Fast convergence of primal-dual dynamics and algorithms with time scaling for linear equality constrained convex optimization problems

Xin He · Rong Hu · Ya-Ping Fang

Received: date / Accepted: date

Abstract We propose a primal-dual dynamic with time scaling for a linear equality constrained convex optimization problem, which consists of a second-order ODE for the primal variable and a first-order ODE for the dual variable. Without assuming strong convexity, we prove its fast convergence property and show that the obtained fast convergence property is preserved under a small perturbation. We also develop an inexact primal-dual algorithm derived by a time discretization, and derive the fast convergence property matching that of the underlying dynamic. Finally, we give numerical experiments to illustrate the validity of the proposed algorithm.

Keywords Linear equality constrained convex optimization problem · Primal-dual dynamic and algorithm · Time scaling · Energy function and energy sequence · Convergence rate.

Mathematics Subject Classification (2010) 34D05 · 37N40 · 46N10 · 49M29 · 65K05 · 65K10 · 90C25

1 Introduction

A basic problem arising in many applications such as compressed sensing, statistical estimation, machine learning, global consensus problem, and image processing, is the linear equality constrained
convex optimization problem:
\[
\min_x f(x), \quad s.t. \ Ax = b,
\]
where \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \), and \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) be a proper, closed, and convex function. See e.g. [14,16,17] for some examples. A benchmark and effective method for the problem (1) is the augmented Lagrangian method (ALM) originally suggested by Hestenes [26] and Powell [35]. ALM and its variants have attracted much attentions for their evident advantages and many researchers have made contributions. Here, we only mentions a few of nice works on the convergence rate analysis.

He and Yuan [25] established a nonergodic \( O(1/k) \) convergence rate of ALM in the case that \( f \) is differentiable. Gu et al. [22] showed that a relaxed ALM enjoys an ergodic \( O(1/k) \) convergence rate in the case that \( f \) is nondifferentiable. Xu [42] derived an ergodic \( O(1/k) \) convergence rate of a linearized ALM with fixed parameters. The nonergodic convergence rates of inexact versions of ALM have been discussed in Lan and Monteiro [29] and Liu et al. [31]. By applying the Nesterov’s acceleration technique to ALM, He and Yuan [25] proposed an accelerated ALM for the problem (1) with \( f \) being differentiable, and showed that it enjoys a nonergodic \( O(1/k^2) \) convergence rate.

On the other hand, dynamical systems have been shown to be efficient tools for solving optimization problems. The dynamical system approaches can not only give more insights into existing numerical methods, but also lead to other possible numerical algorithms by discretization. The gradient flow system
\[
\dot{x}(t) = \nabla \Phi(x(t))
\]
can be understood as the continuous limit of the proximal point algorithm and the gradient method for the unconstrained convex optimization problem
\[
\min \Phi(x),
\]
where \( \Phi(x) \) is a differentiable convex function. More precisely, the implicit and explicit schemes of the gradient flow system lead to the proximal point algorithm and the gradient method respectively. It was shown by Alvarez [1] that an implicit discretization scheme of the heavy ball with friction system due to Polyak [34]
\[
(HBF) \quad \ddot{x}(t) + \gamma \dot{x}(t) + \nabla \Phi(x(t)) = 0,
\]
where \( \gamma > 0 \) is a constant damping, may lead to an inertial proximal-like algorithm for the problem (2). Su et al. [36] showed that the inertial system
\[
(AVD) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla \Phi(x(t)) = 0,
\]
with $\alpha = 3$ can be understood as the continuous limit of the Nesterov’s accelerated gradient algorithm [33], and its widely used successors like the FISTA algorithm of Beck and Teboulle [10] for the problem (2). Su et al. [36] also derived the convergence rate $\Phi(x(t)) - \min \Phi = O(1/t^2)$ of $(AVD_\alpha)$ with $\alpha \geq 3$. Attouch et al. [11] generalized the result of [36] by considering an additional perturbation, and provided a new fast proximal algorithm with a nonergodic $O(1/k^2)$ convergence rate, by discretizing the perturbed $(AVD_\alpha)$. Without assuming strong convexity of $\Phi$, Balhag el al. [8] derived the exponential convergence of the inertial gradient system with time scaling $\beta(t)$

$$\ddot{x}(t) + \gamma \dot{x}(t) + \beta(t)\nabla \Phi(x(t)) = 0 \quad (3)$$

under the assumption $\beta(t) = e^{\beta t}$ with $\beta \leq \gamma$. By discretizing (3), Balhag el al. [8] presented an inertial proximal algorithm and gave similar results to the continuous case. Attouch et al. [5] proposed inertial proximal algorithms for the problem (2) with fast convergence properties, which are obtained by discretizing the following second-order dynamical system

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

The fast convergence properties of $(AVD_{\alpha,\beta})$ as well as its associated discrete algorithms were also discussed by Wibisono et al. [40] and Wilson et al. [41] from a variational perspective. For more results on the dynamical system approaches for unconstrained optimization problems, we refer the reader to [2,3,7,11,12].

As mentioned above, in the literature there are many papers on dynamical system approaches for unconstrained optimization problems. However, it is not rich for the literature on dynamical system approaches designed to solve constrained optimization problems. Moreover, most of them focus on first-order dynamical system approaches for the problem (1) (see e.g.[13,18,21,44]). Second-order dynamical system approaches for the problem (1) are less discussed as yet, and we mention related works as follows. Zeng et al. [43] proposed a second-order dynamical system based on primal-dual framework and proved $L(x(t), \lambda^*) - L(x^*, \lambda^*) = O(1/t^2)$, where $L(x, \lambda)$ is the Lagrangian function of the problem (1) and $(x^*, \lambda^*)$ is a saddle point of $L$ in the sense that

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \quad \forall (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m. \quad (4)$$

He et al. [24] further proposed a second-order primal-dual dynamical system involving time-dependent positive damping terms for the problem (1) with a separable structure, and established similar results to the ones in [43]. Note that neither Zeng et al. [43] nor He et al. [24] discussed how to obtain accelerated algorithms through discretization schemes of the proposed systems. Motivated by [40], Fazlyab et al. [20] developed an Euler-Lagrange equation for (1) and proved that it enjoys an exponential converge under the assumptions that $f$ is strongly convex and twice continuously differentiable. Fazlyab et al. [20] also discussed some accelerated primal-dual algorithms for (1), which are obtained by suitable discretization schemes of the Euler-Lagrange equation.

In this paper, we propose primal-dual dynamics and algorithms with time scaling for the problem (1). Under a scaling condition we prove the fast convergence properties of the proposed dynamics and algorithms by the Lyapunov analysis approach. Our main contributions are summarized as follows:
a) We propose a new primal-dual dynamic with time scaling for the problem (1). Compared with
the existing primal-dual dynamical systems, the novel part in the proposed dynamic lies in its
structure: a primal-dual dynamic with an inertial second-order ODE for the primal variable and a
first-order ODE for the dual variable. By constructing energy functions, we show that the primal-
dual dynamic enjoys an ergodic $O(1/t)$ convergence rate when the scaling coefficient $\beta(t) \equiv \beta > 0$, and a nonergodic $O(1/\beta(t))$ convergence rate when $\lim_{t \to +\infty} \beta(t) = +\infty$. The convergence
properties turn out to be stable under an external perturbation $\epsilon(t)$ with $\int_{0}^{+\infty} ||\epsilon(t)|| dt < +\infty$.
To our best knowledge, this is the first time to consider “second-order” + “first-order” primal-
dual dynamic with time scaling for the problem (1).

b) By a suitable discretization of the proposed dynamic, we present a new primal-dual algorithm
for the problem (1) where the subproblem may be solved inexactly, and prove that the presented
algorithm enjoys a fast convergence rate matching that of the underlying dynamic. We test the
proposed algorithm on a quadratic programming problem to illustrate its validity.

The rest of this paper is organized as follow: In Section 2, we proposed a primal-dual dynamic
with time scaling for the problem (1) as well as its perturbed version, and investigate their fast
convergence properties. In Section 3, we present a fast inexact primal-dual algorithm for the problem
(1) by discretizing the dynamic in Section 2, and derive similar convergence rates to the underlying
dynamic. In Section 4, we test the algorithm in Section 3 on quadratic programming problems.
Finally, we give a concluding remark in Section 5.

2 Primal-dual dynamic with time scaling

In this section we shall propose a primal-dual dynamic with time scaling as well as its perturbed
version for the problem (1) in terms of the augmented Lagrangian function, and discuss their fast
convergence properties. Recall that the augmented Lagrangian function $L^\sigma(x, \lambda)$ of the problem (1)
is defined by

$$L^\sigma(x, \lambda) = L(x, \lambda) + \frac{\sigma}{2} ||Ax - b||^2,$$

where

$$L(x, \lambda) = f(x) + \langle \lambda, Ax - b \rangle$$

is the Lagrangian function of the problem (1) and $\sigma > 0$ is the penalty parameter. As known, $x^*$ is
a solution of the problem (1) if and only if there exists $\lambda^* \in \mathbb{R}^m$ such that $(x^*, \lambda^*)$ is a saddle point
of $L$. It is also known that $(x^*, \lambda^*)$ is a saddle point of $L$ if and only if it is a KKT point of the
problem (1) in the sense that

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

(5)

where $\partial f$ is the classical subdifferential of $f$ defined by

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

Unless otherwise specified, we always assume that the saddle point set $\Omega$ of the Lagrangian $L$ is
nonempty throughout this paper, and $f$ is a proper, convex and continuously differentiable function
in this section. Given a fixed $t_0 \geq 0$, in terms of the augment Lagrangian function $L^\omega$, we propose the following primal-dual dynamic with time scaling:

$$\begin{align*}
\ddot{x}(t) + \gamma \dot{x}(t) &= -\beta(t)\nabla_x L^\omega(x(t), \lambda(t)), \\
\dot{\lambda}(t) &= \beta(t)\nabla_\lambda L^\omega(x(t) + \delta \dot{x}(t), \lambda(t)),
\end{align*}$$

where $\gamma, \delta > 0$ are two constant damping coefficients, and $\beta : [t_0, +\infty) \to (0, +\infty)$ is a time scaling coefficient which plays a crucial role in obtaining fast convergence property. By trivial computation, we can rewrite it as follows:

$$\begin{align*}
\ddot{x}(t) + \gamma \dot{x}(t) &= -\beta(t)(\nabla f(x(t)) + A^T \lambda(t) + \sigma A^T(Ax(t) - b)), \\
\dot{\lambda}(t) &= \beta(t)(A(x(t) + \delta \dot{x}(t)) - b).
\end{align*}$$

(6)

The existence and uniqueness of local solutions of (6) follows directly from the Picard-Lindelof Theorem (see [37] Theorem 2.2]).

**Proposition 1** Let $f$ be such that $\nabla f$ is locally Lipschitz continuous and let $\beta : [t_0, +\infty) \to (0, +\infty)$ be locally integrable. Then for any $(x_0, \lambda_0, u_0)$, there exists a unique solution $(x(t), \lambda(t))$ with $x(t) \in C^2([t_0, T], \mathbb{R}^n)$, $\lambda(t) \in C^1([t_0, T], \mathbb{R}^m)$ of the dynamic (6) satisfying $(x(t_0), \lambda(t_0), \dot{x}(t_0)) = (x_0, \lambda_0, u_0)$ on a maximal interval $[t_0, T) \subseteq [t_0, +\infty)$.

Now we start to discuss the asymptotic properties of (6).

**Theorem 1** Assume that $\nabla f$ is locally Lipschitz continuous, $\beta : [t_0, +\infty) \to (0, +\infty)$ is a continuous differentiable function, and the following scaling condition holds:

$$\dot{\beta}(t) \leq \frac{1}{\delta} \beta(t), \quad \frac{1}{\delta} < \gamma.$$ (7)

Then for any $(x_0, \lambda_0, u_0)$, there exists a unique global solution $(x(t), \lambda(t))$ with $x(t) \in C^2([t_0, +\infty), \mathbb{R}^n)$, $\lambda(t) \in C^1([t_0, +\infty), \mathbb{R}^m)$ of the dynamic (6) satisfying $(x(t_0), \lambda(t_0), \dot{x}(t_0)) = (x_0, \lambda_0, u_0)$. Moreover, $(x(t), \lambda(t))$ is bounded on $[t_0, +\infty)$, and the following conclusions hold for any $(x^*, \lambda^*) \in \Omega$:

1. $\int_{t_0}^{+\infty} \frac{1}{\delta} \beta(t) - \dot{\beta}(t)(L^\omega(x(t), \lambda^*) - L^\omega(x^*, \lambda^*))dt < +\infty$.
2. $\int_{t_0}^{+\infty} \beta(t)\|Ax(t) - b\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$.
3. If $\beta(t) \equiv \beta > 0$, then

$$L(\bar{x}(t), \lambda^*) - L(x^*, \lambda^*) = O\left(\frac{1}{t}\right), \quad \|\dot{\bar{x}}(t)\| = O\left(\frac{1}{\sqrt{t}}\right) \quad \text{with} \quad \bar{x}(t) = \int_{t_0}^{t} x(s)ds.$$  

If $\lim_{t \to +\infty} \beta(t) = +\infty$, then

$$L(x(t), \lambda^*) - L(x^*, \lambda^*) = O\left(\frac{1}{\beta(t)}\right), \quad \|Ax(t) - b\| = O\left(\frac{1}{\sqrt{\beta(t)}}\right).$$

**Proof** By Proposition 1 there exists a unique local solution $(x(t), \lambda(t))$ of (6) defined on a maximal interval $[t_0, T)$ with $T \leq +\infty$ for any initial value.

We first show $T = +\infty$. Fix $(x^*, \lambda^*) \in \Omega$ and let $\lambda \in \mathbb{R}^m$. Consider the energy function $E^\lambda : [t_0, T) \to [0, +\infty)$ defined by

$$E^\lambda(t) = E_0(t) + E_1(t),$$ (8)
\[
\begin{align*}
\mathcal{E}_0(t) &= \beta(t)(\mathcal{L}^\sigma(x(t), \lambda) - \mathcal{L}^\sigma(x^*, \lambda)), \\
\mathcal{E}_1(t) &= \frac{1}{2}\|x(t) - x^*\|^2 + \frac{\delta\gamma - 1}{2\delta}\|x(t) - x^*\|^2 + \frac{1}{\delta}\|\lambda(t) - \lambda\|^2.
\end{align*}
\]

Differentiate \( \mathcal{E}_0(t) \) to get
\[
\dot{\mathcal{E}}_0(t) = \dot{\beta}(t)(f(x(t)) - f(x^*) + \langle \lambda, Ax(t) - b \rangle) + \beta(t)(\nabla f(x(t)) + A^T\lambda + \sigma A^T(Ax(t) - b), \dot{x}(t)).
\]

Since \((x^*, \lambda^*) \in \Omega \) and \( Ax^* = b \), from (6) we have
\[
\begin{align*}
\dot{\mathcal{E}}_1(t) &= \left(\frac{1}{\delta}(x(t) - x^*) + \dot{x}(t), \frac{1}{\delta}\dot{x}(t) + \ddot{x}(t)\right) + \frac{\delta\gamma - 1}{2\delta}\langle x(t) - x^*, \dot{x}(t) \rangle + \frac{1}{\delta}\langle \lambda(t) - \lambda, \dot{\lambda}(t) \rangle \\
&= \left(\frac{1}{\delta}(x(t) - x^*) + \dot{x}(t), \frac{1}{\delta}(x(t) - x^*) - \beta(t)(\nabla f(x(t)) + A^T\lambda(t) + \sigma A^T(Ax(t) - b))\right) \\
&\quad + \frac{\delta\gamma - 1}{\delta}\langle x(t) - x^*, \dot{x}(t) \rangle + \frac{\beta(t)}{\delta}(\lambda(t) - \lambda, A(x(t) - x^*) + \delta A\dot{x}(t)) \\
&= \frac{1 - \delta\gamma}{\delta}\|\dot{x}(t)\|^2 - \frac{\beta(t)}{\delta}(x(t) - x^*, \nabla f(x(t)) + A^T\lambda + \sigma A^T(Ax(t) - b)) \\
&\quad - \beta(t)(\dot{x}(t), \nabla f(x(t)) + A^T\lambda + \sigma A^T(Ax(t) - b)).
\end{align*}
\]

It follows from (8)-(11) that
\[
\begin{align*}
\dot{\mathcal{E}}^\lambda(t) &= \left(\beta(t) - \frac{1}{\delta}\beta(t)))(\mathcal{L}^\sigma(x(t), \lambda) - \mathcal{L}^\sigma(x^*, \lambda)\right) - \frac{\beta(t)}{2\delta}\|Ax(t) - b\|^2 \\
&\quad + \frac{\beta(t)}{\delta}(f(x(t)) - f(x^*) - \langle \lambda, Ax(t) - b \rangle) + \frac{1 - \delta\gamma}{\delta}\|\dot{x}(t)\|^2 \\
&\leq \left(\beta(t) - \frac{1}{\delta}\beta(t))(\mathcal{L}^\sigma(x(t), \lambda) - \mathcal{L}^\sigma(x^*, \lambda)\right) - \frac{\beta(t)}{2\delta}\|Ax(t) - b\|^2 + \frac{1 - \delta\gamma}{\delta}\|\dot{x}(t)\|^2,
\end{align*}
\]

where the inequality is deduced from the convexity of \( f \). Take \( \lambda = \lambda^* \). By (11), \( \mathcal{L}^\sigma(x(t), \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*) \geq 0 \). This implies \( \mathcal{E}^\lambda(t) \geq 0 \). Combining (7) with (11), we get
\[
\begin{align*}
\dot{\mathcal{E}}^{\lambda^*}(t) &\leq \left(\beta(t) - \frac{1}{\delta}\beta(t))(\mathcal{L}^\sigma(x(t), \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)\right) \\
&\quad - \frac{\beta(t)}{2\delta}\|Ax(t) - b\|^2 + \frac{1 - \delta\gamma}{\delta}\|\dot{x}(t)\|^2 \\
&\leq 0
\end{align*}
\]
for any \( t \in [t_0, T) \). As a consequence, \( \mathcal{E}^{\lambda^*}(t) \) is nonincreasing on \([t_0, T), \) and so
\[
0 \leq \mathcal{E}^{\lambda^*}(t) \leq \mathcal{E}^{\lambda^*}(t_0), \quad \forall t \in [t_0, T).
\]

This together with (8) implies
\[
\frac{1}{2}\|x(t) - x^*\|^2 + \frac{\delta\gamma - 1}{2\delta}\|x(t) - x^*\|^2 \leq \mathcal{E}^{\lambda^*}(t_0), \quad \forall t \in [t_0, T).
\]

Since \( \delta\gamma - 1 > 0 \), we get
\[
\|x(t) - x^*\| \leq \sqrt{\frac{2\delta^2\mathcal{E}^{\lambda^*}(t_0)}{\delta\gamma - 1}}, \quad \forall t \in [t_0, T),
\]
and then
\[
\sup_{t \in [t_0, T)} \|\dot{x}(t)\| \leq \sqrt{2\mathcal{E}^{\lambda^*}(t_0)} + \sup_{t \in [t_0, T)} \frac{1}{\delta}\|x(t) - x^*\| < +\infty.
\]
Remark 1. If the scaling condition (7) is replaced with the following stronger one:

\[ \dot{\beta}(t) \leq (1 - \kappa)\frac{1}{\delta}\beta(t), \quad \frac{1}{\delta} < \gamma \]

for some \( \kappa \in (0, 1] \), then by Theorem 1(i) we have

\[ \int_{t_0}^{+\infty} \beta(t)(\mathcal{L}^\sigma(x(t), \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*))dt < +\infty. \]

It is worth mentioning that Balhag et al. [3] discussed the convergence rate analysis of the dynamic (3) by proving \( \Phi(x(t)) - \min \Phi = \mathcal{O}(\frac{1}{\beta(t)}) \) in [3] Theorem 2.1 under the condition \( \dot{\beta}(t) \leq \gamma\beta(t) \), and

\[ \int_{t_0}^{+\infty} \beta(t)(\Phi(x(t)) - \min \Phi)dt < +\infty \]

in [3] Proposition 2.3 under the condition \( \dot{\beta}(t) \leq (1 - \kappa)\gamma\beta(t) \) for the unstained optimization problem (2). In this sense, Theorem 1 extends [3] Theorem 2.1, Proposition 2.3 from the the unstained optimization problem (2) to the linear equality constrained problem (1).
It follows from the definition of $L$ that

$$|f(x(t)) - f(x^*)| \leq L(x(t), \lambda^*) - L(x^*, \lambda^*) + \|\lambda^*\||Ax(t) - b||.$$

Next, we shall derive an improved convergence rate when the scaling coefficient satisfies $\dot{\beta}(t) = \frac{1}{\delta}\beta(t)$. Under the condition $\dot{\beta}(t) = \frac{1}{\delta}\beta(t)$, it is easy to verify that

$$\beta(t) = \mu e^{t/\delta} \text{ with } \mu = e^{-t_0/\delta}.$$

**Theorem 2** Assume that $\beta(t) = \mu e^{t/\delta}$ with $\mu > 0$ and $\frac{1}{\delta} < \gamma$. Let $(x(t), \lambda(t))$ be a global solution of the dynamic (6) and $(x^*, \lambda^*) \in \Omega$. Then

(i) $|f(x(t)) - f(x^*)| = O\left(\frac{1}{\sqrt{\beta(t)}}\right)$, $\|Ax(t) - b\| = O\left(\frac{1}{\sqrt{\beta(t)}}\right)$.

(ii) $\int_{t_0}^{t+\epsilon} e^{t/\delta} ||Ax(t) - b||^2 dt < +\infty$, $\int_{t_0}^{+\infty} \|\hat{x}(t)\|^2 < +\infty$.

**Proof** By Theorem 1 we only need to show (i). Given $\lambda \in \mathbb{R}^m$, define the energy function $\mathcal{E}^\lambda$ as in (8) with $\beta(t) = \mu e^{t/\delta}$. It follows from (11) that

$$\dot{\mathcal{E}}^\lambda(t) \leq -\frac{\sigma \beta(t)}{2\delta} ||Ax(t) - b||^2 + \frac{1 - \delta\gamma}{\delta} \|\hat{x}(t)\|^2 \leq 0.$$

This implies that $\mathcal{E}^\lambda(t)$ is nonincreasing on $[t_0, +\infty)$ for any $\lambda \in \mathbb{R}^m$, and so

$$\mathcal{E}^\lambda(t) \leq \mathcal{E}^\lambda(t_0), \quad \forall t \in [t_0, +\infty), \lambda \in \mathbb{R}^m.$$

It follows from the definition of $\mathcal{L}^\sigma$ and (8) that

$$f(x(t)) - f(x^*) + \langle \lambda, Ax(t) - b \rangle \leq \frac{1}{\mu e^{t/\delta}} \mathcal{E}^\lambda(t_0), \quad \forall \lambda \in \mathbb{R}^m.$$

Take $\rho > \|\lambda^*\|$. By Lemma 4 we have

$$f(x(t)) - f(x^*) + \rho ||Ax(t) - b|| \leq \frac{1}{\mu e^{t/\delta}} \sup_{\|\lambda\| \leq \rho} \mathcal{E}^\lambda(t_0). \quad (16)$$

From (8), we get

$$f(x(t)) - f(x^*) \geq -\|\lambda^*\||Ax(t) - b|| \quad (17)$$

and $\sup_{\|\lambda\| \leq \rho} \mathcal{E}^\lambda(t_0) \geq \mathcal{E}^\lambda(t_0) \geq 0$. This together with (16) yields

$$\|Ax(t) - b|| \leq \frac{1}{\rho - \|\lambda^*\| e^{t/\delta}} \sup_{\|\lambda\| \leq \rho} \mathcal{E}^\lambda(t_0). \quad (18)$$

It follows from (16)–(18) that

$$-\|\lambda^*\| \sup_{\|\lambda\| \leq \rho} \mathcal{E}^\lambda(t_0) \leq f(x(t)) - f(x^*) \leq \frac{1}{\mu e^{t/\delta}} \sup_{\|\lambda\| \leq \rho} \mathcal{E}^\lambda(t_0). \quad (19)$$

It is easy to verify that $\sup_{\|\lambda\| \leq \rho} \mathcal{E}^\lambda(t_0) < +\infty$. By (18) and (19), we get (i).
Next, we consider the following perturbed version of the dynamic (3):
\[
\begin{align*}
\dot{x}(t) + \gamma \dot{x}(t) &= -\beta(t)(\nabla f(x(t)) + A^T \lambda(t)) + \sigma A^T (Ax(t) - b) + \epsilon(t), \\
\dot{\lambda}(t) &= \beta(t)(Ax(t) + \delta \dot{x}(t)) - b),
\end{align*}
\tag{20}
\]
where $\epsilon : [t_0, +\infty) \to \mathbb{R}^n$ is an external perturbation.

To avoid repeating the proof, we take for granted the existence and uniqueness of a global solution of (6). We shall show that the fast convergence properties established in Theorem 1 and Theorem 2 are preserved for (20) as $\epsilon(t)$ decays rapidly.

**Theorem 3** Assume that $\epsilon : [t_0, +\infty) \to \mathbb{R}^n$ is an integrable function such that
\[
\int_{t_0}^{+\infty} \|\epsilon(t)\| dt < +\infty
\]
and the scaling condition (7) holds. Let $(x(t), \lambda(t))$ be a global solution of the dynamic (20) and $(x^*, \lambda^*) \in \Omega$. Then the trajectory $(x(t), \lambda(t))$ is bounded on $[t_0, +\infty)$ and

(i) $\int_{t_0}^{+\infty} (\frac{1}{2} \beta(t) - \beta(t)) (\mathcal{L}^\sigma (x(t), \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*)) dt < +\infty$.

(ii) $\int_{t_0}^{+\infty} \beta(t) \|Ax(t) - b\|^2 dt < +\infty$, $\int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$.

(iii) If $\beta(t) \equiv \beta > 0$, then
\[
\mathcal{L}(\dot{x}(t), \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O(\frac{1}{t}), \quad \|Ax(t) - b\| = O(\frac{1}{\sqrt{t}}) \text{ with } \dot{x}(t) = \frac{\int_{t_0}^{t} x(s) ds}{t - t_0}.
\]

If $\lim_{t \to +\infty} \beta(t) = +\infty$, then
\[
\mathcal{L}(x(t), \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O(\frac{1}{\beta(t)}), \quad \|Ax(t) - b\| = O(\frac{1}{\beta(t)}).
\]

**Proof** Given $\lambda \in \mathbb{R}^m$, define the energy function $\mathcal{E}^{\lambda, \epsilon} : [t_0, +\infty) \to \mathbb{R}$ by
\[
\mathcal{E}^{\lambda, \epsilon}(t) = \mathcal{E}^\lambda(t) - \int_{t_0}^{t} \left(\frac{1}{2} \mathcal{L}(x(s), \lambda) - \mathcal{L}(x^*, \lambda^*)\right) ds,
\tag{21}
\]
where $\mathcal{E}^\lambda(t)$ is defined as in (8). By similar arguments as in the proof of Theorem 1, we have
\[
\dot{\mathcal{E}}^{\lambda, \epsilon}(t) \leq (\frac{1}{\delta} - \frac{1}{\beta(t)}) (\mathcal{L}^\sigma (x(t), \lambda) - \mathcal{L}^\sigma (x^*, \lambda)) - \frac{\sigma \beta(t)}{2\delta} \|Ax(t) - b\|^2 - \frac{\delta \gamma^2}{\delta^2} \|\dot{x}(t)\|^2 \tag{22}
\]
for any $t \geq t_0$. Take $\lambda = \lambda^*$. By (22) and (7), $\mathcal{E}^{\lambda^*, \epsilon}(t)$ is nonincreasing on $[t_0, +\infty)$ and
\[
\mathcal{E}^{\lambda^*, \epsilon}(t) \leq \mathcal{E}^{\lambda^*, \epsilon}(t_0), \quad \forall t \in [t_0, +\infty).
\]

This together with (21) implies
\[
\mathcal{E}^{\lambda^*}(t) \leq \mathcal{E}^{\lambda^*, \epsilon}(t_0) + \int_{t_0}^{t} \left(\frac{1}{\delta} \mathcal{L}(x(s), \lambda) - \mathcal{L}(x^*, \lambda)\right) + \dot{x}(s), \epsilon(s) ds, \quad \forall t \in [t_0, +\infty).
\tag{23}
\]

By the definition of $\mathcal{E}^{\lambda^*}(t)$ and the Cauchy-Schwarz inequality, from (23) we have
\[
\frac{1}{2} \left| \frac{1}{\delta} (x(t) - x^*) + \dot{x}(t) \right|^2 \leq |\mathcal{E}^{\lambda^*, \epsilon}(t_0)| + \int_{t_0}^{t} \frac{1}{\delta} (x(s) - x^*) + \dot{x}(s), \epsilon(s) ds, \quad \forall t \in [t_0, +\infty).
\]
Applying Lemma \ref{lem:2} with \( \mu(t) = \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| \) to the above inequality, we get
\[ \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| \leq 2|\mathcal{E}^{\lambda, \epsilon}(t_0)| + \int_{t_0}^{t} \|\epsilon(s)\|ds, \quad \forall t \in [t_0, +\infty). \]
This together with \( \int_{t_0}^{+\infty} \|\epsilon(s)\|ds < +\infty \) yields
\[ \sup_{t \geq t_0} \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| < +\infty. \]
Since \( \mathcal{E}^{\lambda, \epsilon}(t) \geq 0 \) for all \( t \geq t_0 \), it follows from \eqref{eq:21} and \eqref{eq:23} that
\[ \inf_{t \geq t_0} \mathcal{E}^{\lambda, \epsilon}(t) \geq - \sup_{t \geq t_0} \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| \times \int_{t_0}^{+\infty} \|\epsilon(s)\|ds > -\infty \]
and
\[ \sup_{t \geq t_0} \mathcal{E}^{\lambda, \epsilon}(t) \leq \mathcal{E}^{\lambda, \epsilon}(t_0) + \sup_{t \geq t_0} \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| \times \int_{t_0}^{+\infty} \|\epsilon(s)\|ds < +\infty. \]
This means that \( \mathcal{E}^{\lambda, \epsilon}(t) \) and \( \mathcal{E}^{\lambda}(t) \) are both bounded on \( [t_0, +\infty) \). Integrating \eqref{eq:22} on \( [t_0, +\infty) \), we obtain (i) and (ii). The rest of the proof is similar to the one in Theorem \ref{thm:1}, and so we omit it.

In the next, we discuss the convergence rate of the perturbed dynamic \eqref{eq:20} in the case \( \beta(t) = \mu e^{t/\delta} \).

\begin{theorem}
Assume that \( \beta(t) = \mu e^{t/\delta} \) with \( \mu > 0, \frac{1}{\delta} < \gamma \), and \( \epsilon : [t_0, +\infty) \to \mathbb{R}^n \) is an integrable function satisfying
\[ \int_{t_0}^{+\infty} \|\epsilon(t)\|dt < +\infty. \]
Let \((x(t), \lambda(t))\) be a global solution of \eqref{eq:20} and \((x^*, \lambda^*) \in \Omega \). Then
(i) \( |f(x(t)) - f(x^*)| = O\left(e^{t/\delta}\right), \quad \|Ax(t) - b\| = O\left(e^{t/\delta}\right) \)
(ii) \( \int_{t_0}^{+\infty} e^{t/\delta} \|Ax(t) - b\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 < +\infty. \)
\end{theorem}

\begin{proof}
Since \( \beta(t) = \mu e^{t/\delta} \), we have \( \dot{\beta}(t) = \frac{1}{\delta} \beta(t) \). It follows from \eqref{eq:22} that for any \( \lambda \in \mathbb{R}^m \)
\[ \mathcal{E}'^{\lambda, \epsilon}(t) \leq - \frac{\sigma(t)}{2\delta} \|Ax(t) - b\|^2 + \frac{\delta - \frac{1}{\delta}}{\delta} \|\dot{x}(t)\|^2 \leq 0. \]
As a consequence, we have
\[ \mathcal{E}^{\lambda, \epsilon}(t) \leq \mathcal{E}^{\lambda, \epsilon}(t_0), \quad \forall t \in [t_0, +\infty). \]
This implies
\[ f(x(t)) - f(x^*) + \langle \lambda, Ax(t) - b \rangle \leq \frac{1}{\mu e^{t/\delta}} \left( \mathcal{E}^{\lambda, \epsilon}(t_0) + \sup_{t \geq t_0} \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| \int_{t_0}^{+\infty} \|\epsilon(t)\|dt \right). \]
By similar arguments as in the proof of Theorem \ref{thm:3} we have
\[ \sup_{t \geq t_0} \|\frac{1}{\delta} (x(t) - x^*) + \dot{x}(t)\| < +\infty. \]
The rest of the proof is similar to the one of Theorem \ref{thm:2} and so we omit it here.

\begin{remark}
Some similar results on dynamical systems with perturbations for unstrained optimization problem can be found in \cite{10,15,23}.
\end{remark}
3 Rate-matching primal-dual algorithm

It has been shown in [3,6,10,11] that time discretizations of gradient-based dynamical systems naturally lead to numerical algorithms for solving unconstrained convex problems. On the other hand, Wibisono et al. [10] showed that a naive discretization may not yield a stable numerical algorithm that retains the convergence rate of the underlying dynamical system, and may even lead to divergence. So it is interesting to discuss the discretization scheme of the underlying dynamical system, which leads to a rate-matching algorithm. In this section we attempt to discretize the primal-dual dynamic (20) to obtain a rate-matching primal-dual algorithm with the subproblem being solved inexactly. Inspired by [5,8], we adopt the following discrete scheme: The time step size is fixed to one, and set \( t_k = k \), \( x_k = x(t_k) \), \( \lambda_k = \lambda(t_k) \), \( \beta_k = \beta(t_k) \), \( \epsilon_k = \epsilon(t_k) \), and \( \gamma \dot{x}(t) \approx \theta(x_{k+1} - x_k) + (\gamma - \theta)(x_k - x_{k-1}) \). The above discretization scheme of the dynamic (20) with \( f \) being nondifferentiable leads to:

\[
\begin{aligned}
    x_{k+1} - 2x_k + x_{k-1} + \theta(x_{k+1} - x_k) + (\gamma - \theta)(x_k - x_{k-1}) \\
    \in -\beta_k(\partial f(x_{k+1}) + \sigma A^T (Ax_{k+1} - b) + A^T \lambda_{k+1}) + \epsilon_k, \\
    \lambda_{k+1} = \lambda_k + \beta_k (Ax_{k+1} - b + \delta A(x_{k+1} - x_k)).
\end{aligned}
\]

(25a) (25b)

Based on (25), we propose a new primal-dual algorithm (Algorithm 1) to solve the problem (1).

**Algorithm 1:** Primal-dual algorithm for the problem (1)

Initialization: Choose \( x_0 \in \mathbb{R}^n \), \( \lambda_0 \in \mathbb{R}^m \). Set \( x_1 = x_0 \), \( \lambda_1 = \lambda_0 \), \( \epsilon_0 = 0 \). Choose parameters \( \sigma > 0 \), \( \gamma > 0 \), \( \delta > 0 \), \( \theta > -1 \).

for \( k = 1, 2, \cdots \) do

   Step 1: Compute \( \bar{x}_k = x_k + (1 - \frac{\gamma}{1+\theta})(x_k - x_{k-1}) \).

   Step 2: Choose \( \beta_k > 0 \) and \( \epsilon_k \in \mathbb{R}^n \). Set \( \tau_k = \sigma + (1 + \delta)\beta_k \), \( \eta_k = \frac{1}{\lambda_k} (\delta \beta_k Ax_k + (\sigma + \beta_k)b) \).

   Update

   \[
x_{k+1} = \arg\min_x f(x) + \frac{1 + \theta}{2\beta_k}\|x - \bar{x}_k\|^2 + \frac{\eta_k}{2}\|Ax - \eta_k\|^2 + \langle A^T \lambda_k - \frac{\epsilon_k}{\lambda_k}, x \rangle.
\]

   Step 3: \( \lambda_{k+1} = \lambda_k + \beta_k (Ax_{k+1} - b + \delta A(x_{k+1} - x_k)) \).

if stopping condition is satisfied then

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

end

Remark 3 As known, finding an exact solution of the augmented Lagrangian subproblem may be computationally expensive when the problem dimension \( n \) is very large. As a result, many works focused on inexact versions of ALM and its variants (see e.g. [27,29,31,32]). In Step 2 of Algorithm 1 we solve the subproblem inexactly by finding a quasi-solution. Recall that given \( \epsilon > 0 \), \( x_\epsilon \) is an \( \epsilon \)-quasi-solution of the problem (2) if

\[
\Phi(x_\epsilon) \leq \Phi(x) + \epsilon \|x - x_\epsilon\|, \quad \forall x \in \mathbb{R}^n.
\]

Maybe, a quasi-solution is more suitable as an approximate solution as it has been shown by the known Ekeland’s variational principle.
The sequence \( \{(x_k, \lambda_k)\}_{k \geq 1} \) generated by Algorithm 1 satisfies (25).

**Proof** (25) follows directly from Step 3 of Algorithm 1. By Step 2 of Algorithm 1 and using the optimality condition, we get

\[
0 \in \partial f(x_{k+1}) + \frac{1 + \theta}{\beta_k} (x_{k+1} - x_k) + A^T (\tau_k (Ax_{k+1} - \eta_k) + \lambda_k) - \frac{\epsilon_k}{\beta_k},
\]

which yields

\[
(1 + \theta)(x_{k+1} - x_k) \in -\beta_k (\nabla f(x_{k+1}) + A^T (\tau_k (Ax_{k+1} - \eta_k) + \lambda_k)) + \epsilon_k.
\] (26)

From Step 2 and Step 3 we get

\[
\tau_k (Ax_{k+1} - \eta_k) + \lambda_k = (\sigma + (1 + \delta) \beta_k) Ax_{k+1} - \delta \beta_k Ax_k - (\sigma + \beta_k) b + \lambda_k
\]

\[
= \sigma (Ax_{k+1} - b) + \lambda_k + \beta_k (Ax_{k+1} - b + \delta A(x_{k+1} - x_k))
\] (27)

and

\[
(1 + \theta)(x_{k+1} - x_k) = (1 + \theta)(x_{k+1} - x_k) - (\theta + 1 - \gamma)(x_k - x_{k-1})
\]

\[
= x_{k+1} - 2x_k + x_{k-1} + \theta (x_{k+1} - x_k) + (\gamma - \theta)(x_k - x_{k-1}).
\]

This together with (26) and (27) yields the (25a).

**Remark 5** Since the implication in the proof of Lemma 1 is reversible, it is easy to verify that the sequence generated by (25) also satisfies Algorithm 1.

The following two equalities will be used in the proof of the convergence of Algorithm 1 for any \( x, y \in \mathbb{R}^n \),

\[
\frac{1}{2} \| x \|^2 - \frac{1}{2} \| y \|^2 = \langle x, x - y \rangle - \frac{1}{2} \| x - y \|^2,
\] (28)

\[
2 \langle x - y, x - z \rangle = \| x - y \|^2 + \| x - z \|^2 - \| y - z \|^2.
\] (29)

**Theorem 5** Let \( \{(x_k, \lambda_k)\}_{k \geq 1} \) be the sequence generated by Algorithm 1 and \( (x^*, \lambda^*) \in \Omega \). Assume that

\[
\beta_{k+1} \leq (1 + \frac{1}{\delta}) \beta_k, \quad \sum_{k=1}^{+\infty} \| \epsilon_k \| < +\infty, \quad \theta \in (\gamma - 1, \delta \gamma + \gamma - 1).
\]

Then \( \{(x_k, \lambda_k)\}_{k \geq 1} \) is bounded and the following conclusions hold:

(i) \( \sum_{k=1}^{+\infty} ((1 + \frac{1}{\delta}) \beta_k - \beta_{k+1}) (\mathcal{L}^\sigma (x_{k+1}, \lambda^*) - \mathcal{L}^\sigma (x_k, \lambda^*)) < +\infty. \)

(ii) \( \sum_{k=1}^{+\infty} \| x_{k+1} - x_k \|^2 + \| \lambda_{k+1} - \lambda_k \|^2 < +\infty. \)
\((iii)\) If \(\beta_k \equiv \beta > 0\), then

\[
\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O\left(\frac{1}{k}\right), \quad \|A x_{k+1} - b\| = O\left(\frac{1}{\sqrt{k}}\right),
\]

where \(x_{k+1} = \sum_{k=0}^{k-1} x_{k+1}\).

\((iv)\) If \(\lim_{k \to +\infty} \beta_k = +\infty\), then

\[
\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) = O\left(\frac{1}{\beta_k}\right), \quad \|A x_k - b\| = O\left(\frac{1}{\sqrt{\beta_k}}\right).
\]

**Proof** For notation simplicity, we denote \(s = \theta + 1 - \gamma\). Then \(s > 0\) and \(s - \delta \gamma < 0\). Given \(\lambda \in \mathbb{R}^m\), define energy sequences \(\mathcal{E}^\lambda_k\) and \(\mathcal{E}^{\lambda, \epsilon}_k\) by

\[
\mathcal{E}^\lambda_k = s \beta_k (\mathcal{L}^\sigma(x_k, \lambda) - \mathcal{L}^\sigma(x^*, \lambda)) + \frac{1}{2} \|u_k\|^2 + \frac{s(s - \gamma)}{2 \delta^2} \|x_k - x^*\|^2 + \frac{s}{2 \delta} \|\lambda_k - \lambda\|^2
\]

and

\[
\mathcal{E}^{\lambda, \epsilon}_k = \mathcal{E}^\lambda_k - \sum_{j=1}^{k} (u_j, \epsilon_{j-1}),
\]

where

\[
u_k = \frac{s}{\delta} (x_k - x^*) + s (x_k - x_{k-1}).
\]

By the definition of \(\mathcal{L}^\sigma\), we get

\[
\partial_x \mathcal{L}^\sigma(x, \lambda) = \partial f(x) + A^T \lambda + \sigma A^T (Ax - b),
\]

which together with (25a) and (32) implies

\[
u_{k+1} - u_k = \left(\frac{s}{\delta} - \gamma\right) (x_{k+1} - x_k) + x_{k+1} - 2 x_k + x_{k-1} + \theta (x_{k+1} - x_k) + (\gamma - \theta) (x_k - x_{k-1})
\]

\[
\leq \left(\frac{s}{\delta} - \gamma\right) (x_{k+1} - x_k) - \beta_k (\partial f(x_{k+1}) + \sigma A^T (Ax_{k+1} - b) + A^T \lambda)
\]

\[-\beta_k A^T (\lambda_{k+1} - \lambda) + \epsilon_k
\]

\[-\beta_k \partial_x \mathcal{L}^\sigma(x_{k+1}, \lambda) - \beta_k A^T (\lambda_{k+1} - \lambda) + \epsilon_k + \frac{s - \delta \gamma}{\delta} (x_{k+1} - x_k).
\]

As a consequence, we have

\[
\xi_k := -\frac{1}{\beta_k} (u_{k+1} - u_k) - A^T (\lambda_{k+1} - \lambda) + \frac{\epsilon_k}{\beta_k} + \frac{s - \delta \gamma}{\delta \beta_k} (x_{k+1} - x_k) \in \partial_x \mathcal{L}^\sigma(x_{k+1}, \lambda).
\]

It follows from (28) and (34) that

\[
\frac{1}{2} \|u_{k+1}\|^2 - \frac{1}{2} \|u_k\|^2 = \langle u_{k+1}, u_{k+1} - u_k \rangle - \frac{1}{2} \|u_{k+1} - u_k\|^2
\]

\[
\leq -\beta_k \langle u_{k+1}, \xi_k \rangle - \beta_k \langle u_{k+1}, A^T (\lambda_{k+1} - \lambda) \rangle + \langle u_{k+1}, \epsilon_k \rangle + \frac{s - \delta \gamma}{\delta} \langle u_{k+1}, x_{k+1} - x_k \rangle.
\]

Combing (32) with (34), we get

\[
\langle u_{k+1}, \xi_k \rangle = \frac{s}{\delta} \langle x_{k+1} - x^*, \xi_k \rangle + s \langle x_{k+1} - x_k, \xi_k \rangle
\]

\[
\geq \frac{s}{\delta} (\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x^*, \lambda)) + s (\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x_k, \lambda)),
\]
where the inequality is deduced from the convexity of $\mathcal{L}^\sigma(\cdot, \lambda)$. It follows from (32) and (29) that

$$
\langle u_{k+1}, x_{k+1} - x_k \rangle = \frac{s}{\delta}(x_{k+1} - x^*, x_{k+1} - x_k) + s\|x_{k+1} - x_k\|^2
$$

$$
= \frac{s}{2\delta}(\|x_{k+1} - x^*\|^2 + \|x_{k+1} - x_k\|^2 - \|x_k - x^*\|^2) + s\|x_{k+1} - x_k\|^2
$$

$$
= \frac{s}{2\delta}(\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2) + \frac{(2\delta + 1)s}{2\delta}\|x_{k+1} - x_k\|^2.
$$

Combining (35)–(37) together, we have

$$
\frac{1}{2}\|u_{k+1}\|^2 - \frac{1}{2}\|u_k\|^2 + \frac{s}{\delta}(\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2)
$$

$$
\leq -s\beta_k(\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x^*, \lambda)) - s\beta_k(\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x_k, \lambda))
$$

$$
= \beta_k(\langle u_{k+1}, A\mathbf{1}(\lambda_{k+1} - \lambda) \rangle) + \frac{(2\delta + 1)(s - \delta s)}{2\delta}\|x_{k+1} - x_k\|^2.
$$

Since $(x^*, \lambda^*) \in \Omega$ and $Ax^* = b$, it follows from (25b) and (28) that

$$
\frac{s}{\delta}(\|\lambda_{k+1} - \lambda\|^2 - \|\lambda_k - \lambda\|^2)
$$

$$
= \frac{s}{\delta}(\lambda_{k+1} - \lambda, \lambda_{k+1} - \lambda) - \frac{s}{\delta}\|\lambda_{k+1} - \lambda_k\|^2
$$

$$
= \beta_k(\langle Ax_{k+1}, A\mathbf{1}(\lambda_{k+1} - \lambda) \rangle) - \frac{s}{\delta}\|\lambda_{k+1} - \lambda_k\|^2.
$$

By computation, we get

$$
\beta_{k+1}(\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x^*, \lambda)) - \beta_k(\mathcal{L}^\sigma(x_k, \lambda) - \mathcal{L}^\sigma(x^*, \lambda))
$$

$$
= \beta_{k+1} - \beta_k(\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x^*, \lambda)) + \beta_k(\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x_k, \lambda)).
$$

It follows from (31), (33) and (38–40) that

$$
\mathcal{E}_{k+1}^{\lambda, \epsilon} - \mathcal{E}_k^{\lambda, \epsilon} = \mathcal{E}_k^{\lambda, \epsilon} - \langle u_{k+1}, \epsilon_k \rangle
$$

$$
= \frac{1}{2}\|u_{k+1}\|^2 - \frac{1}{2}\|u_k\|^2 + \frac{s}{\delta}(\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2)
$$

$$
= s\beta_k(\mathcal{L}^\sigma(x_{k+1}, \lambda) - \mathcal{L}^\sigma(x^*, \lambda)) - \beta_k(\mathcal{L}^\sigma(x_k, \lambda) - \mathcal{L}^\sigma(x^*, \lambda))
$$

$$
\leq s\beta_k(1 + \frac{1}{\delta})\beta_k(\|x_{k+1} - x_k\|^2 - \|\lambda_{k+1} - \lambda_k\|^2).
$$

Since $\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*) \geq 0$, $s > 0$, $s - \delta s < 0$ and $\beta_{k+1} \leq (1 + \frac{1}{\delta})\beta_k$, taking $\lambda = \lambda^*$ in (41), we obtain

$$
\mathcal{E}_{k+1}^{\lambda, \epsilon} \leq \mathcal{E}_k^{\lambda, \epsilon}, \quad \forall k \geq 1.
$$

By the definition of $\mathcal{E}_k^{\lambda, \epsilon}$ and $\mathcal{E}_k^{\lambda, \epsilon}$, we have

$$
\frac{1}{2}\|u_k\|^2 \leq \mathcal{E}_k^{\lambda, \epsilon} \leq \mathcal{E}_1^{\lambda, \epsilon} + \sum_{j=1}^{k} \langle u_j, \epsilon_{j-1} \rangle \leq \mathcal{E}_1^{\lambda, \epsilon} + \sum_{j=1}^{k} \|u_j\|\|\epsilon_{j-1}\|.
$$

Applying Lemma 8 to (43), we get

$$
\sup_{k \geq 1}\|u_k\| \leq \sqrt{2\mathcal{E}_1^{\lambda, \epsilon} + 2\sum_{j=1}^{+\infty} \|\epsilon_j\|} < +\infty,
$$

(44)
which together with (43) yields
\[ \sup_{k \geq 1} \mathcal{E}_k^{\lambda^*} \leq \mathcal{E}_1^{\lambda^*} + \sup_{k \geq 1} \| u_k \| \sum_{j=1}^{+\infty} \| \epsilon_j \| < +\infty. \]

So \{\mathcal{E}_k^{\lambda^*}\}_{k \geq 1} is bounded. This together with (40) implies that \{\|x_k - x^*\|\}_{k \geq 1} and \{\|\lambda_k - \lambda^*\|\}_{k \geq 1} are both bounded. As a consequence, \{(x_k, \lambda_k)\}_{k \geq 1} is bounded. On the other hand, by (31) and (44) we have
\[ \inf_{k \geq 1} \mathcal{E}_k^{\lambda^*, \epsilon} \geq - \sup_{k \geq 1} \| u_k \| \sum_{j=1}^{+\infty} \| \epsilon_j \| > -\infty. \]

This together with (42) implies that \{\mathcal{E}_k^{\lambda^*, \epsilon}\}_{k \geq 1} is bounded.

Summing the inequality (41) over \( k = 1, 2, \ldots, n \), we obtain
\[
\sum_{k=1}^{n} (1 + \frac{1}{\delta}) \beta_k - \beta_{k+1} \left( \mathcal{L}^\sigma (x_{k+1}, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*) \right)
+ \frac{(2\delta + 1)(\delta\gamma - s)}{2\delta^2} \sum_{k=1}^{n} \| x_{k+1} - x_k \|^2 + \frac{1}{2\delta} \sum_{k=1}^{n} \| \lambda_{k+1} - \lambda_k \|^2 \]
\[ \leq \frac{1}{8} (\mathcal{E}_1^{\lambda^*, \epsilon} - \mathcal{E}_{n+1}^{\lambda^*, \epsilon}). \]

Since \{\mathcal{E}_k^{\lambda^*, \epsilon}\}_{k \geq 1} is bounded, from (45) we get
\[
\sum_{k=1}^{n} (1 + \frac{1}{\delta}) \beta_k - \beta_{k+1} \left( \mathcal{L}^\sigma (x_{k+1}, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*) \right) < +\infty
\]
and
\[ + \sum_{k=1}^{+\infty} \| x_{k+1} - x_k \|^2 + \| \lambda_{k+1} - \lambda_k \|^2 < +\infty. \]

The proof of (i) and (ii) is complete.

Next, we prove (iii). In the case \( \beta_k \equiv \beta > 0 \), from (i) we have
\[
\sum_{k=1}^{n} \mathcal{L}^\sigma (x_{k+1}, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*) < +\infty. \tag{46}
\]

Since \( \mathcal{L}^\sigma (\cdot, \lambda^*) \) is convex, we have
\[
\mathcal{L}^\sigma (\bar{x}_{k+1}, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*) = \mathcal{L}^\sigma \left( \frac{\sum_{i=1}^{k} x_{i+1} \bar{x}}{k}, \lambda^* \right) - \mathcal{L}^\sigma (x^*, \lambda^*)
\leq \frac{1}{k} \sum_{i=1}^{k} \left( \mathcal{L}^\sigma (x_{i+1}, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*) \right)
\leq \frac{1}{k} \sum_{i=1}^{+\infty} \left( \mathcal{L}^\sigma (x_{i+1}, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*) \right),
\]
which together with the definition of \( \mathcal{L}^\sigma \) and (46) yields (iii).

Now consider the case \( \lim_{k \to +\infty} \beta_k = +\infty \). Since \{\mathcal{E}_k^{\lambda^*}\}_{k \geq 1} is bounded, we have
\[
\sup_{k \geq 1} \beta_k (\mathcal{L}^\sigma (x_k, \lambda^*) - \mathcal{L}^\sigma (x^*, \lambda^*)) = \sup_{k \geq 1} \beta_k (\mathcal{L}(x_k, \lambda^*) - \mathcal{L}(x^*, \lambda^*) + \frac{\sigma}{2} \| Ax_k - b \|^2) \leq \sup_{k \geq 1} \mathcal{E}_k^{\lambda^*} < +\infty,
\]
which yields (iv).
When strengthening the assumption on $\beta_k$, we can improve the convergence rate from $O$ to $o$.

**Theorem 6** Let $\{(x_k, \lambda_k)\}_{k \geq 1}$ be the sequence generated by Algorithm 7 and $(x^*, \lambda^*) \in \Omega$. Assume that
\[
\beta_{k+1} \leq (1 - \kappa)(1 + \frac{1}{\beta})\beta_k, \quad \sum_{k=0}^{+\infty} \|\epsilon_k\| < +\infty, \quad \theta \in (\gamma - 1, \delta \gamma + \gamma - 1),
\]
where $\kappa \in (0, 1)$. If $\lim_{k \to +\infty} \beta_k = +\infty$, then
\[
\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*) = o\left(\frac{1}{\beta_k}\right), \quad \|Ax_{k+1} - b\| = o\left(\frac{1}{\sqrt{\beta_k}}\right).
\]

**Proof** Since $\beta_{k+1} \leq (1 - \kappa)(1 + \frac{1}{\beta})\beta_k$, by Theorem 5 (i) we obtain
\[
\sum_{k=1}^{+\infty} \beta_k(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) < +\infty,
\]
which implies
\[
\lim_{k \to +\infty} \beta_k(\mathcal{L}(x_{k+1}, \lambda^*) - \mathcal{L}(x^*, \lambda^*)) = 0 \quad \text{and} \quad \lim_{k \to +\infty} \beta_k\|Ax_{k+1} - b\|^2 = 0.
\]
This yields the desired results.

Next, we consider the case $\beta_k = \mu(1 + \frac{1}{\beta})^k$ with $\mu > 0$. In this case, $\beta_{k+1} = (1 + \frac{1}{\beta})\beta_k$.

**Theorem 7** Let $\{(x_k, \lambda_k)\}_{k \geq 1}$ be the sequence generated by Algorithm 7 $(x^*, \lambda^*) \in \Omega$ and $\beta_k = \mu(1 + \frac{1}{\beta})^k$ with $\mu > 0$. Assume that
\[
\sum_{k=1}^{+\infty} \|\epsilon_k\| < +\infty, \quad \theta \in (\gamma - 1, \delta \gamma + \gamma - 1).
\]
Then $\{(x_k, \lambda_k)\}_{k \geq 1}$ is bounded and the following conclusions hold:

(i) $\sum_{k=1}^{+\infty} \|x_{k+1} - x_k\|^2 + \|\lambda_{k+1} - \lambda_k\|^2 < +\infty$.

(ii) $|f(x_k) - f(x^*)| = O\left(\frac{1}{(1 + 1/\beta)^k}\right), \quad \|Ax_k - b\| = O\left(\frac{1}{(1 + 1/\beta)^k}\right)$.

**Proof** By Theorem 5 we obtain (i) and the boundedness of $\{(x_k, \lambda_k)\}_{k \geq 1}$. Given $\lambda \in \mathbb{R}^m$, define the $\mathcal{E}_k^\lambda$ and $\mathcal{E}_k^{\lambda, \epsilon}$ respectively as in (30) and (31) with $\beta_k = \mu(1 + \frac{1}{\beta})^k$. Since $\beta_{k+1} = (1 + \frac{1}{\beta})\beta_k$, it follows from (31) that
\[
\mathcal{E}_k^{\lambda, \epsilon} - \mathcal{E}_k^\lambda \leq \frac{(2\delta + 1)(s - \delta \gamma)s}{2\delta s} \|x_{k+1} - x_k\|^2 - \frac{s}{2\delta} \|\lambda_{k+1} - \lambda_k\|^2 \leq 0,
\]
where $s = \theta + 1 - \gamma > 0$. As a consequence, we get
\[
\mathcal{E}_k^\lambda \leq \mathcal{E}_1^{\lambda, \epsilon}, \quad \forall \ k \geq 1, \ \lambda \in \mathbb{R}^m.
\]
By the definitions of $\mathcal{E}_k^\lambda$ and $\mathcal{E}_k^{\lambda, \epsilon}$, we obtain
\[
f(x_k) - f(x^*) + \langle \lambda, Ax_k - b \rangle \leq \frac{1}{s\mu(1 + 1/\beta)^k} \left(\mathcal{E}_1^{\lambda, \epsilon} + \sup_{k \geq 1} \|u_k\| \sum_{j=1}^{+\infty} \|\epsilon_j\|\right).
\]
Take \( \rho > \|\lambda^*\| \). Application of Lemma 4 to the above inequality yields

\[
    f(x_k) - f(x^*) + \rho \|Ax_k - b\| \leq \frac{1}{s\mu(1 + 1/\delta)^k} \left( \sup_{\|\lambda\| \leq \rho} \mathcal{E}_1^{\lambda, \epsilon} + \sup_{k \geq 1} \|u_k\| \sum_{j=1}^{+\infty} \|\epsilon_j\| \right). \tag{47}
\]

By similar arguments as in the proof Theorem 5, we have

\[
    \sup_{k \geq 1} \|u_k\| < +\infty,
\]

this together with assumptions implies

\[
    C := \sup_{\|\lambda\| \leq \rho} \mathcal{E}_1^{\lambda, \epsilon} + \sup_{k \geq 1} \|u_k\| \sum_{j=1}^{+\infty} \|\epsilon_j\| < +\infty. \tag{48}
\]

It follows from (48) that

\[
    f(x_k) - f(x^*) \geq -\|\lambda^*\| \|Ax_k - b\|.
\]

This together with (47) and (48) yields

\[
    \|Ax_k - b\| \leq C \frac{s\mu(\rho - \|\lambda^*\|)(1 + 1/\delta)^k}{s\mu(1 + 1/\delta)^k} \leq f(x_k) - f(x^*) \leq \frac{C}{s\mu(1 + 1/\delta)^k}.
\]

Using the fact \( 0 \leq C < +\infty \), we get (ii).

**Remark 6** As shown in Theorem 5 and Theorem 7, the convergence rate of Algorithm 1 matches that of the underlying dynamic (20).

Now we investigate the convergence of the sequence \( \{(x_k, \lambda_k)\}_{k \geq 1} \) generated by Algorithm 1 under the condition \( \beta_{k+1} \leq (1 - \kappa)(1 + \frac{1}{\delta})\beta_k \).

**Theorem 8** Let \( \{(x_k, \lambda_k)\}_{k \geq 1} \) be the sequence generated by Algorithm 1. Assume that

\[
    \beta_{k+1} \leq (1 - \kappa)(1 + \frac{1}{\delta})\beta_k, \quad \liminf_{k \to +\infty} \beta_k > 0, \quad \theta \in (\gamma - 1, \delta\gamma + \gamma - 1), \quad \sum_{k=1}^{+\infty} \|\epsilon_k\| < +\infty,
\]

where \( \kappa \in (0, 1) \). Then \((x_k, \lambda_k)\) converges to some point of \( \Omega \) as \( k \to +\infty \).

**Proof** Let \( (x^*, \lambda^*) \in \Omega \). By Theorem 5(i) and (ii), we have

\[
    \lim_{k \to +\infty} \|x_{k+1} - x_k\| = 0 \tag{49}
\]

and

\[
    \lim_{k \to +\infty} \beta_k (\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) = 0,
\]

which yields

\[
    \lim_{k \to +\infty} \beta_{k+1}(\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) = \lim_{k \to +\infty} \beta_{k+1} \beta_k(\mathcal{L}^\sigma(x_{k+1}, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) = 0. \tag{50}
\]
Combining (42) with the definition of $\mathcal{E}_k^{\lambda^*}$, we have

$$\mathcal{E}_{k+1}^\ast \leq \mathcal{E}_k^\ast + \langle u_{k+1}, \epsilon_k \rangle \leq \mathcal{E}_k^\ast + \|u_{k+1}\|\|\epsilon_k\|.$$  

(51)

By (44) and the assumption, we know $\sum_{k=0}^{+\infty} \|u_{k+1}\|\|\epsilon_k\| < +\infty$. As an application of Lemma 5 to (31), $\{\mathcal{E}_k^\ast\}_{k \geq 1}$ converges to some point. It follows that

$$\lim_{k \to +\infty} \mathcal{E}_k^\ast = \lim_{k \to +\infty} \left( s\beta_k (\mathcal{L}^\sigma(x_k, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*)) + \frac{1}{2}\|s\| \|x_k - x^*\|^2 + \|x_k - x^*\|^2 \|\lambda_k - \lambda^*\|^2 \right)$$

$$+ s\beta_k (\mathcal{L}^\sigma(x_k, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*))$$

$$= \lim_{k \to +\infty} \left( \frac{s^2}{2\delta} \|x_k - x^*\|^2 + \frac{s^2}{2\delta} \|\lambda_k - \lambda^*\|^2 \|x_k - x^*\|^2 \right)$$

$$= \lim_{k \to +\infty} \frac{s^2}{2\delta} \|x_k - x^*\|^2 + \|\lambda_k - \lambda^*\|^2,$$

where $s = \theta + 1 - \gamma$, the first equality is from the definition of $\mathcal{E}_k^\ast$, the second equality is from (49) and (50), and the last equality is deduced from the fact that $\{x_k\}_{k \geq 1}$ is bounded and $\lim_{k \to +\infty} \|x_{k+1} - x_k\| = 0$. As a consequence,

$$\lim_{k \to +\infty} \gamma \|x_k - x^*\|^2 + \|\lambda_k - \lambda^*\|^2$$

exists for any $(x^*, \lambda^*) \in \Omega$. Since $\liminf_{k \to +\infty} \beta_k > 0$, it follows from (50) that

$$\lim_{k \to +\infty} \mathcal{L}^\sigma(x_k, \lambda^*) - \mathcal{L}^\sigma(x^*, \lambda^*) = 0,$$

which together with the definition of $\mathcal{L}^\sigma$ yields

$$\lim_{k \to +\infty} \|Ax_k - b\| = 0.$$

Let $(\tilde{x}, \bar{\lambda})$ be a cluster point of $\{(x_k, \lambda_k)\}_{k \geq 1}$. Then $\|Ax - b\| = 0$. Since $\lim_{k \to +\infty} \epsilon_k = 0$, it follows from (25a), (49) and Proposition 20.38 that $A^T\bar{\lambda} \in \partial f(\tilde{x})$. By (49), we get $(\tilde{x}, \bar{\lambda}) \in \Omega$. Applying Lemma 3 with $\omega_k = (x_k, \lambda_k)$ and $M = \begin{pmatrix} \gamma I_n & I_m \\ I_m & I_m \end{pmatrix}$, where $I_n$ is the identity matrix of order $n$, we get the desired result.

4 Numerical Experiments

Consider the equality constrained quadratic programming problem (ECQP):

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

where $Q \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$. The optimal value $f(x^*)$ is obtained by Matlab function quadprog with tolerance $10^{-15}$. Let $Q, A, b, q$ be generated by the standard Gaussian distribution. Set the parameters of Algorithm 4 as below:

$$\sigma = 0.1, \ \gamma = 1, \ \delta = \frac{1}{n}, \ \theta = \gamma - 1 + \frac{\delta \gamma}{2}.$$  

(52)

Taking $m = 200$ and $n = 500$, we test Algorithm 4 on ECQP under different choices of $\beta_k$ and $\epsilon_k$. As shown in Fig. 4, a growing $\beta_k$ contributes to achieving a fast convergence. Fig. 4 also shows that
$\epsilon_k$ is smaller, the subproblem is solved more accurately, and then the numerical results are closer to the exact one. Indeed, the term $\epsilon_k$ in Step 2 of Algorithm 1 can be understood as an error when solving the subproblem. More precisely, in Step 2 we find an $\|\epsilon_k\|_{\beta_k}$-quasi-solution of the subproblem, instead of the exact solution.

Fig. 1: Error of the objective function and the constraint for Algorithm 1 with different $\beta_k$ and $\epsilon_k$

In Fig. 2 taking $\epsilon_k \equiv 0$, we compare Algorithm 1 with the accelerated linearized augmented Lagrangian method (AALM) in [42, Algorithm 1] with $\alpha_k = \frac{1}{k^2}$, $\beta_k = \gamma_k = mk$ and $P_k = Id$, as well as the classical augmented Lagrangian method (ALM) [35], which is [42, Algorithm 1] with fixed parameter $\alpha_k = 1$, $\beta_k = \gamma_k = m$ and $P_k = Id$.

Fig. 2: Error of objective function and constraint of Algorithm 1, ALM and AALM
Next, consider the nonnegative linearly constrained quadratic programming problem (NLCQP):

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

In this case, \( f(x) = \frac{1}{2} x^T Qx + q^T x + I_{\{y|y \geq 0\}}(x) \), where \( I_{\{y|y \geq 0\}}(\cdot) \) is the indicator function of the set \( \{y|y \geq 0\} \), i.e.,

\[
I_{\{y|y \geq 0\}}(x) = \begin{cases} 
0, & x \geq 0, \\
+\infty, & \text{otherwise}.
\end{cases}
\]

As a comparison, we test Algorithm 1 and AALM on NLCQP. Take \( m = 100 \) and \( n = 500 \).

Let \( Q = H^T H \) and \( A = [B, Id] \), where \( H \in \mathbb{R}^{n \times n} \) and \( B \) are generated by the standard Gaussian distribution. Let \( q \) be generated by the standard Gaussian distribution and \( b \) be generated by the uniform distribution. Set the parameters of Algorithm 1 as in (52), \( \beta_k = \frac{mk^2}{e^{-t/\delta}} \), and \( \epsilon_k = 0 \), and set the parameters of AALM as \( \alpha_k = \frac{1}{\sqrt{k}} \), \( \beta_k = \gamma_k = mk \) and \( P_k = \frac{2\|Q\|}{k} Id \). Subproblems for both algorithms are solved by interior-point algorithms to a tolerance \( \text{subtol} \).

Algorithm 1 performs better and more stable than AALM under different \( \text{subtol} \).

![Fig. 3: Error of objective function and constraint of Algorithm 1 and AALM with different subtol](image)

5 Conclusion

In this paper, we propose a primal-dual dynamic with time scaling the problem (1), which includes a second-order ODE for the prime variable and a fist-order ODE for the dual variable. Under a scaling condition, we prove that the proposed dynamic as well as its perturbed version enjoys a fast convergence property: \( \mathcal{L}(x(t), \lambda^*) - \mathcal{L}(x^*, \lambda^*) = \mathcal{O}\left(\frac{1}{\beta(t)}\right) \). In the case \( \beta(t) = \mu e^{-t/\delta} \), we derive an improved convergence rate: \( |f(x(t)) - f(x^*)| = \mathcal{O}(\frac{1}{e^{-t/\delta}}) \). By a suitable discrization scheme, we also...
Fast convergence of primal-dual dynamics and algorithms with...

present a fast inexact primal-dual algorithm for the problem (1), and prove that its convergence rate matches that of the underlying dynamic. The fast convergence of our approaches is realized without strong convexity of the objective function.

Appendix: Some auxiliary results

Lemma 2 \[\text{[15, Lemma A.5]}\] Let $\nu : [t_0, T] \to [0, +\infty)$ be integrable and $M \geq 0$. Suppose that $\mu : [t_0, T] \to R$ 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 3 \[\text{[4, Lemma 5.14]}\] Let $\{a_k\}_{k \geq 1}$ and $\{b_k\}_{k \geq 1}$ be two nonnegative sequences such that $\sum_{k=1}^{+\infty} b_k < +\infty$ and

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

where $c \geq 0$. Then

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

for all $k \in \mathbb{N}$.

Lemma 4 \[\text{[4, Lemma 2.2]}\] Let $x^*$ be a solution of the problem (1). Given a function $\phi$ and a fix point $x$, if for any $\lambda$ it holds that

$$f(x) - f(x^*) + \langle \lambda, Ax - b \rangle \leq \phi(\lambda),$$

then for any $\rho > 0$, we have

$$f(x) - f(x^*) + \rho \|Ax - b\| \leq \sup_{\|\lambda\| \leq \rho} \phi(\lambda).$$

Lemma 5 \[\text{[4, Lemma 5.31]}\] Let $\{a_k\}_{k \geq 1}$ and $\{\xi_k\}_{k \geq 1}$ be sequences in $R^+$ satisfying $\sum_{k=1}^{+\infty} \xi_k < +\infty$ and

$$a_{k+1} \leq a_k + \xi_k, \quad \forall k \geq 1.$$

Then $\{a_k\}_{k \geq 1}$ converges.

Lemma 6 \[\text{[4, Theorem 3.3]}\] Let $S$ be a nonempty subset of $R^n$ space and $\{\omega_k\}_{k \geq 1}$ be a sequence in $R^n$, $M \in R^{n \times n}$ be a positive definite matrix. Assume that

(i) for all $\omega^* \in S$, $\lim_{k \to +\infty} \|\omega_k - \omega^*\|_M$ exists;

(ii) every sequential cluster point of $\omega_k$ belongs to $S$.

Then $\omega_k$ converges to a point in $S$ as $k \to +\infty$.

References

1. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control Optim. 38(4), 1102-1119 (2000)
2. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. J. Math. Pures Appl. 81, 747-779 (2002)
3. Attouch, H., Cabot, A.: Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. J. Differential Equations. 263(9), 5412-5458 (2017)
4. 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)
5. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. SIAM J. Optim. 29(3), 2227-2256 (2019)
6. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. ESAIM: Control Optim. Calc. Var. 25(2) (2019)

7. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. 174(1-2), 391-432 (2019)

8. Balhag, A., Chbani, Z., Riahi, H.: Linear convergence of inertial gradient dynamics with constant viscous damping coefficient and time-dependent rescaling parameter (2020). hal-02610699

9. Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces: Second Edition. Springer, New York (2016)

10. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2(1), 183-202 (2009)

11. Bolte, J.: Continuous gradient projection method in Hilbert spaces. J. Optim. Theory Appl. 119(2), 235-259 (2003)

12. Boţ, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control Optim. 54(3), 1423-1443 (2016)

13. Boţ, R.I., Csetnek, E.R., Lásló, S.C.: A primal-dual dynamical approach to structured convex minimization problems. J. Differential Equations. 269(12), 10717-10757 (2020)

14. 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)

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

16. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. 25, 161-319 (2016)

17. Chen, S.S., Donoho, D.L.: Saunders MA. Atomic decomposition by basis pursuit. SIAM Rev. 43(1), 129-159 (2001)

18. Cherukuri, A., Mallada, E., Cortés, J.: Asymptotic convergence of constrained primal-dual dynamics. Systems Control Lett. 87, 10-15 (2016)

19. Combettes, P.L., Vu, B.C.: Variable metric quasi-Fejér monotonicity. Nonlinear Anal. 78, 17-31 (2013)

20. Fazlyab, M., Koppel, A., Preciado, V.M., Ribeiro, A.: A variational approach to dual methods for constrained convex optimization. 2017 American Control Conference (ACC). 5269-5275 (2017)

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

22. 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)

23. Haraux, A., Jendoubi, M.A.: On a second order dissipative ODE in Hilbert space with an integrable source term. Acta Math. Sci. 32(1), 155-163 (2012)

24. He, X., Hu, R., Fang, Y.P.: Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems (2020). [arXiv:2007.12128]

25. He, B., Yuan, X.: On the acceleration of augmented Lagrangian method for linearly constrained optimization. Optimization online. 3 (2010)

26. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. 4(5), 303-320 (1969)

27. Kang, M., Kang, M., Jung, M.: Inexact accelerated augmented Lagrangian methods. Comput. Optim. Appl. 62(2), 373-404 (2015)

28. 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)

29. Lan, G., Monteiro, R.D.C.: Iteration-complexity of first-order augmented Lagrangian methods for convex programming. Math. Program. 155(1-2), 511-547 (2016)

30. Lee, S., Chatzipanagiotis, N., Zavlanos, M.M.: Complexity certification of a distributed augmented Lagrangian method. IEEE Trans. Automat. Control. 63(3), 827-834 (2018)

31. Liu, Y.F., Liu, X., Ma, S.Q.: On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. Math. Oper. Res. 44(2), 632-650 (2019)

32. Nedelcu, V., Necoara, I., Tran-Dinh, Q.: Computational complexity of inexact gradient augmented Lagrangian methods: application to constrained MPC. SIAM J. Control Optim. 52(5), 3109-3134 (2014)
33. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In Sov. Math. Dokl 27(2), 372-376 (1983)
34. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. 4(5), 1-17 (1964)
35. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R., Ed., Optimization, Academic Press, New York (1969)
36. 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)
37. Teschl, G.: Ordinary differential equations and dynamical systems. American Mathematical Soc (2012)
38. Tian, W., Yuan, X.: An alternating direction method of multipliers with a worst-case $O(1/n^2)$ convergence rate. Math. Comp. 88(318), 1685-1713 (2019)
39. Tran-Dinh, Q., Zhu, Y.: Non-stationary first-order primal-dual algorithms with faster convergence rates. SIAM J. Optim. 30(4), 2866-2896 (2020)
40. Wibisono, A., Wilson, A.C., Jordan, M.I.: A variational perspective on accelerated methods in optimization. Proc. Natl. Acad. Sci. USA. 113(47), E7351-E7358 (2016)
41. Wilson, A.C., Recht, B., Jordan, M.I.: A Lyapunov analysis of momentum methods in optimization (2016). arXiv:1611.02635
42. Xu, Y.: Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. SIAM J. Optim. 27(3), 1459-1484 (2017)
43. Zeng, X., Lei, J., Chen, J.: Dynamical primal-dual accelerated method with applications to network optimization (2019). arXiv:1912.03650
44. Zeng, X., Yi, P., Hong, Y.: Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach. IEEE Trans. Automat. Control. 62(10), 5227-5233 (2017)