \theta \geq 2, \quad \beta_{k+1} = \frac{k}{k + 2 - \theta} \beta_k, \quad \forall k \geq k_1, \quad (25) \]
where \( k_1 > \theta - 2 \). Then, we can obtain the following results.

**Lemma 2** Assume that (25) holds. Then the following results hold:

(i) When \( \theta = 2 \): \( \beta_k := \beta_{k_1} \) for all \( k \geq k_1 \).

(ii) When \( \theta > 2 \): for any \( \rho \in [0, \theta - 2) \), there exists \( \mu_1 > 0, \mu_2 > 0 \) such that
\[ \mu_1 k^\rho \leq \beta_k \leq \mu_2 k^{\theta - 2} \]
for \( k \) large enough.

**Proof.** The result (i) is directly derived from (25).

Now consider the case \( \theta > 2 \). Let \( \rho \in [0, \theta - 2) \). It’s easy to verify that
\[ \left(1 + \frac{1}{k}\right)^\rho - 1 \sim \frac{\rho}{k} \quad \text{and} \quad \frac{2 - \theta}{k + 2 - \theta} \sim \frac{2 - \theta}{k}, \]
when \( k \) large enough. Then for any \( \rho \in [0, \theta - 2) \), there exists \( k_2 \geq k_1 \) such that
\[ \left(1 + \frac{1}{k}\right)^\rho - 1 \leq \frac{\theta - 2}{k + 2 - \theta} \]
for all \( k \geq k_2 \). This means that for any \( \mu_1 > 0, \mu_1 k^\rho \leq \frac{\mu_1 k}{k + 2 - \theta} k^\rho, \quad \forall k \geq k_2. \]
Take \( \mu_1 = \frac{\beta_k}{k_2^2} \). It follows from (25) and (26) with \( k = k_2 \) that
\[
\mu_1 (k_2 + 1)^p \leq \beta_{k+1}.
\] (27)
Multiply both sides of (27) by \( \frac{k_2 + 1}{k_2 + 3} \), and combine (25) and (26) to get
\[
\mu_1 (k_2 + 2)^p \leq \beta_{k+2}.
\]
By induction, we obtain
\[
\beta_k \geq \mu_1 k^p \quad \forall k \geq k_2 + 1.
\] (28)
Let \( \lceil \theta \rceil \) represent the biggest integer not more than \( \theta \). By (25), we have
\[
\beta_k = \beta_{k_1} \prod_{j=k_1}^{k-1} \left( 1 + \frac{\theta - 2}{j + 2 - \theta} \right) \quad \forall k > k_1.
\]
It yields
\[
\ln \beta_k \leq \ln \beta_{k_1} + \sum_{j=[k_1+2-\theta]}^{[k+1-\theta]} \ln \left( 1 + \frac{\theta - 2}{j} \right)
\] (29)
\[
\leq \ln \beta_{k_1} + (\theta - 2) \sum_{j=[k_1+2-\theta]}^{[k+1-\theta]} \frac{1}{j}.
\]
for all \( k \geq \lceil k_1 + 2 - \theta \rceil \), where the last inequality follows from \( \ln(1 + x) \leq x \) as \( x > -1 \). Since \( \theta \geq 2 \), by classical comparison between series and integral,
\[
\sum_{j=[k_1+2-\theta]}^{[k+1-\theta]} \frac{1}{j} \leq \sum_{j=1}^{k} \frac{1}{j} \leq \int_{1}^{k} \frac{1}{t} dt = \ln k,
\]
this together with (29) implies
\[
\beta_k \leq \beta_{k_1} k^{\theta - 2} \quad \forall k \geq \lceil k_1 + 2 - \theta \rceil.
\]
Combining this and (28), we get (ii).

(iii) If \( \theta > 2 \), then for any \( \rho \in [2, \theta) \),
\[
|f(x_k) - f(x^*)| = O \left( \frac{1}{k^{2\beta_k}} \right), \quad \|Ax_k - b\| = O \left( \frac{1}{k^{2\beta_k}} \right).
\]

Proof. Since \( \theta \leq \frac{1}{\beta} \), it follows from (25) that the inequality (23) holds, so all results in Theorem 1 are true. It follows from Step 3 of Algorithm 1 that
\[
\lambda_{k+1} - \lambda_k = \sum_{j=k_1}^{k} (\lambda_{j+1} - \lambda_j)
\]
\[
= \sum_{j=k_1}^{k} j \beta_j (Ax_{j+1} - b + \delta(j + 1 - \theta)A(x_{j+1} - x_j))
\]
\[
= \sum_{j=k_1}^{k} \left( (\delta(j + 1 - \theta) + 1)j \beta_j (Ax_{j+1} - b) - (\delta(j + 1 - \theta) + 1)j \beta_j (Ax_j - b) \right)
\]
\[
= \sum_{j=k_1}^{k} \left( (\delta(j + 1 - \theta) + 1)j \beta_j (Ax_{j+1} - b) - (\delta(k_1 + 1 - \theta) + 1)k \beta_k (Ax_{k+1} - b) \right)
\]
\[
= \sum_{j=k_1}^{k} \left( (\delta(j + 1 - \theta) + 1)j \beta_j (Ax_{j+1} - b) - (\delta(j + 1 - \theta) + 1)j \beta_j (Ax_j - b) \right)
\]
\[
+ (1 - \delta) \sum_{j=k_1}^{k} (j - 1) \beta_{j-1} (Ax_j - b)
\]
for any \( k \geq k_1 \), where the last equality follows from (25). Then for any \( k \geq k_1 \), we have
\[
\left\| \left( (\delta(k_1 + 1 - \theta) + 1)k \beta_k (Ax_{k+1} - b) + (1 - \delta) \sum_{j=k_1}^{k} (j - 1) \beta_{j-1} (Ax_j - b) \right) \right\|
\]
\[
\leq \left\| (\delta(k_1 + 1 - \theta) + 1)k \beta_k (Ax_{k+1} - b) \right\| + \|\lambda_{k+1} - \lambda_k\|.
\] (30)
From Theorem 1, we know that dual sequence \( \{\lambda_k\}_{k \geq 1} \) is bounded. Combining this and (30), there exists a constant \( C > 0 \) such that
\[
\left\| (\delta(k_1 + 1 - \theta) + 1)k \beta_k (Ax_{k+1} - b) + (1 - \delta) \sum_{j=k_1}^{k} (j - 1) \beta_{j-1} (Ax_j - b) \right\|
\]
\[
\leq C,
\] (31)
for any \( k \geq k_1 \). When \( \theta = \frac{1}{\beta} \), (31) yields
\[
\|Ax_{k+1} - b\| \leq \frac{C}{\delta(k_1 + 1 - \theta) + 1)k \beta_k}
\]
When \( \theta < \frac{1}{\beta} \), (31) yields
\[
\left\| (\delta(k_1 + 1 - \theta) + 1)k \beta_k (Ax_{k+1} - b) + (1 - \delta) \sum_{j=k_1}^{k} (j - 1) \beta_{j-1} (Ax_j - b) \right\|
\]
\[
\leq \frac{1}{1 - \delta} \left( C + \left\| \sum_{j=1}^{k_1} (j - 1) \beta_{j-1} (Ax_j - b) \right\| \right).
\] (32)
Denote \( c := \frac{1}{1 - \delta} \left( C + \left\| \sum_{j=1}^{k_1} (j - 1) \beta_{j-1} (Ax_j - b) \right\| \right) < \infty \). Applying Lemma 5 with \( a_k = (k - 1) \beta_{k-1} (Ax_k - b) \)
b) and $\zeta = 1 + \frac{\delta}{1 - \delta}$, we have

$$\left\| \sum_{j=1}^{k} (j-1)\beta_{j-1}(Ax_j - b) \right\| \leq c.$$  

Combining this and (32), we obtain

$$\left\| \frac{\delta (k+1 - \theta) + 1}{1 - \delta} \kappa \beta_k (Ax_{k+1} - b) \right\| \leq 2c,$$

and then

$$\|Ax_{k+1} - b\| \leq \frac{2(1 - \delta)c}{(\delta (k+1 - \theta) + 1)\kappa \beta_k}.$$  

From the above discussion and (25), when $\theta \leq \frac{1}{2}$:

$$\|Ax_k - b\| = O\left(\frac{1}{k^2 \beta_k}\right).$$

This together with Theorem 1 (iv) implies

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

$$= O\left(\frac{1}{k^2 \beta_k}\right).$$

Then from Lemma 2, we obtain (i) and (ii).

Remark 5 Take $\theta = 2, \delta \in \left[\frac{1}{\alpha}, \frac{1}{2}\right]$, and $\beta_k \equiv \beta > 0$. When $\alpha \geq \frac{1}{2} + 1 \geq 3$, Step 1 of Algorithm 1 can be regarded as the Nesterov’s acceleration scheme. By Theorem 2 (i), we obtain $|f(x_k) - f(x^*)| = O\left(\frac{1}{k^2}\right)$. In this sense, our results extend the classic results concerning the Nesterov’s accelerated gradient algorithms (Beck & Teboulle, 2009; Nesterov, 2013; Su, Boyd, & Candés, 2016) for the unconstrained optimization problem to the problem (1).

Remark 6 Under the assumption that $f$ is strongly convex and twice continuously differentiable with Lipschitz continuous Hessian, Facchinei, Kanzow, and Varga (2005) proposed an $O(\sqrt{k})$ convergence rate algorithm for the problem (1). Some results on the $O(\sqrt{k})$ convergence rate for the unconstrained optimization problem can be found in (Attouch, Chbani, & Riahi, 2018; Wibisono, Wilson, & Jordan, 2016; Wilson, Recht, & Jordan, 2021).

2.3 Further discussions

Let $I_d$ be the identity matrix and $\mathbb{S}_+(n)$ be the set of all positive semidefinite matrices in $\mathbb{R}^{n \times n}$. For any $M_1, M_2 \in \mathbb{S}_+(n)$, denote

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

It is easy to verify that for any $M \in \mathbb{S}_+(n)$,

$$\frac{1}{2}\|x\|_M^2 - \frac{1}{2}\|y\|_M^2 = \langle x, M(x-y) \rangle - \frac{1}{2}\|x-y\|_M^2. \quad (33)$$

Then, we can replace the subproblem of step 2 with

$$x_{k+1} \in \arg \min_{x} f(x) + \frac{k + \alpha - \theta}{2k \beta_k} \|x - x_k\|_M^2$$

$$+ \frac{\kappa}{2} \|Ax - \eta_k\|^2 + \langle A^T \lambda_k - \epsilon_k, x \rangle,$$

where $M \geq \kappa I$ for some $\kappa > 0$. Redefine (14) as

$$\mathcal{E}_k := k(1 - \theta)\beta_k \left(\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)\right) + \frac{1}{2}\|u_k\|_M^2$$

$$+ \frac{\alpha \delta - \delta - 1}{\kappa} \|x_k - x^*\|_M^2 + \frac{1}{2\kappa} \|\lambda_k - \lambda^*\|^2.$$

Through the arguments similar to the ones in Theorem 1 and Theorem 2, we can get the same convergence rate results. In particular, when the perturbation $\epsilon_k \equiv 0$ (which means that the subproblem can be solved with exact or high precision), we can take $\kappa = 0$. In numerical experiments, we can update the scaling $\beta_k$ by

$$\beta_{k+1} = \begin{cases} \beta_k, & k \leq \theta - 2; \\ \frac{1}{k+2}\beta_k, & k > \theta - 2. \end{cases} \quad (34)$$

3 Inexact linearized primal-dual algorithm

In this section, applying the linearization technique in the linearized ALM (4) to Algorithm 1, we propose an inexact linearized primal-dual algorithm (Algorithm 2) with constant scaling to solve the composite optimization problem (2).

Assumption (H): $\Omega \neq 0, f_1$ is a proper convex function, and $f_2$ is a proper convex differentiable function with an $L$-Lipschitz continuous gradient, i.e.,

$$\|\nabla f_2(x) - \nabla f_2(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

equivalently,

$$f_2(x) \leq f_2(y) + \langle \nabla f_2(y), x - y \rangle + \frac{L}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \quad (35)$$

Note that the Lagrangian function $\mathcal{L}(x, \lambda)$ of the problem (2) becomes

$$\mathcal{L}(x, \lambda) = f_1(x) + f_2(x) + \langle \lambda, Ax - b \rangle.$$  

By arguments similar to those in the proof of Proposition 1, we have the following result.

Lemma 3 The sequence $\{(x_k, \lambda_k)\}_{k \geq 1}$ generated by Algorithm 2 satisfies

$$M \left( x_{k+1} - x_k + \frac{\alpha - 2}{k} (x_k - x_{k-1}) + \frac{2}{k} (x_k - x_{k-1}) \right) \in -\beta \delta f_1(x_k) + f_2(x_k) + A^T \lambda_{k+1} + \epsilon_k. \quad (36)$$
Algorithm 2 Inexact linearized primal-dual algorithm for the problem (2)

Initialization: Choose \(x_0 \in \mathbb{R}^n\), \(\lambda_0 \in \mathbb{R}^m\), \(\epsilon_0 = 0\). Set \(x_1 = x_0\), \(\lambda_1 = \lambda_0\). Choose parameters \(\alpha > 0, \delta > 0, \beta > 0, M \in \mathbb{S}_+(n)\).

For \(k = 1, 2, \ldots\) do

Step 1: Compute \(\bar{x}_k = x_k + \frac{k-2}{k+\alpha} (x_k - x_{k-1})\) and \(\hat{x}_k = x_k + \frac{k - 2}{k + \frac{1}{\delta} - 1} (x_k - x_{k-1})\).

Step 2: Choose \(\epsilon_k \in \mathbb{R}^n\). Set \(\vartheta_k = \beta k (\delta k - \delta - 1)\), \(\eta_k = \frac{1}{\beta k} (\delta k - \delta - 1)\).

Update

\[
x_{k+1} = \arg\min_{x \in \mathbb{R}^n} f_1(x) + \frac{k + \alpha - 2}{2\beta k} \|x - \bar{x}_k\|_M^2 + \frac{\vartheta_k}{2} \|A x - \eta_k\|^2 + \frac{1}{\beta k} \|\lambda_k - \lambda^*\|^2 + \frac{\epsilon_k}{\beta} \langle \lambda_k, \xi \rangle.
\]

Step 3:

\[
\lambda_{k+1} = \lambda_k + \beta k (A x_{k+1} - b + \delta (k-1) A (x_{k+1} - x_k)).
\]

If a stopping condition is satisfied then

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

end

Theorem 3 Assume that Assumption (H) holds, \(2 \leq \frac{1}{\delta} \leq 1\), \(M \succeq \beta \Lambda I d\) 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 the dual sequence \(\{\lambda_k\}_{k \geq 1}\) is bounded and the following statements are true:

(i) When \(\frac{1}{\delta} > 2\): \(\sum_{k=1}^{+\infty} k\beta_k (\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty\);

(ii) \(\sum_{k=1}^{+\infty} \|\lambda_{k+1} - \lambda_k\|^2 < +\infty\);

(iii) When \(\frac{1}{\delta} < \alpha - 1\): the primal sequence \(\{x_k\}_{k \geq 1}\) is bounded and

\[
\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right), \quad \sum_{k=1}^{+\infty} k\|x_{k+1} - x_k\|_M^2 < +\infty;
\]

(iv) Convergence rate:

\[
|f(x_k) + g(x_k) - f(x^*) - g(x^*)| = \mathcal{O}\left(\frac{1}{k^2}\right),
\]

\[
\|Ax_k - b\| = \mathcal{O}\left(\frac{1}{k^2}\right).
\]

Proof. Define sequences \(E_k, E_k^*, u_k\) as

\[
\begin{aligned}
E_k &= \beta k (k - 1) \langle L(x_k, \lambda^*) - L(x^*, \lambda^*) \rangle + \frac{1}{2} \|u_k\|_M^2 + \frac{\alpha \delta - \delta - 1}{2\delta^2} \|x_k - x^*\|_M^2 + \frac{1}{2\delta^2} \|\lambda_k - \lambda^*\|^2, \\
E_k^* &= E_k - \sum_{j=1}^{k} (u_j, (j - 1)\epsilon_{j-1}), \\
u_k &= \frac{1}{\delta} \langle x_k - x^* \rangle + (k - 2) \langle x_k - x_{k-1} \rangle.
\end{aligned}
\]

It is easy to verify that

\[
u_{k+1} - u_k = (k + 1) \langle x_{k+1} - \hat{x}_k \rangle.
\]

Denote \(\mathcal{L}_{f_1}(x) := f_1(x) + \langle \lambda^*, Ax - b \rangle\).

Then

\[
\mathcal{L}(x, \lambda^*) = \mathcal{L}_{f_1}(x) + f_2(x), \quad \partial \mathcal{L}_{f_1}(x) = \partial f_1(x) + \Lambda^T \lambda^*.
\]

This together with (36) and (37c) implies

\[
M(u_{k+1} - u_k) \
\in \frac{1 + \delta - \alpha \delta}{\delta} M(x_{k+1} - x_k) - \beta k \Lambda^T (\lambda_{k+1} - \lambda^*) + k \epsilon_k \\
- \beta k (\partial f_1(x_{k+1}) + \langle \lambda_{k+1} - \lambda^* \rangle + \Lambda^T \lambda^*) + k \epsilon_k \\
= \beta k (\partial \mathcal{L}_{f_1}(x_{k+1}) - \beta k \Lambda^T (\lambda_{k+1} - \lambda^*) - \beta k \langle \lambda_{k+1} - \lambda^* \rangle + k \epsilon_k \\
+ k \epsilon_k + \frac{1 + \delta - \alpha \delta}{\delta} M(x_{k+1} - x_k).
\]

Denote

\[
\xi_k := \frac{1}{\beta k} M(u_{k+1} - u_k) - \beta k \Lambda^T (\lambda_{k+1} - \lambda^*) - \nabla f_2(\hat{x}_k) \\
+ \frac{\epsilon_k}{\beta} + \frac{1 + \delta - \alpha \delta}{\delta} \beta k M(x_{k+1} - x_k).
\]

It follows from (33), (38) and (39) that

\[
\begin{aligned}
\frac{1}{2} \|u_{k+1}\|_M^2 - \frac{1}{2} \|u_k\|_M^2 &\geq \langle u_{k+1}, M(u_{k+1} - u_k) \rangle - \frac{1}{2} \|u_{k+1} - u_k\|_M^2 \\
&= -\beta k \langle u_{k+1}, \xi_k \rangle - \beta k \langle u_{k+1}, \Lambda^T (\lambda_{k+1} - \lambda^*) \rangle \\
&- \beta k \langle u_{k+1}, \nabla f_2(\hat{x}_k) \rangle + \langle u_{k+1}, k \epsilon_k \rangle \\
&+ \frac{1 + \delta - \alpha \delta}{\delta} \langle u_{k+1}, M(x_{k+1} - x_k) \rangle \\
&- \frac{\delta}{2\delta^2} \|\lambda_k - \lambda^*\|^2 - \frac{(\delta k - 1)^2 + 1}{2\delta^2} \|x_{k+1} - \hat{x}_k\|_M^2.
\end{aligned}
\]
Combining (40) with (37c), we get
\[
(u_{k+1}, \xi_k) \geq \frac{1}{\delta} (\mathcal{L}_f (x_{k+1}) - \mathcal{L}_f (x^*))
+ \frac{L}{2} \|x_{k+1} - \check{x}_k\|^2.
\] (42)

Since \( f_2 \) has an \( L \)-Lipschitz continuous gradient, it follows from (35) that
\[
f_2(x_{k+1}) \leq f_2(\check{x}_k) + \langle \nabla f_2(\check{x}_k), x_{k+1} - \check{x}_k \rangle + \frac{L}{2} \|x_{k+1} - \check{x}_k\|^2.
\] (43)

By the convexity of \( f_2 \), we have
\[
\langle \nabla f_2(\check{x}_k), x_{k+1} - \check{x}_k \rangle = \langle \nabla f_2(\check{x}_k), x_{k+1} - x^* \rangle + \langle \nabla f_2(\check{x}_k), x^* - \check{x}_k \rangle
\leq \langle \nabla f_2(\check{x}_k), x_{k+1} - x^* \rangle + f_2(x^*) - f_2(\check{x}_k).
\] (44)

It follows from (43)-(45) that
\[
\langle \nabla f_2(\check{x}_k), x_{k+1} - x^* \rangle \geq f_2(x_{k+1}) - f_2(x^*) - \frac{L}{2} \|x_{k+1} - \check{x}_k\|^2.
\]
and
\[
\langle \nabla f_2(\check{x}_k), x_{k+1} - x_k \rangle \geq f_2(x_{k+1}) - f_2(x_k) - \frac{L}{2} \|x_{k+1} - \check{x}_k\|^2.
\]

Then by computation, we have
\[
\langle u_{k+1}, \nabla f_2(\check{x}_k) \rangle
= \frac{1}{\delta} \langle f_2(x_{k+1}) - f_2(x_k), x_{k+1} - x_k \rangle - \frac{L}{2} \delta (k+1) \|x_{k+1} - \check{x}_k\|^2.
\] (46)

By computation similarto (20), we get
\[
\langle u_{k+1}, M(x_{k+1} - x_k) \rangle
= \frac{1}{2\delta} (\|x_{k+1} - x^*\|^2_M - \|x_k - x^*\|^2_M)
+ (k-1) \frac{1}{2\delta} \|x_{k+1} - x_k\|^2_M.
\] (47)

It follows that
\[
\frac{1}{2} \|u_{k+1}\|_M^2 - \frac{1}{2} \|u_k\|_M^2 + \frac{\alpha \delta - \frac{1}{\delta}}{2\delta^2} \|x_{k+1} - x^*\|^2_M - \|x_k - x^*\|^2_M
\leq \frac{\beta \delta}{\delta} (\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*))
\leq \beta \delta (k+1) (\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x_k, \lambda^*))
\leq \beta \delta (k+1) (\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x_k, \lambda^*))
+ \frac{\beta \delta L}{\delta} (k+1) \|x_{k+1} - x_k\|^2_M
\] (48)

where the first inequality follows from (41), (42), and (46); the second inequality follows from (47) and \( M \succ \beta \delta L Id \succ \frac{\beta \delta L}{\delta} Id \). It follows from (22), (37), and (48) that
\[
\mathcal{E}_{k+1} - \mathcal{E}_k \leq \left( 2 - \frac{1}{\delta} \right) \beta \delta \|x_{k+1} - x^*\|^2_M
+ \frac{2 \beta \delta (k+1) (1 + \alpha \delta)}{\delta^2} \text{ and } \mathcal{E}_k \leq \mathcal{E}_1 \quad \forall \ k \geq 1.
\] (49)

Note that \( 2 \leq \frac{1}{\delta} < \alpha - 1 \). From (49), we have
\[
\mathcal{E}_k \leq \mathcal{E}_1 \quad \forall \ k \geq 1.
\]

By using \( M_k \succ \beta L Id \) and arguments similar to those in Theorem 1, we know \{\mathcal{E}_k\}_{k \geq 1} and \{\mathcal{E}_k\}_{k \geq 1} are both bounded. As a result, we get the boundedness of \{\lambda_k\}_{k \geq 1} (i), (ii), (iii), and
\[
\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*)
= f(x_k) + g(x_k) - f(x^*) - g(x^*) + \langle \lambda^*, A x_k - b \rangle
= O \left( \frac{1}{k^2} \right).
\]

Then by arguments similar to those in Theorem 2, we obtain (iv).

4 Inertial primal-dual dynamic with scaling

The importance of dynamical systems has been recognized as efficient tools for solving optimization problems in the literature. Dynamical systems can not only give more insights into existing numerical methods for optimization problems but also lead to other possible numerical algorithms by discretization (See (Jordan, 2018; Liang & Yin, 2019; Su, Boyd, & Candès, 2016; Wilson, Recht, & Jordan, 2021)). Su, Boyd, & Candès
(2016) showed that the second-order dynamical system

$$(AVD_\alpha) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla \Phi(x(t)) = 0$$

with $\alpha = 3$ can be understood as the continuous limit of the Nesterov’s accelerated gradient algorithm (Nesterov, 1983) and the FISTA algorithm (Beck, Teboulle, 2009) for the problem

$$\min \Phi(x),$$

where $\Phi(x)$ is a differentiable convex function. Attouch, Chbani, & Riahi (2015) showed that suitable discretization schemes of the perturbed second-order dynamical system with scaling

$$(AVD_{\alpha,\beta}) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) \nabla \Phi(x(t)) = \epsilon(t)$$

may lead to inertial proximal algorithms for the problem (50), with fast convergence properties. From a variational perspective, Wibisono, Wilson, & Jordan (2016) showed that the Nesterov’s acceleration technique and many of its generalizations for the unconstrained optimization problem (50) can be viewed as a systematic way to go from the continuous-time curves generated by a Bregman Lagrangian to a family of discrete-time accelerated algorithms.

In the past few years, primal-dual dynamical system methods have been successful in solving constrained convex optimization problem (1), see e.g. (Feijer & Paganini, 2010; He, Hu, & Fang, 2011a; Kia, Cortés, & Martínez, 2015; Tang, Qu, & Li, 2020; Wang et al., 2021; Zhu, Yu, Wen, & Chen, 2020). In the next, recall the inertial primal-dual dynamical system (51) as follows:

$$\begin{cases}
\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) = -\beta(t) (\nabla f(x(t)) + A^T \lambda(t)) + \epsilon(t), \\
\dot{\lambda}(t) = \beta(t) (A(x(t) + \delta \dot{x}(t)) - b).
\end{cases}
$$

From (6) and Proposition 1, the dynamic (51) serves as a guide for the introduction of the proposed fast primal-dual algorithms. For a better understanding of the acceleration schemes of the algorithms proposed in the previous sections, we will investigate the convergence properties of (51). The existence of the solution to (51) in some proper sense is beyond the scope of this paper. Without otherwise specified, in what follows, we always assume that $f$ is a convex and differentiable function.

**Theorem 4** Assume that $\frac{1}{\chi} \leq \alpha - 1$, $\beta : [t_0, +\infty) \rightarrow (0, +\infty)$ is a continuous differentiable function satisfying

$$t \beta(t) \leq \left( \frac{1}{\delta} - 2 \right) \beta(t), \quad \lim_{t \rightarrow +\infty} t^2 \beta(t) = +\infty,$$

and $\epsilon : [t_0, +\infty) \rightarrow \mathbb{R}^n$ is an integrable function satisfying

$$\int_{t_0}^{+\infty} t \| \epsilon(t) \| dt < +\infty.$$

Let $(x(t), \lambda(t))$ is a global solution of the dynamic (51) and $(x^*, \lambda^*) \in \Omega$. Then the dual trajectory $\lambda(t)$ is bounded and the following conclusions hold:

(ii) $L(x(t), \lambda^*) - L(x^*, \lambda^*) = \mathcal{O}(t^{-\chi})$,

(iii) When $\frac{1}{\chi} < \alpha - 1$, the primal trajectory $x(t)$ is bounded and

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right), \quad \int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty.$$

**Proof.** Define the energy function $E^t : [t_0, +\infty) \rightarrow \mathbb{R}$ as

$$E^t(t) = E(t) - \int_{t_0}^{t} \left( \frac{1}{\delta} (x(s) - x^*) + s \dot{x}(s), s \epsilon(s) \right) ds,$$

where

$$E(t) = E_0(t) + E_1(t)$$

with

$$\begin{cases}
E_0(t) = \frac{t^2}{2} \beta(t) (L(x(t), \lambda^*) - L(x^*, \lambda^*)), \\
E_1(t) = \frac{1}{2} \| x(t) - x^* \|^2 + \frac{\alpha}{2} \| \lambda(t) - \lambda^* \|^2.
\end{cases}
$$

By classic differential calculations, we have

$$\dot{E}_0(t) = t^2 \beta(t) \langle \nabla f(x(t)) + A^T \lambda^*, \dot{x}(t) \rangle + \left( 2 t \beta(t) + t^2 \beta(t) \right) (L(x(t), \lambda^*) - L(x^*, \lambda^*))$$

and

$$\begin{align*}
\dot{E}_1(t) &= \frac{1}{\delta} (x(t) - x^*) \cdot \dot{x}(t), (\frac{1}{\delta} + 1) \dot{x}(t) + t \dot{x}(t) \\
&= \frac{\alpha \delta - \delta - 1}{\delta^2} \| x(t) - x^* \|^2 + \frac{\alpha \delta - \delta - 1}{\delta^2} \| \lambda(t) - \lambda^* \|^2 \\
&= \frac{1}{\delta} (x(t) - x^*) \cdot \dot{x}(t), (\frac{1}{\delta} + 1 - \delta) \dot{x}(t) + t \dot{x}(t) \\
&= \frac{\alpha \delta - \delta - 1}{\delta^2} \| x(t) - x^* \|^2 + \frac{\alpha \delta - \delta - 1}{\delta^2} \| \lambda(t) - \lambda^* \|^2 \\
&= \frac{1}{\delta} (x(t) - x^*) \cdot \dot{x}(t), (\frac{1}{\delta} + 1 - \delta) \dot{x}(t) + t \dot{x}(t), t \dot{x}(t) \\
&= \frac{\alpha \delta - \delta - 1}{\delta^2} \| x(t) - x^* \|^2 + \frac{\alpha \delta - \delta - 1}{\delta^2} \| \lambda(t) - \lambda^* \|^2 \\
&= \frac{1}{\delta} (x(t) - x^*) \cdot \dot{x}(t), (\frac{1}{\delta} + 1 - \delta) \dot{x}(t) + t \dot{x}(t), t \dot{x}(t).
\end{align*}
$$

From assumptions, we get

$$\begin{align*}
\dot{E}^t(t) &= \dot{E}(t) - \frac{1}{\delta} (x(t) - x^*) \cdot \dot{x}(t), t \dot{x}(t) \\
&= t (t \beta(t) - (\frac{1}{\delta} - 2) \beta(t)) (L(x(t), \lambda^*) - L(x^*, \lambda^*)) \\
&= \frac{t \beta(t)}{\delta} (f(x(t)) - f(x^*) \cdot (x(t) - x^*), \nabla f(x(t))) \\
&= \frac{1}{\delta} \left( 1 + \alpha \right) t \| \dot{x}(t) \|^2
\end{align*}
$$

(54)
where the inequality follows from the convexity of $f$.

Since $(x^*, \lambda^*) \in \Omega$, it is easy to verify that $\mathcal{L}(x(t), \lambda^*) - \mathcal{L}(x^*, \lambda^*) \geq 0$ and $E(t) \geq 0$. By assumptions and (54), we get that $E'(t) \leq 0$. As a result, $E'(\cdot)$ is nonincreasing on $[t_0, +\infty)$ and

$$E'(t) \leq E'(t_0), \quad \forall t \in [t_0, +\infty). \quad (55)$$

By the definition of $E(\cdot)$ and $E^{\lambda, \epsilon}(\cdot)$, using Cauchy-Schwarz inequality, we get

$$\frac{1}{2} \left\| \frac{1}{\delta}(x(t) - x^*) + t \dot{x}(t) \right\|^2$$

$$\leq \left| E'(t_0) + \int^{t}_{t_0} \left\| \frac{1}{\delta}(x(s) - x^*) + s \dot{x}(s) \right\| \cdot s \epsilon(s) \| ds, \right.$$ for any $t \in [t_0, +\infty)$. Apply Lemma 6 with $\mu(t) = \left\| \frac{1}{\delta}(x(t) - x^*) + t \dot{x}(t) \right\|$, we get

$$\frac{1}{2} \left\| \frac{1}{\delta}(x(t) - x^*) + t \dot{x}(t) \right\| \leq \sqrt{2E'(t_0)} + \int^{+\infty}_{t_0} s \epsilon(s) \| ds < +\infty. \right.$$

for any $t \in [t_0, +\infty)$. This together with (53) and (55) implies

$$\inf_{t \in [t_0, +\infty)} E'(t)$$

$$\geq - \sup_{t \in [t_0, +\infty)} \left| \frac{1}{\delta}(x(t) - x^*) + t \dot{x}(t) \right| \times \int^{+\infty}_{t_0} s \epsilon(s) \| ds \right.$$\n
$$\geq - \infty \right.$$

and

$$\sup_{t \in [t_0, +\infty)} E(t) \leq E^{\lambda, \epsilon}(t_0)$$

$$+ \sup_{t \in [t_0, +\infty)} \left| \frac{1}{\delta}(x(t) - x^*) + t \dot{x}(t) \right| \times \int^{+\infty}_{t_0} s \epsilon(s) \| ds \right.$$\n
$$< +\infty. \right.$$

So $E'(t)$ and $E(t)$ are bounded on $[t_0, \infty)$. By the boundedness of $E'(t)$, we obtain

$$\mathcal{L}(x(t), \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O \left( \frac{1}{t^2 \beta(t)} \right), \quad (56)$$

$\lambda(t)$ is bounded on $[t_0, +\infty)$ and

$$\sup_{t \in [t_0, +\infty)} \left| \frac{1}{\delta}(x(t) - x^*) + t \dot{x}(t) \right| \| < +\infty, \quad (57)$$

When $\alpha \delta - \delta - 1 > 0$, the boundedness of $E'(t)$ also yield that $x(t)$ is bounded on $[t_0, +\infty)$, this together with (57) implies

$$\sup_{t \in [t_0, +\infty)} t \| \dot{x}(t) \| \| < +\infty. \quad (58)$$

By the boundedness of $E^{\lambda, \epsilon}(\cdot)$, integrating (54), we obtain the rest of the results.

**Remark 7** Theorem 4 generalizes (Attouch, Chbani, & Riahi, 2018, Theorem A.1) from the unconstrained optimization problem (50) to the constrained optimization problem (1), and some similar inertial dynamical system methods for the problem (1) can be referred to (Bot & Nguyen, 2021; He, Hu, & Fang, 2021a; Zeng, Lei, & Chen, 2019).

Consider the case $t \beta(t) = \eta t \beta(t)$ with $\eta \leq \frac{1}{3} - 2$. Then the assumption (52) holds, so the all results in Theorem 4 are true. In this case,

$$\frac{\dot{\beta}(t)}{\beta(t)} = \frac{\eta}{t}. \right.$$

Integrating the both sides on $[t_0, t]$, we get

$$\ln(\beta(t) - \ln \beta(t_0) = \eta(\ln t - \ln t_0), \right.$$\n
which implies $\beta(t) = \frac{\beta(t_0)}{\ln t_0 - \ln t}$. Now we derive an optimal convergence rate when $\beta(t) = \mu t^\eta$ with $\mu > 0$.

**Theorem 5** Let $(x(t), \lambda(t))$ be a solution of the dynamic (51), $(x^*, \lambda^*) \in \Omega$ and $\beta(t) = \mu t^\eta$. Assume that $\mu > 0, 0 \leq \eta \leq \frac{1}{3} - 2 \leq \alpha - \delta$ and $\int^{+\infty}_{t_0} t |\epsilon(t)| dt < +\infty. Then

$$f(x(t)) - f(x^*) = O \left( \frac{1}{t^{\eta+2}} \right), \| Ax(t) - b \| = O \left( \frac{1}{t^{\eta+2}} \right). \right.$$

**Proof.** By integrating the second equation of dynamic (51), we have

$$\lambda(t) - \lambda(t_0) = \int^{t}_{t_0} \dot{\lambda}(s) \| ds$$

$$= \int^{t}_{t_0} \lambda(t) \| ds + \delta \dot{s}(s) \| ds$$

$$= \int^{t}_{t_0} \lambda(t) \| ds + \int^{t}_{t_0} \delta s^2 \beta(s) d(Ax(s) - b)$$

$$= \delta s^2 \beta(s)(Ax(s) - b) \bigg|_{s=t_0}^{s=t} \right.$$\n
$$+ \delta \int^{t}_{t_0} s \left( \frac{1}{\delta} - 2 \right) \beta(s) - \delta \beta(s) \right\| (Ax(s) - b) ds$$

$$= \mu \delta t^{2+\eta}(Ax(t) - b) - \mu \delta t^{2+\eta}(Ax(t_0) - b)$$

$$+ \mu \delta \int^{t}_{t_0} s \left( \frac{1}{\delta} - 2 \right) \eta s^{\eta+1} (Ax(s) - b) ds.$$

From Theorem 4, we know that the dual trajectory $\lambda(t)$ is bounded, this together with (59) implies

$$\left\| t^{\eta+2} (Ax(t) - b) + \left( \frac{1}{\delta} - 2 - \eta \right) \int^{t}_{t_0} s^{\eta+1} (Ax(s) - b) ds \right\| \right.$$

$$\leq C, \quad \right.$$\n
where

$$C = \frac{1}{\mu} \sup_{t \geq t_0} \left\| \lambda(t) - \lambda(t_0) \right\| + \left\| \int^{t}_{t_0} s^{\eta+2} (Ax(t) - b) ds \right\| < +\infty.
Now, applying Lemma 7 with \( g(t) = \int_{t_0}^t s^{q+1}(Ax(s) - b)ds \), \( a = \frac{1}{2} - 2 - \eta \), \( b = C \), we obtain
\[
\sup_{t \geq t_0} \| t^{q+2}(Ax(t) - b) \| < +\infty,
\]
which is
\[
\| Ax(t) - b \| = O \left( \frac{1}{t^{q+2}} \right).
\]
This together with Theorem 4 implies
\[
| f(x(t)) - f(x^*) | \
\leq L(x(t), \lambda^*) - L(x^*, \lambda^*) + \| \lambda^* \| \| Ax(t) - b \|
= O \left( \frac{1}{t^{q+2}} \right).
\]

Remark 8 When \( \eta = 0 \) and \( \delta \in \left[ \frac{2}{3}, \frac{1}{2} \right] \), by Lemma 7, we prove the \( O \left( \frac{1}{t^s} \right) \) convergence rates on the objective residual and the feasibility violation. Using Lemma 7, we also can simplify the proof process of (Bot & Nguyen, 2021, Theorem 3.4), which investigated the improved convergence rates of the dynamic in (Zeng, Lei, & Chen, 2019).

Remark 9 In Theorem 4 and Theorem 5, we have established fast convergence properties of the dynamic (51), which are analogous to the ones of the algorithms proposed in previous sections. Doing so, we obtain a dynamic interpretation of the results on the convergence properties of the previous algorithms.

5 Numerical experiment

In this section, we test the proposed primal dual algorithms on solving the linearly constrained \( \ell_1 - \ell_2 \) minimization problem and the nonnegative linearly constrained quadratic programming problem. The numerical results demonstrate the validity and superior performance of our algorithms over some existing accelerated algorithms.

5.1 Linearly constrained \( \ell_1 - \ell_2 \) minimization problem

Consider the \( \ell_1 - \ell_2 \) minimization problem:
\[
\min_{x} \| x \|_1 + \frac{r}{2} \| x \|_2^2 \quad \text{s.t.} \quad Ax = b,
\]
where \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \). Set \( m = 1500 \), \( n = 3000 \). Generate \( A \) by the standard Gaussian distribution and the original solution (signal) \( x^* \in \mathbb{R}^n \) by the Gaussian distribution \( \mathcal{N}(0, 4) \) with 90% nonzero elements. The noise \( \omega \) is generated by the standard Gaussian distribution and normalized to the norm \( \| \omega \| = 10^{-5} \),
\[
b = Ax^* + \omega.
\]
In numerical examples, we solve subproblems by fast iterative shrinkage-thresholding algorithm (FISTA) (Beck & Teboulle, 2009) with the stopping condition:
\[
\frac{\| x_k - x_{k-1} \|_2^2}{\max \{ \| x_{k-1} \|, 1 \} } \leq tol
\]
or the number of iterations exceeds 100, where \( tol \) is precision. Denote the relative error \( Rel = \| x - x^* \|_2 \), the residual error \( Res = \| Ax - b \| \), and the signal-to-noise ratio
\[
SNR = \log_{10} \frac{\| x^* - \text{mean}(x^*) \|_2^2}{\| x - x^* \|_2^2},
\]
where \( x \) is the recovery signal.

Table 1 Numerical experiment with \( tol = 1e^-8 \)

| 7 | 0.1 | 0.5 | 1 | 1.2 |
|---|-----|-----|---|----|
| Iter | IAL | 13 | 15 | 18 |
| IAALM | 44 | 39 | 42 | 46 |
| FIPD-A | 5 | 7 | 10 | 11 |
| FIPD-B | 5 | 6 | 8 | 8 |
| Time | IAL | 18.50 | 20.66 | 23.98 | 22.97 |
| IAALM | 40.21 | 39.96 | 39.71 | 41.30 |
| FIPD-A | 10.98 | 13.36 | 15.94 | 17.04 |
| FIPD-B | 10.21 | 13.16 | 14.94 | 14.42 |
| Res | IAL | 2.69e-5 | 2.95e-5 | 1.00e-5 | 3.91e-5 |
| IAALM | 8.54e-5 | 8.47e-5 | 7.76e-5 | 7.79e-5 |
| FIPD-A | 5.46e-5 | 6.18e-5 | 2.84e-5 | 2.91e-5 |
| FIPD-B | 4.58e-5 | 6.36e-5 | 2.86e-5 | 2.95e-5 |
| SNR | IAL | 5.36e-8 | 5.52e-8 | 2.90e-8 | 7.46e-8 |
| IAALM | 1.48e-7 | 1.46e-7 | 1.61e-7 | 2.18e-7 |
| FIPD-A | 1.68e-7 | 1.85e-7 | 6.65e-8 | 1.92e-8 |
| FIPD-B | 9.41e-8 | 3.30e-7 | 3.79e-7 | 9.94e-8 |

Table 2 Numerical experiment with \( tol = 1e^-6 \)

| 7 | 0.1 | 0.5 | 1 | 1.2 |
|---|-----|-----|---|----|
| Iter | IAL | 38 | 58 | 63 | 62 |
| FIPD-A | 33 | 27 | 32 | 39 |
| FIPD-B | 23 | 35 | 20 | 21 |
| Time | IAL | 17.91 | 16.20 | 16.74 | 24.92 |
| FIPD-A | 9.51 | 10.00 | 12.20 | 16.24 |
| FIPD-B | 8.95 | 10.33 | 11.17 | 13.45 |
| Res | IAL | 9.23e-5 | 7.72e-5 | 9.84e-5 | 9.71e-5 |
| FIPD-A | 9.19e-5 | 9.22e-5 | 9.02e-5 | 8.98e-5 |
| FIPD-B | 8.35e-5 | 9.37e-5 | 9.46e-5 | 8.24e-5 |
| SNR | IAL | 1.70e-7 | 2.56e-7 | 1.98e-7 | 1.97e-7 |
| FIPD-A | 1.72e-7 | 2.75e-7 | 1.70e-8 | 1.55e-7 |
| FIPD-B | 1.64e-7 | 1.87e-7 | 5.08e-7 | 1.65e-8 |
| FIPD-C | 1.35e-7 | 1.34e-2 | 1.34e-2 | 1.34e-2 |

We compare Algorithm 1 (FIPD) with inexact augmented Lagrangian method (IAL (Liu, Liu, & Ma, 2019, Algorithm 1)) and inexact accelerated ALM (IAALM (Kang, Kang, & Jung, 2015, Algorithm 1)). Set the parameters as follows: IAL: \( \beta = 1 \); IAALM: \( \gamma = 1 \); FIPD-A: \( \theta = 2 \), \( \alpha = m \), \( \delta = \frac{1}{m} \); FIPD-B: \( \theta = 3 \), \( \alpha = m \), \( \delta = \frac{1}{n} \); M = \( \frac{1}{2} Id \). In FIPD-A and FIPD-B: take \( \delta_0 = 0.05 \) and use (34) to the scaling \( \beta_k \). We terminate all algorithms when \( \text{res} \leq 1e^-4 \).

In Tables 1, set \( tol = 1e^-8 \). We present the numerical results for various \( \tau \). When \( tol = 1e^-6 \), the IAL method does not work well, Tables 2 shows the numerical results of IAL, FIPD-A and FIPD-B.
5.2 Nonnegative linearly constrained quadratic programming problem

Let $Q \in \mathbb{S}_+(n)$, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Consider the nonnegative linearly constrained quadratic programming problem:

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

Set $m = 500$ and $n = 1000$. Let $Q = 2H^T H$ with $H \in \mathbb{R}^{n \times n}$ generated by the standard Gaussian distribution, $q \in \mathbb{R}^n$ be generated by the standard Gaussian distribution, $A = [B, Id]$ with $B \in \mathbb{R}^{m \times (n-m)}$ generated by the standard Gaussian distribution. In this case, let $f_1(x) = I_{\{y|y \geq 0\}}(x)$ be the indicator function of the set $\{y|y \geq 0\}$ and $f_2(x) = \frac{1}{2} x^T Q x + q^T x$. We compare Algorithm 2 (ILPD) with the accelerated linearized augmented Lagrangian method (AALM (Xu, 2017, Algorithm 1)). Set the parameters as follows: ILPD: $\alpha = 20$, $\beta = \|Q\|$, $\delta = \frac{1}{10}$, $M = \beta \|Q\|$; AALM: $\alpha_k = \frac{1}{k^{\frac{1}{3}}}$, $\beta_k = \gamma_k = \|Q\| k$, $P_k = \frac{2\|Q\|}{k} Id$. Subproblems are solved by interior-point algorithms with a tolerance $tol$. Fig. 1 shows the objective residual and the feasibility violation for the first 500 iterations with different tolerance.

![Numerical results of ILPD and AALM under different tolerance of subproblems](image)

6 Conclusion

By time discretization of the primal-dual dynamical system, we propose a fast primal-dual algorithm for the linear equality constrained optimization problem, and prove that the algorithms enjoy the fast convergence rate $\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O\left(\frac{1}{k^{1/3}}\right)$ under the scaling assumption (23). When (25) holds, we further obtain $|f(x_k) - f(x^*)| = O\left(\frac{1}{k^{1/3}}\right)$ and $\|Ax_k - b\| = O\left(\frac{1}{k^{1/3}}\right)$. We also propose a linearized primal dual algorithm with constant scaling coefficient for the composite problem (2), and prove the $O(1/k)$ convergence rates. Further, we investigate the convergence properties of the dynamic (5) for a better understanding of the proposed algorithms. We show that the dynamic owns a fast convergence properties analogous to that of the previous algorithms. The numerical experiments demonstrate the validity of acceleration and superior performance of the proposed algorithms over some existing ones.

A Some auxiliary results

The following lemmas have been used in the analysis of the convergence properties of the numerical algorithms and dynamical systems.

**Lemma 4** (Attouch, Chbani, & Peypouquet, 2018, Lemma 5.14) Let $\{a_k\}_{k \geq 1}$ and $\{b_k\}_{k \geq 1}$ be two nonnegative sequences. Assume $\sum_{k=1}^{+\infty} b_k < +\infty$ and

$$a_k^2 \leq c^2 + \sum_{j=1}^{k} b_j a_j \quad \forall k \in \mathbb{N},$$

where $c \geq 0$. Then,

$$a_k \leq c + \sum_{j=1}^{+\infty} b_j \quad \forall k \in \mathbb{N}.$$

**Lemma 5** (Lin, Li, & Fang, 2020, Lemma 3.18) Let $\{a_k\}_{k \geq 1}$ be a sequence of vectors in $\mathbb{R}^n$. Assume $\zeta > 1$ and

$$\|\zeta + (\zeta - 1)k\| a_{k+1} \leq \sum_{i=1}^{k} a_i \| \leq C \quad \forall k \geq 1.$$ 

Then $\sum_{i=1}^{k} a_i \| < C$ for all $k \geq 1$.

**Lemma 6** (Brezis, 1973, Lemma A.5) Let $\nu : [t_0, T] \rightarrow [0, +\infty)$ be integrable and $M \geq 0$. Suppose $\mu : [t_0, T] \rightarrow \mathbb{R}$ is continuous and

$$\frac{1}{2} \mu(t)^2 \leq \frac{1}{2} M^2 + \int_{t_0}^{t} \nu(s) \mu(s) ds$$

for all $t \in [t_0, T]$. Then $|\mu(t)| \leq M + \int_{t_0}^{t} \nu(s) ds$ for all $t \in [t_0, T]$.

**Lemma 7** Let $g : [t_0, +\infty) \rightarrow \mathbb{R}^n$ be a continuous differentiable function and $t_0 > 0$, $a \geq 0$, $b \geq 0$. Assume

$$\|Ag(t) + t\dot{g}(t)\| \leq b, \quad \forall t \geq t_0.$$ 

Then

$$\sup_{t \geq t_0} \|t \dot{g}(t)\| < +\infty.$$
Proof. When $a = 0$, the result can be directly obtained from assumption. Otherwise, from assumption, we have

$$
\left\| \frac{dt^a g(t)}{dt} \right\| \leq b t^{a-1},
$$

integrating it from $[t_0, t]$, we obtain

$$
\left\| \int_{t_0}^{t} \frac{ds}{ds} ag(s) \right\| \leq \int_{t_0}^{t} \left\| \frac{ds}{ds} ag(s) \right\| ds \leq \int_{t_0}^{t} bs^{a-1} ds.
$$

It yields

$$
\left\| t^a g(t) - t^a_0 g(t_0) \right\| \leq \frac{b}{a} (t^a - t^a_0),
$$

then we have

$$
\|g(t)\| \leq \frac{b}{a} + \frac{a\|t^a_0 g(t_0)\|}{at^a} - \frac{bt^a_0}{t^a} \forall t \geq t_0.
$$

This together with assumption implies

$$
\|t^a g(t)\| \leq b + \|ag(t)\| < 2b + \frac{a\|t^a_0 g(t_0)\|}{t^a} - \frac{bt^a_0}{t^a},
$$

and then

$$
\sup_{t \geq t_0} \|t^a g(t)\| \leq b + \sup_{t \geq t_0} \frac{a\|t^a_0 g(t_0)\|}{t^a} - \frac{bt^a_0}{t^a} < +\infty.
$$

References
Attouch, H., Chbani, Z., Peypouquet, J., & Redont, P. (2018). Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Mathematical Programming, 168(1), 123-175.

Attouch, H., Chbani, Z., & Riahi, H. (2019). Fast proximal methods via time scaling of damped inertial dynamics. SIAM Journal on Optimization, 29(3), 2227-2256.

Beck, A., & Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences, 2(1), 183-202.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3(1), 1-122.

Boţ, R. I., & Nguyen, D. K. (2021) Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping. Journal of Differential Equations, 303, 369-406.

Brezis, H. (1973). Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. Elsevier, New York.

Candès, E.J., & Wakin, M.B. (2008). An introduction to compressive sampling. IEEE Signal Processing Magazine, 25(2), 21-30.

Fazlyab, M., Koppel, A., Preciado, V. M., & Ribeiro, A. (2017). A variational approach to dual methods for constrained convex optimization. In 2017 American Control Conference (ACC) (pp. 5209-5275). IEEE.

Feijer, D., & Paganini, F. (2010). Stability of primal–dual gradient dynamics and applications to network optimization. Automatica, 46(12), 1974-1981.

He, X., Hu, R., & Fang, Y. P. (2021). Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems. SIAM Journal on Control and Optimization, 59 (5), 3278-3301.

He, X., Hu, R., & Fang, Y. P. (2021). Fast convergence of primal-dual dynamics and algorithms with time scaling for linear equality constrained convex optimization problems. arXiv:2103.12931.

He, X., Hu, R., & Fang, Y. P. (2021). Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems. arXiv:2103.12937.

He, B., & Yuan, X. (2010). On the acceleration of augmented Lagrangian method for linear equality constrained optimization. Optimization online, 3. http://www.optimization-online.org/DBFILE/2010/10/2760.pdf.

Huang, B., Ma, S., & Goldfarb, D. (2013). Accelerated linearized Bregman method. Journal of Scientific Computing, 54(2), 428-453.

Jordan, M. I. (2018). Dynamical, symplectic and stochastic perspectives on gradient-based optimization. In Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018 (pp. 523-549).

Kang, M., Kang, M., & Jung, M. (2015). Inexact accelerated augmented Lagrangian methods. Computational Optimization and Applications, 62(2), 373-404.

Kang, M., Yun, S., Woo, H., & Kang, M. (2013). Accelerated bregman method for linearly constrained $\ell_1$-$\ell_2$ minimization. Journal of Scientific Computing, 56(3), 515-534.

Kim, S. S., Cortés, J., & Martínez, S. (2015). Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. Automatica, 55, 254-264.

Li, X., & Yin, G. (2019). Exponential convergence of distributed primal–dual convex optimization algorithm without strong convexity. Automatica, 105, 298-306.

Lin, Z., Li, H., & Fang, C. (2020). Accelerated optimization for machine learning. Springer Singapore.

Liu, Y. F., Liu, X., & Ma, S. (2019). On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. Mathematics of Operations Research, 44(2), 632-650.

Luo, H. (2021). A primal-dual flow for affine constrained convex optimization. arXiv:2103.06636.

Luo, H. (2021). Accelerated primal-dual methods for linearly constrained convex optimization problems. arXiv:2109.12604.

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In Sov. Math. Dokl, 27(2), 372-376.

Nesterov, Y. (2013). Lectures on Convex Optimization: Second Edition. Springer Optimization and Its Applications 137, Springer, New York.

Su, W., Boyd, S., & Candés, E. J. (2016). A differential equation for modeling nesterov’s accelerated gradient method: Theory and insights. The Journal of Machine Learning Research, 17(153), 5312-5354.
Tang, Y., Qu, G., & Li, N. (2020). Semi–global exponential stability of augmented primal–dual gradient dynamics for constrained convex optimization. *Systems & Control Letters*, 144, 104754.

Wang, Z., Wei, W., Zhao, C., Ma, Z., Zheng, Z., Zhang, Y., & Liu, F. (2021). Exponential stability of partial primal–dual gradient dynamics with nonsmooth objective functions. *Automatica*, 129, 109585.

Wibisono, A., Wilson, A. C., & Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47), E7351-E7358.

Wilson, A. C., Recht, B., & Jordan, M. I. (2021). A Lyapunov Analysis of Accelerated Methods in Optimization. *Journal of Machine Learning Research*, 22(113), 1-34.

Xu, Y. (2017). Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. *SIAM Journal on Optimization*, 27(3), 1459-1484.

Zeng, X., Lei, J., & Chen, J. (2019). Dynamical primal-dual accelerated method with applications to network optimization. arXiv:1912.03681.

Zhang, X., Burger, M., Bresson, X., & Osher, S. (2010). Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3), 253-276.

Zhu, Y., Yu, W., Wen, G., & Chen, G. (2020). Projected primal–dual dynamics for distributed constrained nonsmooth convex optimization. *IEEE Transactions on Cybernetics*, 50(4), 1776-1782.