Fast convergence of inertial primal-dual dynamics and algorithms for a bilinearly coupled saddle point problem*

Ke-wei Ding[a,b], Jörg Fliege[c] and Phan Tu Vuong[c†]

[a] School of Mathematics, Southwest Minzu University, Chengdu, Sichuan 610041, P.R. China

[b] School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, P.R. China

[c] Mathematical Sciences School, University of Southampton, Southampton SO17 1BJ, UK

Abstract. This paper is devoted to study the convergence rates of a second-order dynamical system and its corresponding discretization associated with a continuously differentiable bilinearly coupled convex-concave saddle point problem. First, we consider the second-order dynamical system with asymptotically vanishing damping term and show the existence and uniqueness of the trajectories as global twice continuously differentiable solutions. We derive the convergence rate for the primal-dual gap along the generated trajectories for all damping coefficients $\alpha > 0$ and prove that the primal-dual trajectory of the second-order dynamical system asymptotically weakly converges to a primal-dual optimal solution of the original saddle point problem when $\alpha > 3$. In addition, we obtain a faster convergence rate for the second-order dynamical system with the assumption of strongly convexity. Second, we develop the corresponding inertial algorithm which results from the discretization of the dynamical system and prove convergence properties for the primal-dual gap and the sequence of iterates. We show that the sequence of iterates generated by the inertial algorithm weakly converges to a primal-dual optimal solution which is compatible with the fact in the continuous setting.

Key Words: Saddle point problem; Damped inertial dynamics; Convergence rates; Numerical algorithm; Nesterov's accelerated gradient method; Iterates convergence

2010 Mathematics Subject Classification: 37N40; 49J35; 65K10; 68Q25; 90C25

1 Introduction

Let $\mathcal{X}, \mathcal{Y}$ be two real Hilbert spaces equipped with inner products $\langle \cdot, \cdot \rangle_{\mathcal{X}}, \langle \cdot, \cdot \rangle_{\mathcal{Y}}$ (abbreviated $\langle \cdot, \cdot \rangle$) and norms $\| \cdot \|_{\mathcal{X}} = \langle \cdot, \cdot \rangle_{\mathcal{X}}^{\frac{1}{2}}, \| \cdot \|_{\mathcal{Y}} = \langle \cdot, \cdot \rangle_{\mathcal{Y}}^{\frac{1}{2}}$ (abbreviated $\| \cdot \|$). The mapping $A : \mathcal{X} \to \mathcal{Y}$ is a continuous linear operator with induced norm

$$\|A\| = \max \{ \|Ax\| : x \in \mathcal{X} \text{ with } \|x\| \leq 1 \} .$$

*This work was supported by the Fundamental Research Funds for the Central Universities, Southwest Minzu University (2021128), China Scholarship Council (202100850001).

†To whom all correspondences should be addressed: t.v.phan@soton.ac.uk

In this paper, we consider the following bilinearly coupled convex-concave saddle point problem

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} L(x,y) \equiv f(x) + \langle Ax, y\rangle - g(y), \tag{1.1}$$

where $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$ are two continuously differentiable convex functions. Here we call $\langle Ax, y\rangle$ the bilinear coupling term. Problem (1.1) is an important model for many applications arising in various areas, such as imaging procssing [16, 19], reinforcement learning [31, 21], and robust Learning and empirical risk minimization [26].

A pair $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is called a saddle point of the function L if for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$L(x^*, y) \le L(x^*, y^*) \le L(x, y^*).$$

We denote by $\mathbb{S}$ the set of saddle points of problem (1.1). We assume that problem (1.1) has at least one optimal solution (x^*, y^*), which also satisfies the following KKT conditions

$$\begin{cases} \nabla f(x^*) + A^* y^* = 0, \\ Ax^* - \nabla g(y^*) = 0, \end{cases} \tag{1.2}$$

where A^* is the adjoint operator of A. Define the operator $\mathcal{T}_L : \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y}$ as

$$\mathcal{T}_L(x,y) = \begin{pmatrix} \nabla_x L(x,y) \\ \nabla_y L(x,y) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + A^* y \\ \nabla g(y) - Ax \end{pmatrix}. \tag{1.3}$$

It is obvious that the optimality condition (1.2) can be reformulated as $\mathcal{T}_L(x_*, y_*) = 0$ and $\mathbb{S}$ can be viewed as the set of zeros of the operator $\mathcal{T}_L$. Since $f(\cdot)$ (resp. $g(\cdot)$) is convex and continuously differentiable and A (resp. A^*) is a linear operator, it is obvious that $\mathcal{T}_L(x, y)$ is maximally monotone (see,[10],Corollary 20.28). Moreover, we notice that $\mathbb{S}$ can be interpreted as the set of zeros of the maximally monotone operator $\mathcal{T}_L$ and so $\mathbb{S}$ is closed and convex.

We recall some significant primal-dual algorithms for the saddle point problem (1.1). Chambolle and Pock [16] show an ergodic convergence rate $O(1/k)$ of their celebrated primal-dual algorithm for the primal-dual gap of problem (1.1) with f and g are proper, convex and lower semicontinuous. Based on the primal-dual algorithm described in [16], He et al. [22] propose a generalized primal-dual algorithm and relax the condition for ensuring its convergence, and obtain the convergence rate $O(1/k)$ in both the ergodic and pointwise sense. If f is strongly convex, Chambolle and Pock [17] obtain a faster ergodic convergence rate $O\left(1/k^2\right)$ for the same primal-dual gap. With the assumption that f is a convex and Fréchet differentiable function with L-Lipschitz continuous gradient, Chen et al. [19] present an ergodic convergence rate $O\left(L/k + \|A\|/k^2\right)$ for the primal-dual gap of problem (1.1). On the other hand, with the assumption that both f and g are smooth, Kovalev et al. [25] propose an accelerated primal-dual gradient method for solving the saddle point problem and presented the linear convergence rate when the objective function is strongly convex-concave, convex-strongly concave, or even just convex-concave. Thekumparampil et al. [31] developed a lifted Primal-Dual first order algorithm and show a lower complexity bound under the assumption that f and g are both strongly convex smooth functions. More results regarding (1.1) can be found in [16, 17, 19, 20, 22] and references therein.

1.1 Fast primal-dual algorithm via dynamical system

Recently, continuous-time dissipative dynamical systems have been extensively studied in the context of solving various different optimization problems. Alvarez and Attouch [1, 2, 8] studied second order inertial dynamics with fixed viscous damping coefficient, in line with the seminal work of Polyak on the heavy ball method with friction for unconstrained optimization problems. A decisive step was taken by Su et al. in [30], where, for the minimization of a continuously differentiable convex function $\Phi : X \to \mathbb{R}$, the authors considered the following second order inertial dynamic with asymptotic vanishing viscous damping

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0,\ t > 0. \tag{1.4}$$

The authors successfully link the inertial dynamic (1.4) with the accelerated gradient method of Nesterov [12, 28] in the case $\alpha = 3$. Moreover, Attouch et al. [9] showed that any trajectory of (1.4) converges weakly to a minimizer of Φ when $\alpha > 3$ and establish the strong convergence properties in various practical situations. In addition, Attouch and Peypouquet [4] and May [27] showed that the asymptotic convergence rate of (1.4) is $o\left(1/k^2\right)$ when $\alpha > 3$. When $\alpha \leq 3$, Attouch et al. [5] and Apidopoulos et al. [3] present the convergence rate of the objective values $O\left(1/t^{\frac{2\alpha}{3}}\right)$ for the continuous dynamical system (1.4) as well as a corresponding numerical algorithm.

Subsequetnly, the inertial dynamic method has been generalized to linear equality constrained convex optimization problems by employing the augmented Lagrangian approach. Correspondingly, He [23] investigated asymptotic properties of a second-order continuous primal-dual dynamical system with viscous damping and extrapolation for a separable convex optimization problem with linear equality constraints. Attouch et al. [6] introduced a second-order continuous dynamical system with viscous damping, extrapolation, and temporal scaling for linear equality constrained convex optimization problems and paved the way for developing the corresponding accelerated alternating direction method of multipliers (ADMM) algorithms via temporal discretization. Boţ and Nguyen [14] discussed the convergence behavior of the primal-dual gap, the feasibility measure, the objective function value and trajectory for a second-order dynamical system with asymptotically vanishing damping term. Recently, Boţ et al.[15] presented the corresponding numerical optimization algorithm originating from the second-order dynamical system in [14]. They also provided convergence results regarding the sequence of iterates generated by a fast algorithm for linearly constrained convex optimization problems without additional assumptions such as strong convexity.

It thus seems natural to employ the dynamical system framework to study bilinearly coupled convex-concave saddle point problems. It is worth mentioning here that Li et al. [26] provided a novel first order algorithm based on continuous-time dynamical systems for a smooth bilinearly-coupled strongly-convex-concave saddle point problem and showed matching polynomial convergence behavior in discrete time.

In this paper, we consider the following second order primal dual dynamical system

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla_x L\left(x(t), y + \theta t\dot{y}(t)\right) = 0, \\ \ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t) - \nabla_y L\left(x(t) + \theta t\dot{x}(t), y\right) = 0, \\ (x(t_0), y(t_0)) = (x_0, y_0) \text{ and } (\dot{x}(t_0), \dot{y}(t_0)) = (\dot{x}_0, \dot{y}_0), \end{cases} \tag{1.5}$$

where $t_0 > 0, \alpha > 0, \theta \geq 0$ and $(x_0, y_0), (\dot{x}_0, \dot{y}_0) \in \mathcal{X} \times \mathcal{Y}$.

By unfolding the expressions of the partial gradients of $L(\cdot, \cdot)$ in the dynamical system (1.5), we have the following reformulation of system (1.5):

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x) + A^*\left(y + \theta t\dot{y}(t)\right) = 0, \\ \ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t) - A\left(x(t) + \theta t\dot{x}(t)\right) + \nabla g(y) = 0, \\ (x(t_0), y(t_0)) = (x_0, y_0) \text{ and } (\dot{x}(t_0), \dot{y}(t_0)) = (\dot{x}_0, \dot{y}_0). \end{cases} \tag{1.6}$$

Here we consider problem (1.1) with the assumption that f and g are two convex and continuously differentiable functions with Lipschitz continuous gradients. We establish its second-order dynamical system (1.5) with asymptotically vanishing damping term and design a numerical algorithm based on the discretization of the system. Our main contributions are as follows:

(a) The continuous case. We show convergence rates $O\left(1/t^2\right)$ and $O\left(1/t^{\frac{2\alpha}{3}}\right)$ for the primal-dual gap along the generated trajectories when $\alpha \geq 3$ and $0 < \alpha \leq 3$, respectively. Our dynamical system (1.5) is an extension of the primal-dual dynamical systems with vanishing damping for linearly constrained minimization problems proposed by Boţ and Nguyen [14], where we replace the linear function with a general convex and continuously differentiable function g. We also extend the convergence results of the unconstrained convex optimization problem when $0 < \alpha \leq 3$ given by Attouch et al. [5] to the case of the bilinearly coupled convex-concave saddle point problem. Then, we prove that the primal-dual trajectory asymptotically weakly converges to a primal-dual optimal solution of (1.1). By the assumption of strong convexity, we obtain the strong convergence rate for the primal-dual dynamical system, which extends the results for the unconstrained problem considered by Attouch [9] to the bilinearly coupled convex-concave saddle point problem considered here.

(b) The discrete case. We present the corresponding inertial algorithm based on the discretization of the primal-dual dynamical system with asymptotically vanishing damping term. We consider a general setting for the inertial parameters which covers three classical rules proposed by Nesterov [28], Chambolle-Dossal [18] and Attouch-Cabot [7]. We obtain the convergence rate of $O\left(1/k^2\right)$ for the primal-dual gap under these rules. Moreover, we prove that the sequence of the iterates constructed by this algorithm converges to a primal-dual solution in a general setting which covers the two latter rules, which is an extension of the results in Boţ et al.[15].

1.2 Overview

This paper is organized as follows. We focus on an analysis of the second-order dynamical system with asymptotically vanishing damping term in Section 2. To be specific, we show convergence rates for the primal-dual gap in Subsection 2.1 and the weak convergence of the trajectory to a primal-dual optimal solution in Subsection 2.2. In addition, we show faster convergence rates under the assumption of strong convexity in Subsection 2.3. In Section 3, we present the corresponding numerical algorithm and the convergence of the sequence of iterates that it generates. More precisely, We provide some important estimates in Subsection 3.1 and discuss the boundedness and convergence of the iterates of our algorithm in Subsection 3.2, before we summarize the results in Section 4.

2 Primal-dual dynamical systems: convergence rates

2.1 Fast convergence rates for the primal-dual gap

In this section, we investigate the convergence properties of the dynamical system (1.5), i. e. convergence rates for the primal dual gap, weak convergence of the trajectory to a primal-dual optimal solution and strong convergence rates under the assumption of strongly convexity. The following result, which can be directly proven by the Picard–Lindelof theorem (see [32], Theorem 2.2), establishes the existence and uniqueness of local solutions to the dynamical system (1.5).

Proposition 2.1. *Let f and g be two continuously differentiable convex functions and ∇f and ∇g are l_f- and l_g-Lipschitz continuous, respectively. Suppose $\alpha > 0$. Then for any initial condition $(x(t_0), y(t_0)) = (x_0, y_0)$,*

$(\dot{x}(t_0), \dot{y}(t_0)) = (\dot{x}_0, \dot{y}_0)$, *there exists a unique global twice continuously differentiable solution* $(x, y) : [t_0, T) \to \mathcal{X} \times \mathcal{Y}$.

To prove the existence of a global solution and derive the asymptotic behavior of the dynamical system (1.5), we notice that all the analyses are based on standard techniques of energy (Lyapunov) functions. Many energy functions have been proposed to study dynamcial systems with various damping terms and time scaling terms, see e. g. [30, 5, 29, 6], and choosing an appropriate one is crucial. Motivated by the one introduced in Attouch et al. [5] and Boţ and Nguyen [14], we define the following energy function $\mathcal{E}_{\alpha,\theta,p} : [t_0, +\infty) \to \mathbb{R}$ as

$$\mathcal{E}_{\alpha,\theta,p}(t) = \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{E}_2$$

with

$$\begin{aligned}
\mathcal{E}_0(t) &= \theta^2 t^{2p}(L(x, y^*) - L(x^*, y)),\\
\mathcal{E}_1(t) &= \frac{1}{2}\|\lambda(t)(x(t) - x^*) + \theta t^p \dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - x^*\|^2,\\
\mathcal{E}_2(t) &= \frac{1}{2}\|\lambda(t)(y(t) - y^*) + \theta t^p \dot{y}(t)\|^2 + \frac{\xi(t)}{2}\|y(t) - y^*\|^2,
\end{aligned}$$

where $\lambda(t) = t^{p-1}$ and $\xi(t) = (\theta\alpha - 2\theta p + \theta - 1)t^{2(p-1)}$.

Lemma 2.1. *Let* (x, y) *be a solution of the dynamical system* (1.5) *and* $(x^*, y^*) \in \mathbb{S}$ *. For every* $t \geq t_0$*, it holds that*

$$\begin{aligned}
\dot{\mathcal{E}}_{\alpha,\theta,p}(t) \quad \leq \quad & \theta(2\theta p - 1)t^{2p-1}(L(x, y^*) - L(x^*, y)) + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right)\\
& +(p - 1)\theta(\alpha - 2p + 1)t^{2p-3}(\|x - x^*\|^2 + \|y - y^*\|^2). \qquad (2.1)
\end{aligned}$$

Proof. We have

$$\begin{aligned}
\dot{\mathcal{E}}_0(t) \;&=\; 2p\theta^2 t^{2p-1}(L(x, y^*) - L(x^*, y)) + \theta^2 t^{2p}\left(\langle \nabla f(x), \dot{x}\rangle + \langle A\dot{x}, y^*\rangle - \langle Ax^*, \dot{y}\rangle + \langle \nabla g(y), \dot{y}\rangle\right),\\
\dot{\mathcal{E}}_1(t) \;&=\; \langle t^{p-1}(x - x^*) + \theta t^p \dot{x}, (p - 1)t^{p-2}(x - x^*) + t^{p-1}\dot{x} + \theta p t^{p-1}\dot{x} + \theta t^p \ddot{x}\rangle + \frac{\dot{\xi}}{2}\|x - x^*\|^2 + \xi\langle x - x^*, \dot{x}\rangle\\
&=\; \langle t^{p-1}(x - x^*) + \theta t^p \dot{x}, (p - 1)t^{p-2}(x - x^*) + (\theta(p - \alpha) + 1)t^{p-1}\dot{x} - \theta t^p \nabla f(x) - \theta t^p A^*\left(y + \theta t\dot{y}(t)\right)\rangle\\
&\qquad + \frac{\dot{\xi}}{2}\|x(t) - x^*\|^2 + \xi\langle x - x^*, \dot{x}\rangle\\
&=\; ((p - 1)t^{2p-3} + \frac{\dot{\xi}}{2})\|x - x^*\|^2 + (\xi - \theta\alpha + 2\theta p - \theta + 1)\langle x - x^*, \dot{x}\rangle - \theta t^{2p-1}\langle x - x^*, \nabla f(x)\rangle\\
&\qquad - \theta t^{2p-1}\langle x - x^*, A^* y\rangle - \theta^2 t^{2p}\langle x - x^*, A^*\dot{y}\rangle + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\|\dot{x}\|^2 - \theta^2 t^{2p}\langle \dot{x}, \nabla f(x)\rangle\\
&\qquad - \theta^2 t^{2p}\langle \dot{x}, A^*(y + \theta t\dot{y})\rangle\\
&=\; (p - 1)\theta(\alpha - 2p + 1)t^{2p-3}\|x - x^*\|^2 - \theta t^{2p-1}\langle x - x^*, \nabla f(x)\rangle - \theta t^{2p-1}\langle x - x^*, A^* y\rangle\\
&\qquad - \theta^2 t^{2p}\langle x - x^*, A^*\dot{y}\rangle + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\|\dot{x}\|^2 - \theta^2 t^{2p}\langle \dot{x}, \nabla f(x)\rangle - \theta^2 t^{2p}\langle \dot{x}, A^*(y + \theta t\dot{y})\rangle,\\
\dot{\mathcal{E}}_2(t) \;&=\; (p - 1)\theta(\alpha - 2p + 1)t^{2p-3}\|y - y^*\|^2 + \theta t^{2p-1}\langle y - y^*, \nabla g(y)\rangle - \theta t^{2p-1}\langle y - y^*, Ax\rangle\\
&\qquad - \theta^2 t^{2p}\langle y - y^*, A\dot{x}\rangle + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\|\dot{y}\|^2 - \theta^2 t^{2p}\langle \dot{y}, \nabla g(y)\rangle + \theta^2 t^{2p}\langle \dot{y}, A(x + \theta t\dot{x})\rangle.
\end{aligned}$$

Combining these terms, we arrive at

$$\begin{aligned}
\dot{\mathcal{E}}_{\alpha,\theta,p}(t) &= 2p\theta^2 t^{2p-1}(L(x,y^*)-L(x^*,y)) + (p-1)\theta(\alpha-2p+1)t^{2p-3}(\|x-x^*\|^2+\|y-y^*\|^2) \\
&\quad -\theta t^{2p-1}\left(\langle x-x^*, \nabla f(x)\rangle - \langle x^*, A^*y\rangle + \langle y-y^*, \nabla g(y)\rangle + \langle Ax, y^*\rangle\right) \\
&\quad +\theta(1+\theta p-\theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2+\|\dot{y}\|^2\right) \\
&\le 2p\theta^2 t^{2p-1}(L(x,y^*)-L(x^*,y)) + (p-1)\theta(\alpha-2p+1)t^{2p-3}(\|x-x^*\|^2+\|y-y^*\|^2) \\
&\quad +\theta t^{2p-1}\left(f(x^*)-f(x)+\langle x^*, A^*y\rangle + g(y^*)-g(y)-\langle Ax, y^*\rangle\right) \\
&\quad +\theta(1+\theta p-\theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2+\|\dot{y}\|^2\right) \\
&= \theta(2\theta p-1)t^{2p-1}(L(x,y^*)-L(x^*,y)) + \theta(1+\theta p-\theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2+\|\dot{y}\|^2\right) \\
&\quad +(p-1)\theta(\alpha-2p+1)t^{2p-3}(\|x-x^*\|^2+\|y-y^*\|^2), \qquad (2.2)
\end{aligned}$$

where the inequality follows from the convexity of f and g. This completes the proof. □

To guarantee the energy function $\mathcal{E}_{\alpha,\theta,p}$ is nonnegative and nonincreasing, we impose the following conditions:

$$\begin{aligned}
&(A1):\ \theta\alpha-2\theta p+\theta-1\ge 0, \\
&(A2):\ 2\theta p-1\le 0, \\
&(A3):\ 1+\theta p-\theta\alpha\le 0, \\
&(A4):\ (p-1)\theta(\alpha-2p+1)\le 0.
\end{aligned}$$

Then, we immediately have the following result.

Proposition 2.2. *Let f and g be two continuously differentiable convex functions and let ∇f and ∇g bet l_f- and l_g-Lipschitz continuous, respectively. Suppose $\alpha>0$ and assumptions $(A1)-(A4)$ hold. Let $(x^*,y^*)\in\mathbb{S}$ and (x,y) be a solution of the dynamical system (1.5) defined on $[t_0,T)$ for some initial value. Then, $\dot{\mathcal{E}}_{\alpha,\theta,p}\le 0$. In addtion, if $\theta\alpha-2\theta p+\theta-1>0$, we can use $T=+\infty$.*

Proof. By the assumptions $(A1)-(A4)$, it is obvious that $\dot{\mathcal{E}}_{\alpha,\theta,p}\le 0$ and so the energy function $\mathcal{E}(t)_{\alpha,\theta,p}$ is nonincreasing on $[t_0,T)$. Then, $\mathcal{E}_{\alpha,\theta,p}(t)\le\mathcal{E}_{\alpha,\theta,p}(t_0)$, $\forall\, t\in[t_0,T)$. This implies that the energy function $\mathcal{E}(t)_{\alpha,\theta,p}$ is bounded on $[t_0,T)$ and so with

$$\frac{1}{2}\|t^{p-1}(x(t)-x^*)+\theta t^p\dot{x}(t)\|^2+\frac{(\theta\alpha-2\theta p+\theta-1)t^{2(p-1)}}{2}\|x(t)-x^*\|^2\le\mathcal{E}_{\alpha,\theta,p}(t_0),\ \forall\, t\in[t_0,T),$$

it follows that

$$t^{p-1}\|x(t)-x^*\|\le\sqrt{\frac{2\mathcal{E}_{\alpha,\theta,p}(t_0)}{\theta\alpha-2\theta p+\theta-1}}\ \text{and}\ \|t^{p-1}(x(t)-x^*)+\theta t^p\dot{x}(t)\|\le\sqrt{2\mathcal{E}(t_0)_{\alpha,\theta,p}}$$

for $\forall\, t\in[t_0,T)$. Thus,

$$\begin{aligned}
\theta t^p\|\dot{x}(t)\| &\le \|t^{p-1}(x(t)-x^*)+\theta t^p\dot{x}(t)\|+t^{p-1}\|x(t)-x^*\| \\
&\le \sqrt{\frac{2\mathcal{E}_{\alpha,\theta,p}(t_0)}{\theta\alpha-2\theta p+\theta-1}}+\sqrt{2\mathcal{E}_{\alpha,\theta,p}(t_0)} \\
&\le \left(1+\frac{1}{\theta\alpha-2\theta p+\theta-1}\right)\sqrt{2\mathcal{E}_{\alpha,\theta,p}(t_0)},\forall\, t\in[t_0,T), \qquad (2.3)
\end{aligned}$$

which yields $\sup_{t\in[t_0,T)}\|\dot{x}(t)\|<+\infty$ with the fact $t^p\ge t_0^p>0$.

Similarly, we have $\sup_{t\in[t_0,T)}\|\dot{y}(t)\|<+\infty$. Assume that $T<+\infty$ and so the trajectory $(x(t),y(t))$ is bounded on $[t_0,T)$. Consequently, $(x(t),y(t),\dot{x}(t),\dot{y}(t))$ has a limit at $t=T$ and therefore can be continued, which is a contradiction. In conclusion, $T=+\infty$, which completes the proof. $\square$

We will now simplify the conditions $(A1)-(A4)$. From $(A1)$, we have $\theta(\alpha-2p+1)\geq 1$ and so $\theta>0$. In addition, condition $(A4)$ becomes $p\leq 1$. By transforming $(A2)$ into $-\frac{1}{\theta}\leq -2p$, we have $p\leq \alpha-\frac{1}{\theta}\leq \alpha-2p$ and $p\leq \frac{\alpha+1}{2}-\frac{1}{2\theta}\leq \frac{\alpha+1}{2}-p$, from which it follows that $p\leq\frac{\alpha}{3}$ and $p\leq\frac{\alpha+1}{4}$. Combining the results above, we have $p\leq\min\{1,\frac{\alpha}{3},\frac{\alpha+1}{4}\}$.

It is obvious that $\frac{\alpha}{3}\leq\frac{\alpha+1}{4}\leq 1$ when $\alpha\leq 3$ and $1\leq\frac{\alpha+1}{4}\leq\frac{\alpha}{3}$ when $\alpha\geq 3$. Therefore, the formula $p\leq\min\{1,\frac{\alpha}{3},\frac{\alpha+1}{4}\}$ can be simplified to $p\leq\min\{1,\frac{\alpha}{3}\}$. In other words, we can set $p=1$ when $\alpha\geq 3$ and $p=\frac{\alpha}{3}$ when $\alpha\leq 3$. Based on the energy functions $\mathcal{E}_{\alpha,\theta,1}(t)$ and $\mathcal{E}_{\alpha,\theta,\frac{\alpha}{3}}(t)$, we now have the following results.

Theorem 2.1. *Let f and g be two continuously differentiable convex functions, (x,y) be a solution of dynamical system (1.5) and $(x^*,y^*)\in\mathbb{S}$. The following statements are true:*

(1) When $\alpha\geq 3$ and $p=1$, conditions $(A1)-(A4)$ can be further simplified to $\frac{1}{\alpha-1}\leq\theta\leq\frac{1}{2}$. Then,

$$(1-2\theta)\int_{t_0}^{+\infty} t(L(x(t),y^*)-L(x^*,y(t)))dt\leq\frac{\mathcal{E}_{\alpha,\theta}(t_0)}{\theta}<+\infty, \tag{2.4}$$

$$(\theta\alpha-\theta-1)\int_{t_0}^{+\infty} t\left(\|\dot{x}(t)\|^2+\|\dot{y}(t)\|^2\right)dt\leq\frac{\mathcal{E}_{\alpha,\theta}(t_0)}{\theta}<+\infty. \tag{2.5}$$

Moreover, if $\frac{1}{\alpha-1}<\theta\leq\frac{1}{2}$ then $\|\dot{x}(t)\|=O(\frac{1}{t})$.

(2) When $\alpha\leq 3$ and $p\leq\frac{\alpha}{3}$, conditions $(A1)-(A4)$ can be further simplified to $\frac{1}{\alpha-p}\leq\theta\leq\frac{1}{2p}$, and we have

$$(1-2\theta p)\int_{t_0}^{+\infty} t^{2p-1}(L(x(t),y^*)-L(x^*,y(t)))dt\leq\frac{\mathcal{E}_{\alpha,\theta,p}(t_0)}{\theta}<+\infty, \tag{2.6}$$

$$(\theta\alpha-\theta p-1)\int_{t_0}^{+\infty} t^{2p-1}\left(\|\dot{x}(t)\|^2+\|\dot{y}(t)\|^2\right)dt\leq\frac{\mathcal{E}_{\alpha,\theta,p}(t_0)}{\theta}<+\infty. \tag{2.7}$$

Moreover, if $\frac{1}{\alpha-p}<\theta\leq\frac{1}{2p}$ then

$$\|\dot{x}(t)\|=O(\frac{1}{t^p}). \tag{2.8}$$

Proof. When $p=1$, by (2.2) and $\frac{1}{\alpha-1}\leq\theta\leq\frac{1}{2}$, we have

$$\dot{\mathcal{E}}_{\alpha,\theta,1}(t)=\theta(2\theta-1)t(L(x,y^*)-L(x^*,y))+\theta(1+\theta-\theta\alpha)t\left(\|\dot{x}\|^2+\|\dot{y}\|^2\right)\leq 0. \tag{2.9}$$

Then, the energy function $\mathcal{E}_{\alpha,\theta,1}(t)$ is nonincreasing on $[t_0,+\infty)$. For every $t\geq t_0$, it holds that

$$\begin{aligned}\mathcal{E}_{\alpha,\theta,1}(t) \quad &= \quad \theta^2t^2(L(x,y^*)-L(x^*,y))+\frac{1}{2}\|(x(t)-x^*)+\theta t\dot{x}(t)\|^2+\frac{\theta\alpha-\theta-1}{2}\|x(t)-x^*\|^2\\ &\quad\quad +\frac{1}{2}\|(y(t)-y^*)+\theta t\dot{y}(t)\|^2+\frac{\theta\alpha-\theta-1}{2}\|y(t)-y^*\|^2\\ &\leq \quad \mathcal{E}_{\alpha,\theta,1}(t_0).\end{aligned} \tag{2.10}$$

By integrating (2.9) from t_0 to t, we have

$$\theta(1-2\theta)\int_{t_0}^{t} s(L(x(s),y^*)-L(x^*,y(s)))ds+\theta(\theta\alpha-\theta-1)\int_{t_0}^{t} s\left(\|\dot{x}(s)\|^2+\|\dot{y}(s)\|^2\right)ds\leq\mathcal{E}(t_0)_{\alpha,\theta,1}.$$

All items inside the integrals are nonnegative. Thus, we arrive at (2.4) and (2.5) by passing $t \to +\infty$. By setting $p = 1$ in (2.3), we get

$$\|\dot{x}(t)\| \leq \frac{1}{\theta t}\left(1 + \frac{1}{\theta\alpha - \theta - 1}\right)\sqrt{2\mathcal{E}_{\alpha,\theta,p}(t_0)}, \forall\, t \in [t_0, +\infty),$$

which yields $\|\dot{x}(t)\| = O(\frac{1}{t})$. When $\alpha \leq 3$ and $p \leq \frac{\alpha}{3}$, we clearly have $p \leq 1$ and $\alpha - 2p + 1 \geq \alpha - p > 0$. Thus, condition $(A4)$ is satisfied. From condition $(A1 - A3)$, It is obvious that $\frac{1}{\alpha-2p+1} \leq \frac{1}{\alpha-p} \leq \theta \leq \frac{1}{2p}$. Since

$$\begin{aligned}\dot{\mathcal{E}}_{\alpha,\theta,p}(t) &= \theta(2\theta p - 1)t^{2p-1}(L(x, y^*) - L(x^*, y)) + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right)\\ &\quad + (p-1)\theta(\alpha - 2p + 1)t^{2p-3}(\|x - x^*\|^2 + \|y - y^*\|^2) \leq 0,\end{aligned}$$

similarly to the proof in the case $\alpha \geq 3$, we get (2.6) and (2.7) and (2.8). □

Remark 2.1. *If ∇f and ∇g are l_f- and l_g-Lipschitz continuous, respectively, formula* (2.2) *in the proof of Lemma 2.1 can be sharpened to*

$$\begin{aligned}\dot{\mathcal{E}}_{\alpha,\theta,p}(t) \;\leq\;& 2p\theta^2 t^{2p-1}(L(x, y^*) - L(x^*, y)) + (p-1)\theta(\alpha - 2p + 1)t^{2p-3}(\|x - x^*\|^2 + \|y - y^*\|^2)\\ & + \theta t^{2p-1}\Bigg(f(x^*) - f(x) + \langle x^*, A^*y\rangle + g(y^*) - g(y) - \langle Ax, y^*\rangle - \frac{1}{2l_f}\|\nabla f(x) - \nabla f(x^*)\|^2\\ & - \frac{1}{2l_g}\|\nabla g(y) - \nabla g(y^*)\|^2\Bigg) + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right)\\ =\;& \theta(2\theta p - 1)t^{2p-1}(L(x, y^*) - L(x^*, y)) + (p-1)\theta(\alpha - 2p + 1)t^{2p-3}(\|x - x^*\|^2 + \|y - y^*\|^2)\\ & + \theta(1 + \theta p - \theta\alpha)t^{2p-1}\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right) - \frac{\theta t^{2p-1}}{2l_f}\|\nabla f(x) - \nabla f(x^*)\|^2 - \frac{\theta t^{2p-1}}{2l_g}\|\nabla g(y) - \nabla g(y^*)\|^2.\end{aligned}$$

Similar to the proof in Theorem 2.1, the following statements then hold:

$$\text{If } \alpha \geq 3, \int_{t_0}^{+\infty} t\|\nabla f(x) - \nabla f(x^*)\|^2 dt < +\infty, \int_{t_0}^{+\infty} t\|\nabla g(y) - \nabla g(y^*)\|^2 dt < +\infty; \tag{2.11}$$

$$\text{If } \alpha \leq 3 \text{ and } p \leq \tfrac{\alpha}{3}, \int_{t_0}^{+\infty} t^{2p-1}\|\nabla f(x) - \nabla f(x^*)\|^2 dt < +\infty, \int_{t_0}^{+\infty} t^{2p-1}\|\nabla g(y) - \nabla g(y^*)\|^2 dt < +\infty.$$

Remark 2.2. *When $a \geq 3$, from* (2.10) *we obtain a convergence rate for the primal-dual gap of*

$$L(x, y^*) - L(x^*, y) = O\left(1/t^2\right).$$

When $\alpha \leq 3$, we set $p = \frac{\alpha}{3}$ to obtain the best decay rate. From $\frac{1}{\alpha-p} \leq \theta \leq \frac{1}{2p}$, it follows that $\theta = \frac{3}{2\alpha}$. Similarly, we then obtain a convergence rate for the primal-dual gap of

$$L(x, y^*) - L(x^*, y) = O\left(1/t^{\frac{2\alpha}{3}}\right).$$

With this we extend (Attouch et al., [5], Theorem 2.1 and Theorem 2.4) from the unconstrained convex optimization case with $0 < \alpha \leq 3$ to the case of the bilinearly coupled convex-concave saddle point problem. In addition, it is easy to verify that (2.6) *and* (2.7) *still hold, but* (2.8) *fails in this case.*

2.2 Weak convergence of the trajectory to a primal-dual optimal solution

In this subsection, we assume that f and g be two continuously differentiable functions and ∇f and ∇g are l_f- and l_g-Lipschitz continuous, respectively. In addition, we suppose $\alpha > 3$, and $\frac{1}{\alpha-1} < \theta < \frac{1}{2}$. To discuss weak convergence, we define the following two mappings on $[t_0, +\infty)$:

$$W(t) := L(x, y^*) - L(x^*, y) + \frac{1}{2}\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right),$$
$$\varphi(t) := \frac{1}{2}\left(\|x(t) - x^*\|^2 + \|y(t) - y^*\|^2\right).$$

From inequalites (2.4) and (2.5), it is obvious that $tW(t)$ belongs to the Lebesgue space $\mathbb{L}^1([t_0, +\infty))$.

Lemma 2.2. *Let (x, y) be a solution of the dynamical system* (1.5) *and $(x^*, y^*) \in \mathbb{S}$. Then, we have*

$$\ddot{\varphi}(t) + \frac{\alpha}{t}\dot{\varphi}(t) + \theta t\dot{W}(t) + \frac{1}{2l_f}\|\nabla f(x) - \nabla f(x^*)\|^2 + \frac{1}{2l_g}\|\nabla g(y) - \nabla g(y^*)\|^2 \leq 0. \tag{2.12}$$

Proof. For any fixed $t \geq t_0$, by differentiating W with respect to t, we arrive at

$$\dot{W}(t) = (\langle \nabla f(x), \dot{x}\rangle + \langle A\dot{x}, y^*\rangle - \langle Ax^*, \dot{y}\rangle + \langle \nabla g(y), \dot{y}\rangle) + \langle \ddot{x}, \dot{x}\rangle + \langle \ddot{y}, \dot{y}\rangle, \tag{2.13}$$

Substituting the expressions for $\ddot{x}$ and $\ddot{y}$ in the dynamical system (1.5) into (2.13), we also have

$$\dot{W}(t) = \langle A\dot{x}, y^* - y\rangle + \langle A(x - x^*), \dot{y}\rangle - \frac{\alpha}{t}\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right). \tag{2.14}$$

This leads to

$$\begin{aligned}
& \ddot{\varphi}(t) + \frac{\alpha}{t}\dot{\varphi}(t) \\
= \;& \langle x - x^*, \ddot{x} + \frac{\alpha}{t}\dot{x}\rangle + \langle y - y^*, \ddot{y} + \frac{\alpha}{t}\dot{y}\rangle + \|\dot{x}\|^2 + \|\dot{y}\|^2 \\
= \;& \langle x - x^*, -\nabla f(x) - A^*(y + \theta t\dot{y})\rangle + \langle y - y^*, A(x + \theta t\dot{x}) - \nabla g(y)\rangle + \|\dot{x}\|^2 + \|\dot{y}\|^2 \\
= \;& -\langle x - x^*, \nabla f(x)\rangle + \langle Ax^*, y\rangle - \langle y - y^*, \nabla g(y)\rangle - \langle y^*, Ax\rangle \\
& + \theta t\langle y - y^*, A\dot{x}\rangle - \theta t\langle A(x - x^*), \dot{y}\rangle + \|\dot{x}\|^2 + \|\dot{y}\|^2.
\end{aligned}$$

By the Lipschitz continuity of $\nabla f(x)$ and $\nabla g(y)$, we obtain

$$\begin{aligned}
& \ddot{\varphi}(t) + \frac{\alpha}{t}\dot{\varphi}(t) \\
\leq \;& -\left(f(x) - f(x^*) - \langle Ax^*, y\rangle + g(y) - g(y^*) + \langle y^*, Ax\rangle\right) - \frac{1}{2l_f}\|\nabla f(x) - \nabla f(x^*)\|^2 \\
& - \frac{1}{2l_g}\|\nabla g(y) - \nabla g(y^*)\|^2 + \theta t\langle y - y^*, A\dot{x}\rangle - \theta t\langle A(x - x^*), \dot{y}\rangle + \|\dot{x}\|^2 + \|\dot{y}\|^2 \\
= \;& -\left(L(x, y^*) - L(x^*, y)\right) - \frac{1}{2l_f}\|\nabla f(x) - \nabla f(x^*)\|^2 \\
& - \frac{1}{2l_g}\|\nabla g(y) - \nabla g(y^*)\|^2 + \theta t\langle y - y^*, A\dot{x}\rangle - \theta t\langle A(x - x^*), \dot{y}\rangle + \|\dot{x}\|^2 + \|\dot{y}\|^2.
\end{aligned} \tag{2.15}$$

Then, adding $\theta t\dot{W}(t)$ to (2.15) yields

$$\begin{aligned}
& \ddot{\varphi}(t) + \frac{\alpha}{t}\dot{\varphi}(t) + \theta t\dot{W}(t) \\
\leq \;& -\left(L(x, y^*) - L(x^*, y)\right) - \frac{1}{2l_f}\|\nabla f(x) - \nabla f(x^*)\|^2 - \frac{1}{2l_g}\|\nabla g(y) - \nabla g(y^*)\|^2 + (1 - \theta\alpha)\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right).
\end{aligned}$$

With the fact that $1 - \theta\alpha < -\theta$, we obtain (2.12) which completes the proof. $\square$

The following lemma provides one of the two conditions of the Opial Lemma (see Lemma A.3 in [14]) which is a classical tool to prove weak convergence of the trajectory.

Lemma 2.3. *Let (x, y) be a solution of the dynamical system* (1.5) *and $(x^*, y^*) \in \mathbb{S}$. Then, the positive part $[\dot{\varphi}]_+$ of $\dot{\varphi}$ belongs to $\mathbb{L}^1([t_0, +\infty))$ and $\lim_{t\to+\infty} \varphi(t)$ exists.*

Proof. As this proof is similar to the one in Lemma 4.4 in [14], we omit it here. □

Lemma 2.4. *Let (x, y) be a solution of the dynamical system* (1.5) *and $(x^*, y^*) \in \mathbb{S}$. Then, for all $t \geq t_0$, the following statement holds:*

$$\begin{aligned}
& (1-\theta)\left(\|A(x-x^*)\|^2 + \|A^*(y-y^*)\|^2\right) + \theta\frac{d}{dt}\left(t\|A(x-x^*)\|^2 + t\|A^*(y-y^*)\|^2\right) \\
& +\frac{\alpha}{t}\frac{d}{dt}\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right) + 2\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(y-y^*)\rangle - 2\langle \ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t), A(x-x^*)\rangle \\
\leq\ & \|\nabla f(x) - \nabla f(x^*)\|^2 + \|\nabla g(y) - \nabla g(y^*)\|^2.
\end{aligned} \tag{2.16}$$

Proof. From (1.5), we have

$$\begin{aligned}
& \|\nabla f(x) + A^* y^*\|^2 + \|\nabla g(y) - Ax^*\|^2 \\
=\ & \|\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A^*(y - y^* + \theta t\dot{y}(t))\|^2 + \|\ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t) - A^*(x - x^* + \theta t\dot{x}(t))\|^2 \\
=\ & \|\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t)\|^2 + \|\ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t)\|^2 + \|A^*(y - y^* + \theta t\dot{y}(t))\|^2 + \|A^*(x - x^* + \theta t\dot{x}(t))\|^2 \\
& +2\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(y-y^*)\rangle - 2\langle \ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t), A(x-x^*)\rangle \\
& +2\theta t\langle \ddot{x}(t), A^*\dot{y}\rangle - 2\theta t\langle \ddot{y}(t), A\dot{x}\rangle.
\end{aligned} \tag{2.17}$$

Accordingly, we arrive at

$$2\theta t\langle \ddot{x}(t), A^*\dot{y}\rangle - 2\theta t\langle \ddot{y}(t), A\dot{x}\rangle \geq -\|\ddot{x}(t)\|^2 - \|\ddot{y}(t)\|^2 - \theta^2 t^2\|A^*\dot{y}\|^2 - \theta^2 t^2\|A\dot{x}\|^2, \tag{2.18}$$

$$\|\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t)\|^2 + \|\ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t)\|^2 - \|\ddot{x}(t)\|^2 - \|\ddot{y}(t)\|^2 \geq \frac{\alpha}{t}\frac{d}{dt}\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right), \tag{2.19}$$

and

$$\begin{aligned}
& \|A^*(y - y^* + \theta t\dot{y}(t))\|^2 + \|A^*(x - x^* + \theta t\dot{x}(t))\|^2 - \theta^2 t^2\|A^*\dot{y}\|^2 - \theta^2 t^2\|A\dot{x}\|^2 \\
=\ & \|A(x-x^*)\|^2 + \|A^*(y-y^*)\|^2 + 2\theta t\langle A(x-x^*), A\dot{x}\rangle + 2\theta t\langle A^*(y-y^*), A^*\dot{y}\rangle \\
=\ & (1-\theta)\left(\|A(x-x^*)\|^2 + \|A^*(y-y^*)\|^2\right) + \theta\frac{d}{dt}\left(t\|A(x-x^*)\|^2 + t\|A^*(y-y^*)\|^2\right).
\end{aligned} \tag{2.20}$$

Then, combining (2.17) with (2.18), (2.19) and (2.20), we have

$$\begin{aligned}
& \|\nabla f(x) + A^* y^*\|^2 + \|\nabla g(y) - Ax^*\|^2 \\
\geq\ & (1-\theta)\left(\|A(x-x^*)\|^2 + \|A^*(y-y^*)\|^2\right) + \theta\frac{d}{dt}\left(t\|A(x-x^*)\|^2 + t\|A^*(y-y^*)\|^2\right) \\
& +\frac{\alpha}{t}\frac{d}{dt}\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right) + 2\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(y-y^*)\rangle - 2\langle \ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t), A(x-x^*)\rangle.
\end{aligned}$$

We notice that $A^*y^* = -\nabla f(x^*)$ and $Ax^* = \nabla g(y^*)$, which completes the proof. □

The following theorem provides a further important integrability result.

Theorem 2.2. *Let (x, y) be a solution of the dynamical system* (1.5) *and $(x^*, y^*) \in \mathbb{S}$. Then,*

$$\int_{t_0}^{+\infty} t\|A(x(t)-x^*)\|^2 dt < +\infty \quad \textit{and} \quad \int_{t_0}^{+\infty} t\|A^*(y(t)-y^*)\|^2 dt < +\infty. \tag{2.21}$$

Moreover, $\|A(x-x^*)\| = o\left(\frac{1}{\sqrt{t}}\right)$ and $\|A(y-y^*)\| = o\left(\frac{1}{\sqrt{t}}\right)$ for $t \to +\infty$.

Proof. Summing (2.12) and (2.16), for every $t \geq t_0$, we have

$$\begin{aligned}
&\ddot{\varphi}(t) + \frac{\alpha}{t}\dot{\varphi}(t) + \theta t\dot{W}(t) + \theta\frac{d}{dt}\left(t\|A\left(x-x^*\right)\|^2 + t\|A^*\left(y-y^*\right)\|^2\right) \qquad (2.22)\\
&+\frac{\alpha}{t}\frac{d}{dt}\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right) + 2\langle\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(y-y^*)\rangle - 2\langle\ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t), A(x-x^*)\rangle\\
\leq\ & C_{31}\|\nabla f(x) - \nabla f(x^*)\|^2 + C_{32}\|\nabla g(y) - \nabla g(y^*)\|^2 + (\theta-1)\left(\|A\left(x-x^*\right)\|^2 + \|A^*\left(y-y^*\right)\|^2\right),
\end{aligned}$$

where $C_{31} := \left[1 - \frac{1}{2l_f}\right]_+ \geq 0$ and $C_{32} := \left[1 - \frac{1}{2l_g}\right]_+ \geq 0$.

Multiplying (2.22) by t^α and integrating, for every $t \geq t_0$, we get

$$\begin{aligned}
I_1(t) + I_2(t) + I_3(t) + I_4(t) + I_5(t) \leq\ & C_{31}\int_{t_0}^t s^\alpha\|\nabla f(x(s)) - \nabla f(x^*)\|^2 ds + C_{32}\int_{t_0}^t s^\alpha\|\nabla g(y(s)) - \nabla g(y^*)\|^2 ds\\
&+(\theta-1)\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds, \qquad (2.23)
\end{aligned}$$

where

$$\begin{aligned}
I_1(t) &:= \int_{t_0}^t \left(s^\alpha\ddot{\varphi}(s) + \alpha s^{\alpha-1}\dot{\varphi}(s)\right)ds,\\
I_2(t) &:= \int_{t_0}^t \theta s^{\alpha+1}\dot{W}(s)ds,\\
I_3(t) &:= \int_{t_0}^t \theta s^\alpha\frac{d}{ds}\left(s\|A\left(x(s)-x^*\right)\|^2 + s\|A^*\left(y(s)-y^*\right)\|^2\right)ds,\\
I_4(t) &:= \int_{t_0}^t \alpha s^{\alpha-1}\frac{d}{ds}\left(\|\dot{x}(s)\|^2 + \|\dot{y}(s)\|^2\right)ds,\\
I_5(t) &:= \int_{t_0}^t 2s^\alpha\left(\langle\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), A^*(y(s)-y^*)\rangle - 2\langle\ddot{y}(t) + \frac{\alpha}{t}\dot{y}(t), A(x(s)-x^*)\rangle\right)ds.
\end{aligned}$$

Next, we will focus on the integrals $I_i(t)$ $(i = 1, 2, \ldots, 5)$ separately. First, $I_1(t) = \int_{t_0}^t \frac{d}{ds}\left(s^\alpha\dot{\varphi}(s)\right)ds = t^\alpha\dot{\varphi}(t) - t_0^\alpha\dot{\varphi}(t_0)$ yields

$$0 = I_1(t) + t_0^\alpha\dot{\varphi}(t_0) - t^\alpha\dot{\varphi}(t). \qquad (2.24)$$

Using integration by parts, we obtain

$$I_2(t) = \theta\int_{t_0}^t s^{\alpha+1}dW(s) = \theta t^{\alpha+1}W(t) - \theta t_0^{\alpha+1}W(t_0) - \theta(\alpha+1)\int_{t_0}^t s^\alpha W(s)ds$$

and so

$$0 \leq \theta t^{\alpha+1}W(t) \leq I_2(t) + \theta t_0^{\alpha+1}W(t_0) + \theta(\alpha+1)\int_{t_0}^t s^\alpha W(s)ds. \qquad (2.25)$$

By integration by parts again, we get

$$\begin{aligned}
I_3(t) =\ & \theta t^{\alpha+1}\left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) - \theta t_0^{\alpha+1}\left(\|A\left(x_0-x^*\right)\|^2 + \|A^*\left(y_0-y^*\right)\|^2\right)\\
&-\theta\alpha\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds,
\end{aligned}$$

by which it follows that

$$\begin{aligned}
&\theta t^{\alpha+1}\left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right)\\
=\ & I_3(t) + \theta t_0^{\alpha+1}\left(\|A\left(x_0-x^*\right)\|^2 + \|A^*\left(y_0-y^*\right)\|^2\right) + \theta\alpha\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds.
\end{aligned}$$

Similarly,

$$I_4(t) = \alpha t^{\alpha-1}\left(\|\dot{x}(t)\|^2+\|\dot{y}(t)\|^2\right) - \alpha t_0^{\alpha-1}\left(\|\dot{x}_0\|^2+\|\dot{y}_0)\|^2\right) - \alpha(\alpha-1)\int_{t_0}^t s^{\alpha-2}\left(\|\dot{x}(s)\|^2+\|\dot{y}(s)\|^2\right)ds.$$

which yields

$$0 \le I_4(t) + \alpha t_0^{\alpha-1}\left(\|\dot{x}(t_0)\|^2+\|\dot{y}_0)\|^2\right) + \frac{\alpha(\alpha-1)}{t_0^2}\int_{t_0}^t s^{\alpha}\left(\|\dot{x}(s)\|^2+\|\dot{y}(s)\|^2\right)ds. \tag{2.26}$$

Then, again by integration by parts, we have

$$\begin{aligned} I_5(t) &= 2\int_{t_0}^t \left(\left\langle \frac{d}{ds}\left(s^\alpha \dot{x}(s)\right), A^*(y(s)-y^*)\right\rangle - \left\langle \frac{d}{ds}\left(s^\alpha \dot{y}(s)\right), K(x(s)-x^*)\right\rangle\right)ds \\ &= 2t^\alpha\langle \dot{x}(t), A^*(y(t)-y^*)\rangle - 2t^\alpha\langle \dot{y}(t), A(x(t)-x^*)\rangle + 2t_0^\alpha\langle \dot{x}_0, A^*(y_0-y^*)\rangle - 2t_0^\alpha\langle \dot{y}_0, A(x_0-x^*)\rangle \end{aligned}$$

and so

$$0 = I_5(t) - 2t^\alpha\langle \dot{x}, A^*(y(t)-y^*)\rangle + 2t^\alpha\langle \dot{y}, A(x(t)-x^*)\rangle - 2t_0^\alpha\langle \dot{x}_0, A^*(y_0-y^*)\rangle + 2t_0^\alpha\langle \dot{y}_0, A(x_0-x^*)\rangle. \tag{2.27}$$

Summing (2.24), (2.25), (2.26), (2.26) and (2.27), we arrive at

$$\begin{aligned} & \theta t^{\alpha+1}\left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) \\ \le\; & I_1(t)+I_2(t)+I_3(t)+I_4(t)+I_5(t) - t^\alpha\dot{\varphi}(t) + \theta(\alpha+1)\int_{t_0}^t s^\alpha W(s)ds \\ & +\theta\alpha\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds + \frac{\alpha(\alpha-1)}{t_0^2}\int_{t_0}^t s^\alpha\left(\|\dot{x}(s)\|^2+\|\dot{y}(s)\|^2\right)ds \\ & -2t^\alpha\langle \dot{x}(t), A^*(y(t)-y^*)\rangle + 2t^\alpha\langle \dot{y}(t), A(x(t)-x^*)\rangle + C_{33}. \end{aligned}$$

Combining this with (2.23) yields

$$\begin{aligned} & \theta t^{\alpha+1}\left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) \\ \le\; & C_{31}\int_{t_0}^t s^\alpha\|\nabla f(x(s))-\nabla f(x^*)\|^2 ds + C_{32}\int_{t_0}^t s^\alpha\|\nabla g(y(s))-\nabla g(y^*)\|^2 ds \\ & +(\theta-1)\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds - t^\alpha\dot{\varphi}(t) + \theta(\alpha+1)\int_{t_0}^t s^\alpha W(s)ds \\ & +\theta\alpha\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right) + \frac{\alpha(\alpha-1)}{t_0^2}\int_{t_0}^t s^\alpha\left(\|\dot{x}(s)\|^2+\|\dot{y}(s)\|^2\right)ds \\ & -2t^\alpha\langle \dot{x}(t), A^*(y(t)-y^*)\rangle + 2t^\alpha\langle \dot{y}(t), A(x(t)-x^*)\rangle + C_{33} \\ \le\; & \int_{t_0}^t s^\alpha V(s)ds + (\theta\alpha+\theta-1)\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds - t^\alpha\dot{\varphi}(t) \\ & -2t^\alpha\langle \dot{x}(t), A^*(y(t)-y^*)\rangle + 2t^\alpha\langle \dot{y}(t), A(x(t)-x^*)\rangle + C_{33}, \end{aligned} \tag{2.28}$$

where

$$\begin{aligned} C_{33} &:= t_0^\alpha\dot{\varphi}(t_0) + \theta t_0^{\alpha+1}W(t_0) + \theta t_0^{\alpha+1}\left(\|K\left(x(t_0)-x^*\right)\|^2 + \|A^*\left(y(t_0)-y^*\right)\|^2\right) \\ &\quad +\alpha t_0^{\alpha-1}\left(\|\dot{x}(t_0)\|^2+\|\dot{y}_0)\|^2\right) - 2t_0^\alpha\langle \dot{x}_0, A^*(y_0-y^*)\rangle + 2t_0^\alpha\langle \dot{y}_0, A(x_0-x^*)\rangle; \\ V(s) &:= C_{31}\|\nabla f(x(s))-\nabla f(x^*)\|^2 + C_{32}\|\nabla g(y(s))-\nabla g(y^*)\|^2 + \theta(\alpha+1)W(s) + \frac{\alpha(\alpha-1)}{t_0^2}\left(\|\dot{x}(s)\|^2+\|\dot{y}(s)\|^2\right). \end{aligned}$$

Now, dividing (2.28) by t^α, we have

$$\begin{aligned} & \theta t\left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) \\ \le\; & \frac{1}{t^\alpha}\int_{t_0}^t s^\alpha V(s)ds + \frac{(\theta\alpha+\theta-1)}{t^\alpha}\int_{t_0}^t s^\alpha\left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right)ds - \dot{\varphi}(t) \\ & -2\langle \dot{x}(t), A^*(y(t)-y^*)\rangle + 2\langle \dot{y}(t), A(x(t)-x^*)\rangle + \frac{C_{33}}{t^\alpha}. \end{aligned} \tag{2.29}$$

By integrating (2.29) from t_0 to r for any $r \geq t_0$, we have

$$\begin{aligned}
&\theta \int_{t_0}^{r} t \left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) dt \\
\leq\ & \int_{t_0}^{r} \frac{1}{t^\alpha} \left(\int_{t_0}^{t} s^\alpha V(s) ds\right) dt + (\theta\alpha + \theta - 1) \int_{t_0}^{r} \frac{1}{t^\alpha} \left(\int_{t_0}^{t} s^\alpha \left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right) ds\right) dt \\
& -\varphi(r) + \varphi(t_0) - 2\int_{t_0}^{r} \left(\langle \dot{x}(t), A^*(y(t)-y^*)\rangle - \langle \dot{y}(t), A(x(t)-x^*)\rangle\right) dt + C_{33} \int_{t_0}^{r} \frac{1}{t^\alpha} dt. \qquad (2.30)
\end{aligned}$$

Since $V(t)$ and $\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2$ are continuous on $[t_0, +\infty)$, by Lemma A.1 in [14], we arrive at

$$\int_{t_0}^{r} \frac{1}{t^\alpha} \left(\int_{t_0}^{t} s^\alpha V(s) ds\right) dt \leq \frac{1}{\alpha - 1} \int_{t_0}^{r} tV(t) dt \qquad (2.31)$$

and

$$\begin{aligned}
&\int_{t_0}^{r} \frac{1}{t^\alpha} \left(\int_{t_0}^{t} s^\alpha \left(\|A\left(x(s)-x^*\right)\|^2 + \|A^*\left(y(s)-y^*\right)\|^2\right) ds\right) dt \\
\leq\ & \frac{1}{\alpha-1} \int_{t_0}^{r} t \left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) dt. \qquad (2.32)
\end{aligned}$$

By $\frac{1-2\theta}{\alpha-1} > 0$, we have

$$\begin{aligned}
& -2\int_{t_0}^{r} \left(\langle \dot{x}(t), A^*(y(t)-y^*)\rangle - \langle \dot{y}(t), A(x(t)-x^*)\rangle\right) dt \\
\leq\ & \int_{t_0}^{r} \left(\frac{(\alpha-1)t}{1-2\theta}\|\dot{x}(t)\|^2 + \frac{1-2\theta}{(\alpha-1)t}\|A^*(y(t)-y^*)\|^2 + \frac{(\alpha-1)t}{1-2\theta}\|\dot{y}(t)\|^2 + \frac{1-2\theta}{(\alpha-1)t}\|A(x(t)-x^*)\|^2\right) dt \\
=\ & \frac{(\alpha-1)}{1-2\theta} \int_{t_0}^{r} t\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right) dt + \frac{1-2\theta}{\alpha-1} \int_{t_0}^{r} \frac{1}{t} \left(\|A^*(y(t)-y^*)\|^2 + \|A(x(t)-x^*)\|^2\right) dt. \qquad (2.33)
\end{aligned}$$

In addition, it holds that

$$\int_{t_0}^{r} \frac{1}{t^\alpha} dt = \frac{t_0^{1-\alpha}}{\alpha-1} - \frac{r^{1-\alpha}}{\alpha-1} \leq \frac{t_0^{1-\alpha}}{\alpha-1}. \qquad (2.34)$$

Since $\varphi(t) \geq 0$, by (2.31), (2.32), (2.33), (2.34) and (2.30) we get

$$\begin{aligned}
&\frac{1-2\theta}{\alpha-1} \int_{t_0}^{r} \left(t - \frac{1}{t}\right) \left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) dt \\
\leq\ & \frac{1}{\alpha-1} \int_{t_0}^{r} tV(t) dt + \frac{(\alpha-1)}{1-2\theta} \int_{t_0}^{r} t\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right) dt + \varphi(t_0) + C_{33}\frac{t_0^{1-\alpha}}{\alpha-1}.
\end{aligned}$$

From (2.5), (2.11) and the fact $tW(t) \in \mathbb{L}^1([t_0, +\infty))$, we notice that $tV(t)$ and $t\left(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2\right)$ belong to $\mathbb{L}^1([t_0, +\infty))$. By passing $r \to +\infty$ in (2.35), we have the following statement

$$\int_{t_0}^{+\infty} \left(t - \frac{1}{t}\right) \left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) dt < +\infty.$$

Due to $\lim_{t\to+\infty} \frac{t-\frac{1}{t}}{t} = 1$ and using that $\|K\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2$ is nonnegative, we have

$$\int_{t_0}^{+\infty} t \left(\|A\left(x(t)-x^*\right)\|^2 + \|A^*\left(y(t)-y^*\right)\|^2\right) dt < +\infty,$$

which is nothing else than (2.21). In addition,

$$\frac{d}{dt}\left(t\|A\left(x(t)-x^*\right)\|^2\right) = \|A\left(x(t)-x^*\right)\|^2 + 2t\langle A^*A(x-x^*), \dot{x}\rangle \leq (1 + \|A\|^2 t)\|A\left(x(t)-x^*\right)\|^2 + t\|\dot{x}\|^2.$$

Since $t\|A\left(x(t)-x^*\right)\|^2, t\|\dot{x}\|^2 \in \mathbb{L}^1([t_0, +\infty))$, by Lemma A.2 in [14], we have $\|A(x-x^*)\| = o\left(\frac{1}{\sqrt{t}}\right)$. Similarly, we get $\|A(y-y^*)\| = o\left(\frac{1}{\sqrt{t}}\right)$. □

Theorem 2.3. *Let $(x(t), y(t))$ be a solution of the dynamical system* (1.5) *and $(x^*, y^*) \in \mathbb{S}$. Then, it holds that*

$$\|\nabla f(x) - \nabla f(x^*)\| = o\left(\frac{1}{\sqrt{t}}\right) \text{ and } \|\nabla g(y) - \nabla g(y^*)\| = o\left(\frac{1}{\sqrt{t}}\right) \tag{2.35}$$

for $t \to +\infty$. Consequently,

$$\|\nabla_x L(x, y)\| = o\left(1/\sqrt{t}\right) \text{ and } \|\nabla_y L(x, y)\| = o\left(1/\sqrt{t}\right).$$

Proof. We note that

$$\begin{aligned} \frac{d}{dt}\left(t\|\nabla f(x) - \nabla f(x^*)\|^2\right) &= \|\nabla f(x) - \nabla f(x^*)\|^2 + 2t\langle \nabla f(x) - \nabla f(x^*), \frac{d}{dt}\nabla f(x)\rangle \\ &\leq (1+t)\|\nabla f(x) - \nabla f(x^*)\|^2 + l_f^2 t\|\dot{x}\|^2, \end{aligned}$$

where the inequality follows from the fact that ∇f is l_f-Lipschitz continuous. From (2.11) we have $t\|\nabla f(x) - \nabla f(x^*)\|^2 \in \mathbb{L}^1([t_0, +\infty))$. By Lemma A.2 in [14] again, we obtain $\|\nabla f(x) - \nabla f(x^*)\| = o\left(\frac{1}{\sqrt{t}}\right)$. Furthermore, since

$$t\|\nabla_x L(x, y)\| = t\|\nabla f(x) - \nabla f(x^*) - A^*y^* + A^*y\| \leq \|t\nabla f(x) - \nabla f(x^*)\| + t\|A^*y^* - A^*y\|,$$

it follows that $\|\nabla_x L(x, y)\| = o\left(1/\sqrt{t}\right)$. In a similar fashion it follows that $\|\nabla_y L(x, y)\| = o\left(1/\sqrt{t}\right)$. This completes the proof. □

Theorem 2.4. *Let $(x(t), y(t))$ be a solution of the dynamical system* (1.5) *and $(x^*, y^*) \in \mathbb{S}$. Then $(x(t), y(t))$ converges weakly to a primal-dual optimal solution of* (1.1) *as $t \to +\infty$.*

Proof. Suppose $(\bar{x}, \bar{y})$ is an arbitrary weak sequential cluster point of $(x(t), y(t))$ as $t \to +\infty$, and so there exists a sequence $(x(t_n), y(t_n))$ such that $(x(t_n), y(t_n)) \to (\bar{x}, \bar{y})$ as $n \to +\infty$. By Theorem 2.3, we get

$$\nabla f(x(t_n)) - A^*y(t_n) \to \nabla f(x^*) - A^*y^* = 0 \text{ and } \nabla g(y(t_n)) - Ax(t_n) \to \nabla g(y^*) - Ax^* = 0, \quad \text{as } n \to +\infty,$$

respectively. On the one hand, since the graph of the operator $\mathcal{T}_L$ in (1.3) is sequentially closed (Proposition 20.38, [10]), we conclude that

$$\nabla f(\bar{x}) - A^*\bar{y} \to \nabla f(x^*) - A^*y^* = 0 \text{ and } \nabla g(\bar{y}) - A\bar{x} \to \nabla g(y^*) - Ax^* = 0, \text{as } n \to +\infty,$$

which means that $(\bar{x}, \bar{y}) \in \mathbb{S}$. On the other hand, from Lemma 2.3 we notice that $\lim_{t\to+\infty}(\|x - x^*\| + \|y - y^*\|)$ exists for every $(x^*, y^*) \in \mathbb{S}$. With this, we complete the proof via Opial's Lemma as given in [14]. □

2.3 Strong convergence results

In this subsection we assume that f (respectively, g) is a μ_f-strongly (respectively, μ_g-strongly) convex continuously differentiable function. Let $\mu = \min\{\mu_f, \mu_g\}$. We show that the convergence rate of the dynamical system (1.5) with strongly convex functions rapidly increase while α increases. The following theorem extends the corresponding result for the unconstrained problem from Attouch et al. [9] to the bilinearly coupled convex-concave saddle point problem.

Theorem 2.5. *Let f and g be two continuously differentiable and strongly convex functions, and (x, y) be a solution of the dynamical system (1.5) with $\alpha > 3$. Then, for $t \to +\infty$, $(x(t), y(t))$ converges strongly to the unique element $(x^*, y^*) \in \mathbb{S}$. Moreover,*

$$L(x(t), y^*) - L(x^*, y(t)) = O(\frac{1}{t^{\frac{2}{3}\alpha}}), \tag{2.36}$$

$$\|x(t) - x^*\|^2 = O(\frac{1}{t^{\frac{2}{3}\alpha}}),\ \|y(t) - y^*\|^2 = O(\frac{1}{t^{\frac{2}{3}\alpha}}), \tag{2.37}$$

$$\|\dot{x}(t)\|^2 = O(\frac{1}{t^{\frac{2}{3}\alpha}}),\ \|\dot{y}(t)\|^2 = O(\frac{1}{t^{\frac{2}{3}\alpha}}). \tag{2.38}$$

Proof. First, we define the following energy function:

$$\mathcal{E}(t) = t^p \left(\theta^2 t^2 (L(x, y^*) - L(x^*, y)) + \frac{1}{2}\|(x(t) - x^*) + \theta t \dot{x}(t)\|^2 + \frac{1}{2}\|(y(t) - y^*) + \theta t \dot{y}(t)\|^2\right). \tag{2.39}$$

By a discussion similar to the one in the proof of in Lemma (2.1), we have

$$\begin{aligned}
\dot{\mathcal{E}}(t) \quad &= \quad \theta^2(p+2)t^{p+1}(L(x, y^*) - L(x^*, y)) + \frac{p}{2}t^{p-1}(\|x - x^*\|^2 + \|y - y^*\|^2) \\
&\quad\quad (1 + \theta + \theta p - \theta\alpha)\, t^p \left(\langle x - x^*, \dot{x}\rangle + \langle y - y^*, \dot{y}\rangle\right) \\
&\quad\quad +\theta\left(1 + \theta - \theta\alpha + \frac{1}{2}\theta p\right) t^{p+1}(\|\dot{x}\|^2 + \|\dot{y}\|^2) \\
&\quad\quad -\theta t^{p+1}\left(\langle x - x^*, \nabla f(x)\rangle - \langle x^*, A^* y\rangle + \langle y - y^*, \nabla g(y)\rangle + \langle Ax, y^*\rangle\right) \\
&\leq \quad \theta^2(p+2)t^{p+1}(L(x, y^*) - L(x^*, y)) + \frac{p}{2}t^{p-1}(\|x - x^*\|^2 + \|y - y^*\|^2) \\
&\quad\quad (1 + \theta + \theta p - \theta\alpha)\, t^p \left(\langle x - x^*, \dot{x}\rangle + \langle y - y^*, \dot{y}\rangle\right) \\
&\quad\quad +\theta\left(1 + \theta - \theta\alpha + \frac{1}{2}\theta p\right) t^{p+1}(\|\dot{x}\|^2 + \|\dot{y}\|^2) \\
&\quad\quad -\theta t^{p+1}\left(f(x) - f(x^*) + \frac{\mu_f}{2}\|x - x^*\|^2 - \langle x^*, A^* y\rangle + g(y) - g(y^*) + \frac{\mu_g}{2}\|y - y^*\|^2 + \langle Ax, y^*\rangle\right) \\
&\leq \quad \theta\left(\theta(p+2) - 1\right) t^{p+1}(L(x, y^*) - L(x^*, y)) - \frac{1}{2}\left(\mu\theta t^2 - p\right) t^{p-1}(\|x - x^*\|^2 + \|y - y^*\|^2) \\
&\quad\quad (1 + \theta + \theta p - \theta\alpha)\, t^p \left(\langle x - x^*, \dot{x}\rangle + \langle y - y^*, \dot{y}\rangle\right) + \theta\left(1 + \theta - \theta\alpha + \frac{1}{2}\theta p\right) t^{p+1}(\|\dot{x}\|^2 + \|\dot{y}\|^2).
\end{aligned}$$

The first inequality follows f and g being strongly convex, while the second one follows from the expansion of $L(x, y^*) - L(x^*, y)$. To deduce the best decay rate of the dynamical system (1.5), we now fix $p = \frac{2}{3}(\alpha - 3)$ and $\theta = \frac{3}{2\alpha}$, from which it follows that $\theta(p+2) - 1 = 1 + \theta - \theta\alpha + \frac{1}{2}\theta p = 0$ and $1 + \theta + \theta p - \theta\alpha = \frac{p}{2}$. With this, we arrive at

$$\dot{\mathcal{E}}(t) \leq -\frac{1}{2}\left(\mu\theta t^2 - p\right) t^{p-1}(\|x - x^*\|^2 + \|y - y^*\|^2) + \frac{\theta p}{2} t^p \left(\langle x - x^*, \dot{x}\rangle + \langle y - y^*, \dot{y}\rangle\right).$$

Let $t_1 := \max\left\{t_0, \sqrt{\frac{p}{\theta\mu}}\right\} = \max\left\{t_0, \frac{2}{3}\sqrt{\frac{\alpha(\alpha-3)}{\mu}}\right\}$. For $t \geq t_1$, we have

$$\dot{\mathcal{E}}(t) \leq \frac{\theta p}{2} t^p \left(\langle x - x^*, \dot{x}\rangle + \langle y - y^*, \dot{y}\rangle\right).$$

Using integration by parts we see that

$$\mathcal{E}(t) \leq \mathcal{E}(t_1) + \frac{\theta p}{4}\left(t^p(\|x - x^*\|^2 + \|y - y^*\|^2) - p\int_{t_1}^{t} s^{p-1}(\|x(s) - x^*\|^2 + \|y(s) - y^*\|^2)ds\right),$$

by which follows that

$$\mathcal{E}(t) \leq \mathcal{E}(t_1) + \frac{\theta p}{4} t^p \left(\|x - x^*\|^2 + \|y - y^*\|^2\right).$$

Since f is μ_f-strongly convex and $\langle x, A^*y^*\rangle$ is linear in x, so $f(\cdot)+\langle x, A^*y^*\rangle$ is μ_f-strongly convex. By the fact that x^* is the minimum of $f(\cdot)+\langle x, A^*y^*\rangle - g(y^*)$ and Theorem 5.25 [13], we have

$$f(x)+\langle x, A^*y^*\rangle - f(x^*) - \langle x^*, A^*y^*\rangle \geq \frac{\mu_f}{2}\|x-x^*\|^2 \tag{2.40}$$

for all $x \in \mathcal{X}$. Similarly,

$$g(y)-\langle x^*, A^*y\rangle - g(y^*) + \langle x^*, A^*y^*\rangle \geq \frac{\mu_g}{2}\|y-y^*\|^2 \tag{2.41}$$

for all $y \in \mathcal{Y}$. Summing (2.40) and (2.41), we obtain

$$L(x,y^*)-L(x^*,y) \geq \frac{\mu_f}{2}\|x-x^*\|^2 + \frac{\mu_g}{2}\|y-y^*\|^2 \geq \frac{\mu}{2}(\|x-x^*\|^2+\|y-y^*\|^2) \tag{2.42}$$

and so

$$\mathcal{E}(t) \leq \mathcal{E}(t_1) + \frac{\theta p}{2\mu}t^p\left(L(x,y^*)-L(x^*,y)\right). \tag{2.43}$$

By the definition of $\mathcal{E}(t)$, we have

$$t^{p+2}\left(L(x,y^*)-L(x^*,y)\right) \leq \mathcal{E}(t) \leq \mathcal{E}(t_1) + \frac{\theta p}{2\mu}t^p\left(L(x,y^*)-L(x^*,y)\right). \tag{2.44}$$

Dividing (2.44) by t^{p+2} for all $t \geq t_1$, we have

$$\begin{aligned} L(x,y^*)-L(x^*,y) &\leq \mathcal{E}(t_1)t^{-p-2} + \frac{\theta p}{2\mu}t^{-2}\left(L(x,y^*)-L(x^*,y)\right) \\ &\leq \mathcal{E}(t_1)t^{-p-2} + \frac{\theta p}{2\mu}t_1^{-2}\left(L(x,y^*)-L(x^*,y)\right) \\ &\leq \mathcal{E}(t_1)t^{-p-2} + \frac{1}{8}\left(L(x,y^*)-L(x^*,y)\right), \end{aligned}$$

where the first inequality follows from the definition of t_1 and the last one follows from $\theta = \frac{3}{2\alpha} < \frac{1}{2}$. Obviously,

$$L(x,y^*)-L(x^*,y) \leq \frac{8}{7}\mathcal{E}(t_1)t^{-p-2} = \frac{8}{7}\mathcal{E}(t_1)t^{-\frac{2\alpha}{3}}. \tag{2.45}$$

By (2.42), it follows that

$$\|x-x^*\|^2+\|y-y^*\|^2 \leq \frac{2}{\mu}(L(x,y^*)-L(x^*,y)) \leq \frac{16}{7\mu}\mathcal{E}(t_1)t^{-\frac{2\alpha}{3}}.$$

With this we arrive at (2.36) and (2.37). Revisiting (2.43) and (2.45), for every $t \geq t_1$ we have

$$\mathcal{E}(t) \leq \mathcal{E}(t_1) + \frac{4\theta p}{7\mu}t^{-2}\mathcal{E}(t_1) \leq \left(1+\frac{4\theta p}{7\mu}t_1^{-2}\right)\mathcal{E}(t_1) \leq \left(1+\frac{4\theta^2}{7}\right)\mathcal{E}(t_1) \leq \frac{8}{7}\mathcal{E}(t_1).$$

Then, the definition of $\mathcal{E}(t)$ gives

$$\|(x(t)-x^*)+\theta t\dot{x}(t)\|^2 \leq \frac{16}{7}t^{-p}\mathcal{E}(t_1).$$

Therefore,

$$\begin{aligned} \|t\dot{x}(t)\| &\leq \frac{1}{\theta}\left(\|(x(t)-x^*)+\theta t\dot{x}(t)\| + \|(x(t)-x^*)\|\right) \\ &\leq \frac{1}{\theta}\left(\sqrt{\frac{16}{7}t^{-p}\mathcal{E}(t_1)} + \sqrt{\frac{16}{7\mu}\mathcal{E}(t_1)t^{-p-2}}\right) \\ &\leq \frac{4}{\theta}\sqrt{\frac{1}{7}\mathcal{E}(t_1)}\left(1+\sqrt[4]{\frac{\theta}{\mu p}}\right)t^{-\frac{p}{2}}, \end{aligned}$$

by which it follows that

$$\|\dot{x}(t)\|^2 \leq \frac{64\alpha^2}{63}\mathcal{E}(t_1)\left(1+\sqrt[4]{\frac{9\alpha}{4\mu(\alpha-3)}}\right)^2 t^{-p-2}.$$

Similarly, we have

$$\|\dot{y}(t)\|^2 \leq \frac{64\alpha^2}{63}\mathcal{E}(t_1)\left(1+\sqrt[4]{\frac{9\alpha}{4\mu(\alpha-3)}}\right)^2 t^{-p-2},$$

Since $p+2=\frac{2}{3}\alpha$, we see that (2.38) holds, which completes the proof. □

3 A fast algorithm for bilinearly coupled saddle point problems

Fast gradient algorithms originating from various second order dynamical systems in the spirit of Nesterov's accelerated gradient method have been proposed in [9, 15, 24, 30]. In the following, we will investigate the convergence properties of a discretized version of the dynamical system (1.5), i. e. the convergence rate for the primal dual gap and the convergence of the trajectory to a primal-dual optimal solution. For convenience, we suppose $\alpha \geq 3$ and $\frac{1}{\alpha-1} \leq \theta \leq \frac{1}{2}$. In order to provide a reasonable time discretization of the dynamical system (1.5), we follow the techniques mentioned in Boţ et al. [15]. Let

$$\begin{cases} u := x + \frac{t}{\alpha-1}\dot{x}, \\ v := y + \frac{t}{\alpha-1}\dot{y}, \end{cases} \quad \text{and} \quad \begin{cases} u^\gamma := \gamma\left(x+\theta t\dot{x}\right), \\ v^\gamma := \gamma\left(y+\theta t\dot{y}\right), \end{cases}$$

where $\gamma := \frac{1}{\theta(\alpha-1)} \in \left[\frac{2}{\alpha-1}, 1\right]$. Then, (1.5) can be reformulated as

$$\begin{cases} \dot{u} = -\frac{t}{\alpha-1}\nabla f(x) - \frac{t}{\gamma(\alpha-1)}A^*v^\gamma, \\ \dot{v} = \frac{t}{\gamma(\alpha-1)}Au^\gamma - \frac{t}{\alpha-1}\nabla g(y), \\ u = x + \frac{t}{\alpha-1}\dot{x}, \\ u^\gamma = \gamma\left(x+\theta t\dot{x}\right), \\ v = y + \frac{t}{\alpha-1}\dot{y}, \\ v^\gamma = \gamma\left(y+\theta t\dot{y}\right), \end{cases} \tag{3.1}$$

Let $\sigma, \rho > 0$. For every $k \geq 1$, we take for x and y two different time steps

$$\sigma_k := \sigma\left(1+\frac{\alpha-1}{k}\right) \text{ and } \rho_k := \sigma\left(1+\frac{\alpha-1}{k}\right) \tag{3.2}$$

respectively, and set $x(\sqrt{\sigma_k}k) \approx x_{k+1}$, $u(\sqrt{\sigma_k}k) \approx u_{k+1}$ and $u^\gamma(\sqrt{\sigma_k}k) \approx u^\gamma_{k+1}$, which follows from the fact that $\sqrt{\sigma_k}k$ is closer to $\sqrt{\sigma}(k+1)$ than $\sqrt{\sigma}k$. Similarly, we set $y(\sqrt{\rho_k}k) \approx y_{k+1}$, $v(\sqrt{\rho_k}k) \approx v_{k+1}$ and $v^\gamma(\sqrt{\rho_k}k) \approx v^\gamma_{k+1}$. The implicit discretization scheme for (3.1) at time $t := \sqrt{\sigma_k}k$ for x, u, u^γ and at time $t := \sqrt{\rho_k}k$ for y, v, v^γ gives then

$$\begin{cases} \frac{u_{k+1}-u_k}{\sqrt{\sigma_k}} = -\frac{\sqrt{\sigma_k}k}{\alpha-1}\nabla f(z_k) - \frac{\sqrt{\sigma_k}k}{\gamma(\alpha-1)}A^*\tilde{v}^\gamma_{k+1}, \\ \frac{u_{k+1}-u_k}{\sqrt{\rho_k}} = \frac{\sqrt{\rho_k}k}{\gamma(\alpha-1)}Au^\gamma_{k+1} - \frac{\sqrt{\rho_k}k}{\alpha-1}\nabla g(\lambda_k), \\ u_{k+1} = x_{k+1} + \frac{\sqrt{\sigma_k}k}{\alpha-1}\frac{x_{k+1}-x_k}{\sqrt{\sigma_k}}, \\ u^\gamma_{k+1} = \gamma x_{k+1} + \frac{\sqrt{\sigma_k}k}{\alpha-1}\frac{x_{k+1}-x_k}{\sqrt{\sigma_k}}, \\ v_{k+1} = y_{k+1} + \frac{\sqrt{\rho_k}k}{\alpha-1}\frac{y_{k+1}-y_k}{\sqrt{\rho_k}}, \\ v^\gamma_{k+1} = \gamma y_{k+1} + \frac{\sqrt{\rho_k}k}{\alpha-1}\frac{y_{k+1}-y_k}{\sqrt{\rho_k}}, \end{cases} \tag{3.3}$$

where z_k and λ_k are obtained by the constructions of the proximal operator, and the term $\tilde{v}_{k+1}^{\gamma}$ first constructed in [15] is appropriately chosen to achieve an easily implementable iterative scheme. Here we choose z_k, λ_k and $\tilde{v}_{k+1}^{\gamma}$ as follows:

$$z_k := x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}),\ \lambda_k := y_k + \frac{k-1}{k+\alpha-1}(y_k - y_{k-1}),\ \tilde{v}_{k+1}^{\gamma} := v_{k+1}^{\gamma} + (1-\gamma)(y_{k+1} - y_k).$$

To derive another acceleration scheme for the original bilinearly coupled convex-concave saddle point problem (1.1), we consider the following change of variables for every $k \geq 1$: $t_k = 1 + \frac{k-1}{\alpha-1}$. Then, by recalling the definitions of z_k, λ_k, we can transform (3.3) into the following:

$$\begin{cases} z_k := x_k + \frac{t_k-1}{t_{k+1}}(x_k - x_{k-1}), \\ x_{k+1} = z_k - \sigma\nabla f(z_k) - \sigma A^*\tilde{v}_{k+1}^{\gamma}, \\ u_{k+1}^{\gamma} = \gamma x_{k+1} + (t_{k+1}-1)(x_{k+1} - x_k), \\ \lambda_k := y_k + \frac{t_k-1}{t_{k+1}}(y_k - y_{k-1}), \\ y_{k+1} = \lambda_k - \rho\nabla g(\lambda_k) + \frac{\rho}{\gamma}Au_{k+1}^{\gamma}, \\ v_{k+1}^{\gamma} = \gamma y_{k+1} + (t_{k+1}-1)(y_{k+1} - y_k), \end{cases} \tag{3.4}$$

By the relations given in (3.4), we have

$$\begin{aligned} \tilde{v}_{k+1}^{\gamma} &= \gamma y_k + t_{k+1}(y_{k+1} - y_k) \\ &= \gamma y_k + (t_k - 1)(y_k - y_{k-1}) + t_{k+1}\left(y_{k+1} - y_k - \frac{t_k-1}{t_{k+1}}(y_k - y_{k-1})\right) \\ &= v_k^{\gamma} + \frac{\rho t_{k+1}}{\gamma}\left(A(\gamma x_{k+1} + (t_{k+1}-1)(x_{k+1} - x_k)) - \gamma\nabla g(\lambda_k)\right). \end{aligned} \tag{3.5}$$

Substituting (3.5) into the second equation of (3.4), we arrive at

$$\frac{x_{k+1}}{\sigma} + \frac{\rho}{\gamma}t_{k+1}(t_{k+1}+\gamma-1)A^*Ax_{k+1} = \frac{z_k}{\sigma} - \nabla f(z_k) - A^*v_k^{\gamma} + +\frac{\rho}{\gamma}t_{k+1}(t_{k+1}-1)A^*Ax_k + \rho t_{k+1}A^*\nabla g(\lambda_k).$$

Now we are in a position to present our main algorithm as follows:

Algorithm 1 Choose $\gamma, \sigma, \rho, m > 0$ be such that

$$0 < \max\{m, \sigma l_f, \rho l_g\} \leq \gamma \leq 1. \tag{3.6}$$

Choose $\{t_k\}_{k\geq 1}$ as a nondecreasing sequence such that

$$t_1 := 1 \text{ and } t_{k+1}^2 - mt_{k+1} - t_k^2 \leq 0,\ \forall\, k \geq 1. \tag{3.7}$$

Given $x_0 = x_1, y_0 = y_1$. For every $k \geq 1$, we set

$$z_k := x_k + \frac{t_k-1}{t_{k+1}}(x_k - x_{k-1}), \tag{3.8}$$

$$\lambda_k := y_k + \frac{t_k-1}{t_{k+1}}(y_k - y_{k-1}), \tag{3.9}$$

$$v_k^{\gamma} := \gamma y_k + (t_k - 1)(y_k - y_{k-1}), \tag{3.10}$$

$$s_{k+1} := \frac{\rho}{\gamma}t_{k+1}(t_{k+1}+\gamma-1), \tag{3.11}$$

$$\eta_k := \frac{1}{t_{k+1}+\gamma-1}\left((t_{k+1}-1)Ax_k + \gamma\nabla g(\lambda_k)\right), \tag{3.12}$$

$$x_{k+1} := \arg\min_{x\in X}\left\{\frac{1}{2\sigma}\|x - z_k\|^2 + \frac{s_{k+1}}{2\gamma}\|Ax - \eta_k\|^2 + \langle\nabla f(z_k), x\rangle + \frac{1}{\gamma}\langle v_k^{\gamma}, Ax\rangle\right\}, \tag{3.13}$$

$$u_{k+1}^{\gamma} := \gamma x_{k+1} + (t_{k+1}-1)(x_{k+1} - x_k), \tag{3.14}$$

$$y_{k+1} := \lambda_k - \rho\nabla g(\lambda_k) + \frac{\rho}{\gamma}Au_{k+1}^{\gamma}. \tag{3.15}$$

3.1 Preliminary estimates

To prove the convergence of Algorithm 1, we first provide some important and useful estimates.

Lemma 3.1. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1. Then, for every $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and every $k \geq 1$ it holds that*

$$\begin{aligned} f(x_{k+1}) \leq f(x) - \frac{1}{\gamma}\langle v_{k+1}^{\gamma}, Ax_{k+1} - Ax\rangle + \frac{1-\gamma}{\gamma}\langle y_k - y_{k+1}, Ax_{k+1} - Ax\rangle \\ + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x\rangle + \frac{l_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x)\|^2 \end{aligned} \tag{3.16}$$

and

$$g(y_{k+1}) \leq g(y) + \langle \nabla g(\lambda_k), y_{k+1} - y\rangle + \frac{l_g}{2}\|y_{k+1} - \lambda_k\|^2 - \frac{1}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y)\|^2. \tag{3.17}$$

Proof. By (3.13), we have

$$\nabla f(z_k) + \frac{1}{\gamma}A^* v_k^{\gamma} + \frac{1}{\sigma}(x_{k+1} - z_k) + \frac{s_{k+1}}{\gamma}A^*(Ax_{k+1} - \eta_k) = 0. \tag{3.18}$$

According to (3.11), (3.12) and (3.14), (3.15), we also have

$$\begin{aligned} \frac{s_{k+1}}{\gamma}A^*(Ax_{k+1} - \eta_k) &= \frac{\rho}{\gamma^2}t_{k+1}(t_{k+1} + \gamma - 1)\left(Ax_{k+1} - \frac{1}{t_{k+1} + \gamma - 1}\left((t_{k+1} - 1)Ax_k + \gamma\nabla g(\lambda_k)\right)\right) \\ &= \frac{\rho}{\gamma^2}t_{k+1}\left(Au_{k+1}^{\gamma} - \gamma\nabla g(\lambda_k)\right) \\ &= \frac{1}{\gamma}\left(v_{k+1}^{\gamma} - v_k^{\gamma} + (1-\gamma)(y_{k+1} - y_k)\right). \end{aligned} \tag{3.19}$$

Then, by substituting (3.18) in (3.19), we see that

$$\nabla f(z_k) = -\frac{1}{\gamma}A^* v_{k+1}^{\gamma} + \frac{1-\gamma}{\gamma}(A^* y_k - A^* y_{k+1}) + \frac{1}{\sigma}(z_k - x_{k+1}) \tag{3.20}$$

holds Let $(x,y) \in \mathcal{X} \times \mathcal{Y}$ be fixed. By the Descent Lemma, we obtain

$$f(x_{k+1}) \leq f(z_k) + \langle \nabla f(z_k), x_{k+1} - z_k\rangle + \frac{l_f}{2}\|x_{k+1} - z_k\|^2,$$

and

$$f(z_k) \leq f(x) + \langle \nabla f(z_k), z_k - x\rangle - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x)\|^2.$$

Summing the above two inequalities yields

$$\begin{aligned} f(x_{k+1}) &\leq f(x) + \langle \nabla f(z_k), x_{k+1} - x\rangle + \frac{l_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x)\|^2 \\ &\leq f(x) - \frac{1}{\gamma}\langle v_{k+1}^{\gamma}, Ax_{k+1} - Ax\rangle + \frac{1-\gamma}{\gamma}\langle y_k - y_{k+1}, Ax_{k+1} - Ax\rangle \\ &\quad + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x\rangle + \frac{l_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x)\|^2, \end{aligned}$$

which completes the proof of inequality (3.16). Employing the Descent Lemma again, we have

$$g(y_{k+1}) \leq g(\lambda_k) + \langle \nabla g(\lambda_k), y_{k+1} - \lambda_k\rangle + \frac{l_g}{2}\|y_{k+1} - \lambda_k\|^2$$

and

$$g(\lambda_k) \leq g(y) + \langle \nabla g(\lambda_k), \lambda_k - y\rangle - \frac{1}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y)\|^2.$$

With this we obtain (3.17) by summing the above two inequalities. $\square$

Lemma 3.2. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1. Then, for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and every $k \geq 1$ the following inequalities hold*

$$
\begin{aligned}
& L(x_{k+1}, y) - L(x, y_{k+1}) \\
\leq \;& \frac{1-\gamma}{\gamma}\langle y_k - y_{k+1}, Ax_{k+1} - Ax\rangle + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x\rangle + \frac{1}{\rho}\langle \lambda_k - y_{k+1}, y_{k+1} - y\rangle \\
& \frac{l_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x)\|^2 + \frac{l_g}{2}\|y_{k+1} - \lambda_k\|^2 - \frac{1}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y)\|^2 \\
& -\frac{t_{k+1}-1}{\gamma}\langle y_{k+1} - y_k, Ax_{k+1} - Ax\rangle + \frac{t_{k+1}-1}{\gamma}\langle Ax_{k+1} - Ax_k, y_{k+1} - y\rangle
\end{aligned} \tag{3.21}
$$

and

$$
\begin{aligned}
& L(x_{k+1}, y) - L(x, y_{k+1}) - (L(x_k, y) - L(x, y_k)) \\
\leq \;& \frac{1-\gamma}{\gamma}\langle y_k - y_{k+1}, Ax_{k+1} - Ax_k\rangle + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x_k\rangle + \frac{1}{\rho}\langle \lambda_k - y_{k+1}, y_{k+1} - y_k\rangle \\
& +\frac{l_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2 + \frac{l_g}{2}\|y_{k+1} - \lambda_k\|^2 - \frac{1}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y_k)\|^2 \\
& +\langle y - y_{k+1}, Ax_{k+1} - Ax_k\rangle + \langle Ax_{k+1} - Ax, y_{k+1} - y_k\rangle.
\end{aligned} \tag{3.22}
$$

Proof. From Lemma 3.1, we conclude that

$$
\begin{aligned}
& L(x_{k+1}, y) - L(x, y_{k+1}) \\
= \;& f(x_{k+1}) + \langle Ax_{k+1}, y\rangle - g(y) - f(x) - \langle Ax, y_{k+1}\rangle + g(y_{k+1}) \\
\leq \;& \langle y - \frac{1}{\gamma}v^{\gamma}_{k+1}, Ax_{k+1} - Ax\rangle + \frac{1-\gamma}{\gamma}\langle y_k - y_{k+1}, Ax_{k+1} - Ax\rangle + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x\rangle \\
& +\langle \nabla g(\lambda_k) - Ax, y_{k+1} - y\rangle + \frac{l_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2l_f}\|\nabla f(z_k) - \nabla f(x)\|^2 \\
& +\frac{l_g}{2}\|y_{k+1} - \lambda_k\|^2 - \frac{1}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y)\|^2.
\end{aligned} \tag{3.23}
$$

According to (3.15), we have

$$
0 = \lambda_k - y_{k+1} - \rho\nabla g(\lambda_k) + \frac{\rho}{\gamma}Au^{\gamma}_{k+1},
$$

which further gives

$$
0 = \frac{1}{\rho}\langle \lambda_k - y_{k+1}, y_{k+1} - y\rangle + \frac{1}{\gamma}\langle Au^{\gamma}_{k+1} - \gamma\nabla g(\lambda_k), y_{k+1} - y\rangle. \tag{3.24}
$$

From (3.10) and (3.14), we further see that

$$
\begin{aligned}
& \langle y - \frac{1}{\gamma}v^{\gamma}_{k+1}, Ax_{k+1} - Ax\rangle + \frac{1}{\gamma}\langle Au^{\gamma}_{k+1} - \gamma\nabla g(\lambda_k), y_{k+1} - y\rangle \\
= \;& \langle y - y_{k+1} - \frac{t_{k+1}-1}{\gamma}(y_{k+1} - y_k), Ax_{k+1} - Ax\rangle \\
& +\frac{1}{\gamma}\langle \gamma Ax_{k+1} + (t_{k+1}-1)(Ax_{k+1} - Ax_k) - \gamma\nabla g(\lambda_k), y_{k+1} - y\rangle \\
= \;& \langle y - y_{k+1}, Ax_{k+1} - Ax\rangle - \frac{t_{k+1}-1}{\gamma}\langle y_{k+1} - y_k, Ax_{k+1} - Ax\rangle \\
& +\langle Ax_{k+1} - \nabla g(\lambda_k), y_{k+1} - y\rangle + \frac{t_{k+1}-1}{\gamma}\langle Ax_{k+1} - Ax_k, y_{k+1} - y\rangle \\
= \;& \langle y - y_{k+1}, \nabla g(\lambda_k) - Ax\rangle - \frac{t_{k+1}-1}{\gamma}\langle y_{k+1} - y_k, Ax_{k+1} - Ax\rangle + \frac{t_{k+1}-1}{\gamma}\langle Ax_{k+1} - Ax_k, y_{k+1} - y\rangle.
\end{aligned} \tag{3.25}
$$

Summing (3.23),(3.24) and (3.25) yields inequality (3.21). By recalling inequality (3.16) with $x := x_k$ and inequality (3.17) with $y := y_k$, we obtain

$$\begin{aligned}
& L(x_{k+1}, y) - L(x, y_{k+1}) - (L(x_k, y) - L(x, y_k)) \\
\leq \quad & \langle y - \frac{1}{\gamma} v_{k+1}^{\gamma}, Ax_{k+1} - Ax_k \rangle + \frac{1-\gamma}{\gamma} \langle y_k - y_{k+1}, Ax_{k+1} - Ax_k \rangle \\
& + \frac{1}{\sigma} \langle z_k - x_{k+1}, x_{k+1} - x_k \rangle + \frac{l_f}{2} \|x_{k+1} - z_k\|^2 - \frac{1}{2l_f} \|\nabla f(z_k) - \nabla f(x_k)\|^2 \\
& + \langle \nabla g(\lambda_k) - Ax, y_{k+1} - y_k \rangle + \frac{l_g}{2} \|y_{k+1} - \lambda_k\|^2 - \frac{1}{2l_g} \|\nabla g(\lambda_k) - \nabla g(y_k)\|^2. \qquad (3.26)
\end{aligned}$$

In additon, from (3.10) and (3.14), we have

$$\begin{aligned}
& \langle y - \frac{1}{\gamma} v_{k+1}^{\gamma}, Ax_{k+1} - Ax_k \rangle + \frac{1}{\gamma} \langle Au_{k+1}^{\gamma} - \gamma \nabla g(\lambda_k), y_{k+1} - y_k \rangle \\
= \quad & \langle y - y_{k+1}, Ax_{k+1} - Ax_k \rangle - \frac{t_{k+1} - 1}{\gamma} \langle y_{k+1} - y_k, Ax_{k+1} - Ax_k \rangle \\
& \langle Ax_{k+1} - \nabla g(\lambda_k), y_{k+1} - y_k \rangle + \frac{t_{k+1} - 1}{\gamma} \langle Ax_{k+1} - Ax_k, y_{k+1} - y_k \rangle \\
= \quad & \langle y - y_{k+1}, Ax_{k+1} - Ax_k \rangle + \langle Ax_{k+1} - \nabla g(\lambda_k), y_{k+1} - y_k \rangle. \qquad (3.27)
\end{aligned}$$

By summing (3.26), (3.27) and (3.24) with $y := y_k$ we obtain (3.22), this completes the proof. $\square$

For $(x, y) \in X \times Y$ and $k \geq 1$, we introduce the following energy function:

$$\begin{aligned}
\mathcal{E}_k(x, y) \quad := \quad & t_k(t_k - 1 + \gamma) \left(L(x_k, y) - L(x, y_k)\right) + \frac{1}{2\sigma} \|u_k^{\gamma} - \gamma x\|^2 + \frac{1}{2\rho} \|v_k^{\gamma} - \gamma y\|^2 \\
& + \frac{\gamma(1-\gamma)}{2\sigma} \|x_k - x\|^2 + \frac{\gamma(1-\gamma)}{2\rho} \|y_k - y\|^2 + \frac{(1-\gamma)(t_k - 1)}{2\rho} \|y_k - y_{k-1}\|^2.
\end{aligned}$$

It is obvious that $\mathcal{E}_k(x^*, y^*) \geq 0$ for every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$. Next we show an important inequality for this family of energy functions which will play a significant role in the following analysis.

Proposition 3.1. *Let $\{(x_k, y_k)\}_{k \geq 0}$ be the sequence generated by Algorithm 1. Then, for every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$, we have*

$$\begin{aligned}
& \mathcal{E}_{k+1}(x^*, y^*) - \mathcal{E}_k(x^*, y^*) \\
\leq \quad & \left(t_{k+1}(t_{k+1} - 1) - t_k(t_k - 1 + \gamma)\right) \left(L(x_k, y^*) - L(x^*, y_k)\right) \\
& - \left(\frac{t_{k+1}^2}{2\sigma} (\gamma - l_f \sigma) + (1-\gamma) \frac{l_f t_{k+1}}{2} \right) \|x_{k+1} - z_k\|^2 - \left(\frac{t_{k+1}^2}{2\rho} (\gamma - l_g \rho) + (1-\gamma) \frac{l_g t_{k+1}}{2} \right) \|y_{k+1} - \lambda_k\|^2 \\
& - \frac{\gamma t_{k+1}}{2l_f} \|\nabla f(z_k) - \nabla f(x^*)\|^2 - \frac{t_{k+1}(t_{k+1} - 1)}{2l_f} \|\nabla f(z_k) - \nabla f(x_k)\|^2 - \frac{t_{k+1}(t_{k+1} - 1)}{2l_g} \|\nabla g(\lambda_k) - \nabla g(y_k)\|^2 \\
& - \frac{t_{k+1}}{2l_g} \left(\gamma - \rho l_g (1-\gamma)\right) \|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 - \frac{(1-\gamma) t_{k+1}}{2\rho} \|y_{k+1} - y_k - \rho \left(\nabla g(\lambda_k) - \nabla g(y^*)\right)\|^2 \\
& - \frac{(1-\gamma)(t_{k+1} - 1)}{\sigma} \|x_{k+1} - x_k\|^2 - \frac{(1-\gamma)(t_{k+1} - 1)}{2\rho} \|y_{k+1} - y_k\|^2. \qquad (3.28)
\end{aligned}$$

Proof. By taking $x := x^*$ in (3.21) and $y := y^*$ in (3.22), we have

$$\begin{aligned}
& t_{k+1}(t_{k+1}-1+\gamma)\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})\right)-t_k(t_k-1+\gamma)\left(L(x_k,y^*)-L(x^*,y_k)\right)\\
= \; & \gamma t_{k+1}\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})\right)+t_{k+1}(t_{k+1}-1)\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})-\left(L(x_k,y^*)-L(x^*,y_k)\right)\right)\\
& +\left(t_{k+1}(t_{k+1}-1)-t_k(t_k-1+\gamma)\right)\left(L(x_k,y^*)-L(x^*,y_k)\right)\\
\leq \; & \left(t_{k+1}(t_{k+1}-1)-t_k(t_k-1+\gamma)\right)\left(L(x_k,y^*)-L(x^*,y_k)\right)\\
& +\frac{(1-\gamma)t_{k+1}}{\gamma}\langle y_k-y_{k+1},\gamma(Ax_{k+1}-Ax^*)+(t_{k+1}-1)\left(Ax_{k+1}-Ax_k\right)\rangle\\
& +\frac{t_{k+1}}{\sigma}\langle z_k-x_{k+1},\gamma(x_{k+1}-x^*)+(t_{k+1}-1)\left(x_{k+1}-x_k\right)\rangle\\
& +\frac{t_{k+1}}{\rho}\langle \lambda_k-y_{k+1},\gamma(y_{k+1}-y^*)+(t_{k+1}-1)\left(y_{k+1}-y_k\right)\rangle\\
& +\frac{l_f t_{k+1}}{2}\left(t_{k+1}-1+\gamma\right)\|x_{k+1}-z_k\|^2+\frac{l_g t_{k+1}}{2}\left(t_{k+1}-1+\gamma\right)\|y_{k+1}-\lambda_k\|^2\\
& -\frac{\gamma t_{k+1}}{2l_f}\|\nabla f(z_k)-\nabla f(x^*)\|^2-\frac{t_{k+1}(t_{k+1}-1)}{2l_f}\|\nabla f(z_k)-\nabla f(x_k)\|^2\\
& -\frac{\gamma t_{k+1}}{2l_g}\|\nabla g(\lambda_k)-\nabla g(y^*)\|^2-\frac{t_{k+1}(t_{k+1}-1)}{2l_g}\|\nabla g(\lambda_k)-\nabla g(y_k)\|^2. \qquad (3.29)
\end{aligned}$$

Then, by (3.9), (3.14) and (3.15) we arrive at

$$\begin{aligned}
& \frac{(1-\gamma)t_{k+1}}{\gamma}\langle y_k-y_{k+1},\gamma(Ax_{k+1}-Ax^*)+(t_{k+1}-1)\left(Ax_{k+1}-Ax_k\right)\rangle\\
= \; & \frac{(1-\gamma)t_{k+1}}{\rho}\langle y_k-y_{k+1},y_{k+1}-\lambda_k\rangle+(1-\gamma)t_{k+1}\langle y_k-y_{k+1},\nabla g(\lambda_k)-Ax^*\rangle\\
= \; & \frac{(1-\gamma)t_{k+1}}{2\rho}\left(-\|y_{k+1}-y_k\|^2-\|y_{k+1}-\lambda_k\|^2+\|y_k-\lambda_k\|^2\right)+(1-\gamma)t_{k+1}\langle y_k-y_{k+1},\nabla g(\lambda_k)-\nabla g(y^*)\rangle\\
\leq \; & \frac{(1-\gamma)t_{k+1}}{2\rho}\left(-\|y_{k+1}-y_k\|^2+\frac{(t_k-1)^2}{t_{k+1}^2}\|y_k-y_{k-1}\|^2\right)+(1-\gamma)t_{k+1}\langle y_k-y_{k+1},\nabla g(\lambda_k)-\nabla g(y^*)\rangle\\
\leq \; & -\frac{(1-\gamma)t_{k+1}}{2\rho}\|y_{k+1}-y_k\|^2+\frac{(1-\gamma)(t_k-1)}{2\rho}\|y_k-y_{k-1}\|^2+(1-\gamma)t_{k+1}\langle y_k-y_{k+1},\nabla g(\lambda_k)-\nabla g(y^*)\rangle,
\end{aligned} \qquad (3.30)$$

where the last inequality follows from the fact that t_k is nondecreasing and $t_k \geq 1$. We notice that

$$\begin{aligned}
t_{k+1}(z_k-x_{k+1}) &= t_{k+1}(x_k-x_{k+1})+u_k^\gamma-u_{k+1}^\gamma+(1-t_{k+1})(x_k-x_{k+1})-\gamma(x_k-x_{k+1})\\
&= u_k^\gamma-u_{k+1}^\gamma+(1-\gamma)(x_k-x_{k+1}),
\end{aligned}$$

which we combine with (3.8) and (3.14) to see that

$$\begin{aligned}
& \frac{t_{k+1}}{\sigma}\langle z_k-x_{k+1},\gamma(x_{k+1}-x^*)+(t_{k+1}-1)\left(x_{k+1}-x_k\right)\rangle\\
= \; & \frac{1}{\sigma}\left(\langle u_k^\gamma-u_{k+1}^\gamma,u_{k+1}^\gamma-\gamma x^*\rangle-(1-\gamma)(t_{k+1}-1)\|x_{k+1}-x_k\|^2+(1-\gamma)\gamma\langle(x_k-x_{k+1}),(x_{k+1}-x^*)\rangle\right)\\
= \; & -\frac{1}{2\sigma}\|u_k^\gamma-u_{k+1}^\gamma\|^2-\frac{1}{2\sigma}\|u_{k+1}^\gamma-\gamma x\|^2+\frac{1}{2\sigma}\|u_k^\gamma-\gamma x\|^2-\frac{(1-\gamma)(t_{k+1}-1)}{\sigma}\|x_{k+1}-x_k\|^2\\
& -\frac{(1-\gamma)\gamma}{2\sigma}\|x_k-x_{k+1}\|^2-\frac{(1-\gamma)\gamma}{2\sigma}\|x_{k+1}-x^*\|^2+\frac{(1-\gamma)\gamma}{2\sigma}\|x_k-x^*\|^2. \qquad (3.31)
\end{aligned}$$

By (3.14) and Corollary 2.14 in [10] we deduce now

$$\begin{aligned}
-\frac{1}{2\sigma}\|u_k^\gamma - u_{k+1}^\gamma\|^2 &= -\frac{1}{2\sigma}\|u_k - u_{k+1} + (\gamma-1)(x_{k+1}-x_k)\|^2 \\
&= -\frac{\gamma}{2\sigma}\|u_k - u_{k+1}\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_{k+1}-x_k\|^2 - \frac{1-\gamma}{2\sigma}\|u_k - u_{k+1} - x_{k+1} + x_k\|^2 \\
&\leq -\frac{\gamma t_{k+1}^2}{2\sigma}\|x_{k+1}-z_k\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_{k+1}-x_k\|^2. \qquad (3.32)
\end{aligned}$$

Summing (3.31) and (3.32) yields

$$\begin{aligned}
&\frac{t_{k+1}}{\sigma}\langle z_k - x_{k+1}, \gamma(x_{k+1}-x^*) + (t_{k+1}-1)(x_{k+1}-x_k)\rangle \\
\leq\ & -\frac{1}{2\sigma}\|u_{k+1}^\gamma - \gamma x^*\|^2 + \frac{1}{2\sigma}\|u_k^\gamma - \gamma x^*\|^2 - \frac{(1-\gamma)(t_{k+1}-1)}{\sigma}\|x_{k+1}-x_k\|^2 \\
& -\frac{(1-\gamma)\gamma}{2\sigma}\|x_{k+1}-x^*\|^2 + \frac{(1-\gamma)\gamma}{2\sigma}\|x_k - x^*\|^2 - \frac{\gamma t_{k+1}^2}{2\sigma}\|x_{k+1}-z_k\|^2. \qquad (3.33)
\end{aligned}$$

By a similar discussion, we obtain the inequality

$$\begin{aligned}
&\frac{t_{k+1}}{\rho}\langle \lambda_k - y_{k+1}, \gamma(y_{k+1}-y^*) + (t_{k+1}-1)(y_{k+1}-y_k)\rangle \\
\leq\ & -\frac{1}{2\rho}\|v_{k+1}^\gamma - \gamma y^*\|^2 + \frac{1}{2\rho}\|v_k^\gamma - \gamma y^*\|^2 - \frac{(1-\gamma)(t_{k+1}-1)}{\rho}\|y_{k+1}-y_k\|^2 \\
& -\frac{(1-\gamma)\gamma}{2\rho}\|y_{k+1}-y^*\|^2 + \frac{(1-\gamma)\gamma}{2\rho}\|y_k - y^*\|^2 - \frac{\gamma t_{k+1}^2}{2\rho}\|y_{k+1}-\lambda_k\|^2. \qquad (3.34)
\end{aligned}$$

Summing (3.29), (3.30), (3.33) and (3.34), we have

$$\begin{aligned}
&\mathcal{E}_{k+1}(x^*,y^*) - \mathcal{E}_k(x^*,y^*) \\
\leq\ & (t_{k+1}(t_{k+1}-1) - t_k(t_k - 1 + \gamma))\left(L(x_k,y^*) - L(x^*,y_k)\right) \\
& -\left(\frac{t_{k+1}^2}{2\sigma}(\gamma - l_f\sigma) + (1-\gamma)\frac{l_f t_{k+1}}{2}\right)\|x_{k+1}-z_k\|^2 - \left(\frac{t_{k+1}^2}{2\rho}(\gamma - l_g\rho) + (1-\gamma)\frac{l_g t_{k+1}}{2}\right)\|y_{k+1}-\lambda_k\|^2 \\
& -\frac{\gamma t_{k+1}}{2l_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 - \frac{t_{k+1}(t_{k+1}-1)}{2l_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2 - \frac{t_{k+1}(t_{k+1}-1)}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y_k)\|^2 \\
& -\frac{\gamma t_{k+1}}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 - \frac{(1-\gamma)t_{k+1}}{2\rho}\|y_{k+1}-y_k\|^2 + (1-\gamma)t_{k+1}\langle y_k - y_{k+1}, \nabla g(\lambda_k) - \nabla g(y^*)\rangle \\
& -\frac{(1-\gamma)(t_{k+1}-1)}{\sigma}\|x_{k+1}-x_k\|^2 - \frac{(1-\gamma)(t_{k+1}-1)}{2\rho}\|y_{k+1}-y_k\|^2, \qquad (3.35)
\end{aligned}$$

and thus

$$\begin{aligned}
& -\frac{\gamma t_{k+1}}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 - \frac{(1-\gamma)t_{k+1}}{2\rho}\|y_{k+1}-y_k\|^2 + (1-\gamma)t_{k+1}\langle y_k - y_{k+1}, \nabla g(\lambda_k) - \nabla g(y^*)\rangle \\
=\ & -\frac{t_{k+1}}{2l_g}\left(\gamma - \rho l_g(1-\gamma)\right)\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 - \frac{(1-\gamma)t_{k+1}}{2\rho}\|y_{k+1}-y_k - \rho\left(\nabla g(\lambda_k) - \nabla g(y^*)\right)\|^2 \qquad (3.36)
\end{aligned}$$

By replacing the corresponding term in (3.35) with (3.36), we arrive at (3.28), which completes the proof. □

Proposition 3.2. *Let $\{(x_k,y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1. Then, for $(x^*,y^*)\in\mathbb{S}$ and* *every $k \geq 1$, the sequence $\{\mathcal{E}_k(x^*, y^*)\}_{k\geq 1}$ is nonincreasing and we have the following statements:*

$$\sum_{k\geq 1}(\gamma - m)t_k\left(L(x_k, y^*) - L(x^*, y_k)\right) < +\infty,$$

$$\sum_{k\geq 1}\left(\frac{t_{k+1}^2}{2\sigma}(\gamma - l_f\sigma) + (1-\gamma)\frac{l_f t_{k+1}}{2}\right)\|x_{k+1} - z_k\|^2 < +\infty, \sum_{k\geq 1}\frac{(1-\gamma)(t_{k+1}-1)}{2\rho}\|y_{k+1} - y_k\|^2 < +\infty,$$

$$\sum_{k\geq 1}\left(\frac{t_{k+1}^2}{2\rho}(\gamma - l_g\rho) + (1-\gamma)\frac{l_g t_{k+1}}{2}\right)\|y_{k+1} - \lambda_k\|^2 < +\infty, \sum_{k\geq 1}\frac{(1-\gamma)(t_{k+1}-1)}{\sigma}\|x_{k+1} - x_k\|^2 < +\infty,$$

$$\sum_{k\geq 1}\frac{\gamma t_{k+1}}{2l_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 < +\infty, \sum_{k\geq 1}\frac{t_{k+1}(t_{k+1}-1)}{2l_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2 < +\infty,$$

$$\sum_{k\geq 1}\frac{t_{k+1}(t_{k+1}-1)}{2l_g}\|\nabla g(\lambda_k) - \nabla g(y_k)\|^2 < +\infty, \sum_{k\geq 1}\frac{t_{k+1}}{2l_g}\left(\gamma - \rho l_g(1-\gamma)\right)\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 < +\infty.$$

Proof. Due to $\gamma - l_g\rho \geq 0$ and $0 \leq \gamma \leq 1$, it follows that $l_g\rho(1-\gamma) \leq \gamma(1-\gamma) \leq \gamma$ and $\frac{t_{k+1}^2}{2\rho}(\gamma - l_g\rho) + (1-\gamma)\frac{l_g t_{k+1}}{2} \geq 0$. Similarly, we get $\frac{t_{k+1}^2}{2\sigma}(\gamma - l_f\sigma) + (1-\gamma)\frac{l_f t_{k+1}}{2} \geq 0$. According to (3.7), we have

$$t_{k+1}(t_{k+1}-1) - t_k(t_k - 1 + \gamma) \leq (m-1)t_{k+1} + (1-\gamma)t_k \leq (m-\gamma)t_k \leq 0.$$

Thus, all the coeffcients in the right-hand side of (3.28) are nonpositive, it follows that the sequence $\{\mathcal{E}_k(x^*, y^*)\}_{k\geq 1}$ is nonincreasing and so $\mathcal{E}_{k+1}(x^*, y^*) \leq \mathcal{E}_k(x^*, y^*)$. We complete the proof via Lemma 1.1 in [15]. □

Remark 3.1. *Since $t_k \geq 1$, we have*

$$\gamma t_k^2\left(L(x_k, y^*) - L(x^*, y_k)\right) \quad \leq \quad t_k(t_k - 1 + \gamma)\left(L(x_k, y^*) - L(x^*, y_k)\right) \leq \mathcal{E}_k(x^*, y^*) \leq \mathcal{E}_1(x^*, y^*)$$

and so

$$L(x_k, y^*) - L(x^*, y_k) \leq \frac{1}{\gamma t_k^2}\mathcal{E}_1(x^*, y^*).$$

In Boţ et al. [15], the authors show that three most prominent choices for the sequence $\{t_k\}_{k\geq 1}$, i.e. Nesterov rule [28], Chambolle-Dossal rule [18] and Attouch-Cabot rule [7] (here this rule requires $k \geq [\alpha] + 1$) are all satified the conditions (3.7) *in Algorithm 1. Since $t_k \geq \frac{k+1}{2}$ in the Nesterov rule and $t_k = 1 + \frac{k-1}{\alpha-1}$ in the Chambolle-Dossal rule and the Attouch-Cabot rule, we see that*

$$L(x_k, y^*) - L(x^*, y_k) = O\left(1/k^2\right).$$

holds.

3.2 Boundedness and convergence of the iterates

In this subsection, we show the boundedness of the sequence generated by Algorithm 1. under some additional assumptions on the parameters of the algorithm. By Proposition 3.2, we notice that the sequence $\{\mathcal{E}_k(x^*, y^*)\}_{k\geq 1}$ is nonincreasing, and so by$\mathcal{E}_k(x^*, y^*) \leq \mathcal{E}_1(x^*, y^*)$ it follows that

$$\frac{1}{2\sigma}\|u_k^\gamma - \gamma x^*\|^2 + \frac{1}{2\rho}\|v_k^\gamma - \gamma y^*\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_k - x^*\|^2 + \frac{\gamma(1-\gamma)}{2\rho}\|y_k - y^*\|^2 \leq \mathcal{E}_1(x^*, y^*) < +\infty.$$

We obtain that the sequences $\{u_k^\gamma\}_{k\geq 1}$ and $\{v_k^\gamma\}_{k\geq 1}$ are bounded. In addition, if $\gamma < 1$, the sequences $\{x_k\}_{k\geq 1}$ and $\{y_k\}_{k\geq 1}$ also are bounded. From (3.14) and (3.10), we have

$$
\begin{aligned}
t_k(x_k - x_{k-1}) &= u_k^\gamma - \gamma x^* + (1-\gamma)(x_k - x^*) - (x_{k-1} - x^*),\\
t_k(y_k - y_{k-1}) &= v_k^\gamma - \gamma y^* + (1-\gamma)(y_k - y^*) - (y_{k-1} - y^*),
\end{aligned}
$$

which yield the sequences $\{t_k(x_k - x_{k-1})\}_{k\geq 1}$ and $\{t_k(y_k - y_{k-1})\}_{k\geq 1}$ are bounded.

We can now show the boundedness of the sequences $\{x_k\}_{k\geq 1}$ and $\{y_k\}_{k\geq 1}$ under the mild condition (3.37) below, which has been proposed in [11, 15]. Moreover, Boţ et al. [15] prove that some classical inertial parameters rules satisfy (3.37), i.e., the Nesterov rule [28], the Chambolle-Dossal rule [18] and the Attouch-Cabot rule [7]. For brevity, we define on $\mathcal{X}\times\mathcal{Y}$ the inner product

$$
\langle h, h'\rangle_{\mathcal{W}} = \langle (x,y),(x',y')\rangle_{\mathcal{W}} = \frac{1}{\sigma}\langle x, x'\rangle_{\mathcal{X}} + \frac{1}{\rho}\langle y, y'\rangle_{\mathcal{Y}}, \quad \forall\, h := (x,y), h' := (x',y') \in \mathcal{X}\times\mathcal{Y},
$$

and the norm induced by this scalar product,

$$
\|u\|_{\mathcal{W}} = \sqrt{\frac{1}{\sigma}\|x\|^2 + \frac{1}{\rho}\|y\|^2}, \qquad \forall\, h := (x,y) \in \mathcal{X}\times\mathcal{Y}.
$$

Proposition 3.3. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1. Suppose that*

$$
\tau := \inf_{k\geq 1} \frac{t_k}{k} > 0. \tag{3.37}
$$

Then, the sequences $\{x_k\}_{k\geq 1}$, $\{y_k\}_{k\geq 1}$, $\{t_k(x_k - x_{k-1})\}_{k\geq 1}$ and $\{t_k(y_k - y_{k-1})\}_{k\geq 1}$ are bounded.

Proof. Let $(x^*, y^*) \in \mathbb{S}$ be fixed. We denote

$$
h^* := (x^*, y^*) \in \mathbb{S}, \text{ and } h_k := (x_k, y_k) \in \mathcal{X}\times\mathcal{Y},\ \forall\, k \geq 1.
$$

By (3.14) in Algorithm 1 and Corollary 2.14 in [10], for every $k \geq 1$, we see that

$$
\begin{aligned}
\|u_k^\gamma - \gamma x^*\|^2 &= \|(t_k - 1 + \gamma)(x_k - x^*) - (t_k - 1)(x_{k-1} - x^*)\|^2\\
&= \gamma(t_k - 1 + \gamma)\|x_k - x^*\|^2 - \gamma(t_k - 1)\|x_{k-1} - x^*\|^2 + (t_k - 1 + \gamma)(t_k - 1)\|x_k - x_{k-1}\|^2
\end{aligned}
$$

holds. Similarly, we have

$$
\|v_k^\gamma - \gamma y^*\|^2 = \gamma(t_k - 1 + \gamma)\|y_k - y^*\|^2 - \gamma(t_k - 1)\|y_{k-1} - y^*\|^2 + (t_k - 1 + \gamma)(t_k - 1)\|y_k - y_{k-1}\|^2,
$$

and so the energy function can be rewritten as

$$
\begin{aligned}
\mathcal{E}_k(x,y) &= t_k(t_k - 1 + \gamma)\left(L(x_k, y) - L(x, y_k)\right) + \frac{\gamma}{2}t_k\|h_k - h^*\|_{\mathcal{W}}^2 - \frac{\gamma}{2}(t_k - 1)\|h_{k-1} - h^*\|_{\mathcal{W}}^2\\
&\quad + \frac{1}{2}(t_k - 1 + \gamma)(t_k - 1)\|h_k - h_{k-1}\|_{\mathcal{W}}^2 + \frac{(1-\gamma)(t_k - 1)}{2\rho}\|y_k - y_{k-1}\|^2.
\end{aligned} \tag{3.38}
$$

From the fact that $\mathcal{E}_k(x^*, y^*)$ is nonincreasing, for every $k \geq 1$ we get

$$
\frac{\gamma}{2}t_k\|h_k - h^*\|_{\mathcal{W}}^2 - \frac{\gamma}{2}(t_k - 1)\|h_{k-1} - h^*\|_{\mathcal{W}}^2 \leq \mathcal{E}_k(x^*, y^*) \leq \mathcal{E}_1(x^*, y^*),
$$

from which it follows that

$$
\frac{\gamma}{2}t_k\|h_k - h^*\|_{\mathcal{W}}^2 \leq \frac{\gamma}{2}(t_k - 1)\|h_{k-1} - h^*\|_{\mathcal{W}}^2 + \mathcal{E}_1(x^*, y^*) \leq \frac{\gamma}{2}t_{k-1}\|h_{k-1} - h^*\|_{\mathcal{W}}^2 + \mathcal{E}_1(x^*, y^*),
$$

where the second inequality holds due to$t_k - t_{k-1} < 1$ which is a straightforward result via (3.7). (The proof is also provided in Lemma 3.5 in [15].) We set $t_0 := 0$ for convention. After summing up (3.39) from 1 to k, we have for every $k \geq 1$

$$\frac{\gamma}{2} t_k \|h_k - h^*\|^2_{\mathcal{W}} \leq k\mathcal{E}_1(x^*, y^*) \tag{3.39}$$

and so

$$\|h_k - h^*\|^2_{\mathcal{W}} \leq \frac{2k}{\gamma t_k}\mathcal{E}_1(x^*, y^*) \leq \frac{2}{\gamma\tau}\mathcal{E}_1(x^*, y^*) < +\infty. \tag{3.40}$$

With this, we conclude that $\{x_k\}_{k\geq 1}$, $\{y_k\}_{k\geq 1}$ are bounded and consequently $\{t_k(x_k - x_{k-1})\}_{k\geq 1}$, $\{t_k(y_k - y_{k-1})\}_{k\geq 1}$ are bounded. ☐

Proposition 3.4. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1 with* (3.37) *and $(x^*, y^*) \in \mathbb{S}$. In addition, we assume that $0 < \max\{m, l_f\sigma, l_g\rho\} < \gamma \leq 1$. Then it holds that*

$$\sum_{k\geq 1}\left(\left(1 - \frac{m}{\gamma}\right)t_{k+1} - 1\right)\|A^*\left(y_k - y^*\right)\|^2 < +\infty, \tag{3.41}$$

$$\sum_{k\geq 1}\frac{(t_{k+1} - 1)^3}{\gamma^2}\|A^*\left(y_{k+1} - y_k\right)\|^2 < +\infty, \tag{3.42}$$

$$\left(1 - \frac{m}{\gamma}\right)\sum_{k\geq 1} t_k\|A\left(x_k - x^*\right)\|^2 < +\infty, \tag{3.43}$$

$$\sum_{k\geq 1} t_k\frac{(t_{k+1} - 1)^2}{\gamma^2}\|A\left(x_{k+1} - x_k\right)\|^2 < +\infty. \tag{3.44}$$

Moreover, there exists $M > 0$ such that

$$\|A^*\left(y_k - y^*\right)\| \leq \frac{M}{t_k - 1}, \text{ and } \|A\left(x_k - x^*\right)\| \leq \frac{M}{t_k}.$$

Proof. From (3.20), we have

$$A^*\left(\frac{1}{\gamma}v^\gamma_{k+1} - y^*\right) = -\nabla f(z_k) - A^*y^* + \frac{1-\gamma}{\gamma}A^*(y_k - y_{k+1}) + \frac{1}{\sigma}(z_k - x_{k+1}).$$

By Theorem 3.29 and $t_k \geq 1$, it follows that

$$\begin{aligned}\sum_{k\geq 1}(t_{k+1} - 1)\left\|A^*\left(\frac{1}{\gamma}v^\gamma_{k+1} - y^*\right)\right\|^2 &\leq 3\sum_{k\geq 1}(t_{k+1} - 1)\|\nabla f(z_k) - \nabla f(x^*)\|^2 \\ &+ \frac{3(1-\gamma)^2\|A\|^2}{\gamma^2}\sum_{k\geq 1}(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 + \frac{3}{\sigma^2}\sum_{k\geq 1}(t_{k+1} - 1)\|z_k - x_{k+1}\|^2 < +\infty.\end{aligned}$$

According to (3.10), for every $k \geq 1$, we have

$$\begin{aligned}A^*\left(\frac{1}{\gamma}v^\gamma_{k+1} - y^*\right) &= A^*\left(y_{k+1} + \frac{t_{k+1} - 1}{\gamma}(y_{k+1} - y_k) - y^*\right) \\ &= \left(1 + \frac{t_{k+1} - 1}{\gamma}\right)A^*\left(y_{k+1} - y^*\right) - \frac{t_{k+1} - 1}{\gamma}A^*\left(y_k - y^*\right).\end{aligned}$$

Then, by Corollary 2.14 in [10], it follows that

$$\begin{aligned}\left\|A^*\left(\frac{1}{\gamma}v^\gamma_{k+1} - y^*\right)\right\|^2 &= \left(1 + \frac{t_{k+1} - 1}{\gamma}\right)\|A^*\left(y_{k+1} - y^*\right)\|^2 - \frac{t_{k+1} - 1}{\gamma}\|A^*\left(y_k - y^*\right)\|^2 \\ &+ \frac{t_{k+1} - 1}{\gamma}\left(1 + \frac{t_{k+1} - 1}{\gamma}\right)\|A^*\left(y_{k+1} - y_k\right)\|^2.\end{aligned}$$

But by (3.7), we see that

$$\begin{aligned}\frac{(t_{k+1}-1)^2}{\gamma}-(t_k-1)\left(1+\frac{t_k-1}{\gamma}\right) &\le \left(1-\frac{2}{\gamma}\right)(t_{k+1}-t_k)+1-\left(1-\frac{m}{\gamma}\right)t_{k+1}\\ &\le -\left(\left(1-\frac{m}{\gamma}\right)t_{k+1}-1\right),\end{aligned}$$

where the last inequality follows from the fact that $\left(1-\frac{2}{\gamma}\right)<0$ and $\{t_k\}$ is nondecreasing. Therefore,

$$\begin{aligned}&(t_{k+1}-1)\left(1+\frac{t_{k+1}-1}{\gamma}\right)\|A^*(y_{k+1}-y^*)\|^2\\ =\;&(t_k-1)\left(1+\frac{t_k-1}{\gamma}\right)\|A^*(y_k-y^*)\|^2+(t_{k+1}-1)\left\|A^*\left(\frac{1}{\gamma}v_{k+1}^{\gamma}-y^*\right)\right\|^2\\ &+\left(\frac{(t_{k+1}-1)^2}{\gamma}-(t_k-1)\left(1+\frac{t_k-1}{\gamma}\right)\right)\|A^*(y_k-y^*)\|^2\\ &-\frac{(t_{k+1}-1)^2}{\gamma}\left(1+\frac{t_{k+1}-1}{\gamma}\right)\|A^*(y_{k+1}-y_k)\|^2\\ \le\;&(t_k-1)\left(1+\frac{t_k-1}{\gamma}\right)\|A^*(y_k-y^*)\|^2+(t_{k+1}-1)\left\|A^*\left(\frac{1}{\gamma}v_{k+1}^{\gamma}-y^*\right)\right\|^2\\ &-\left(\left(1-\frac{m}{\gamma}\right)t_{k+1}-1\right)\|A^*(y_k-y^*)\|^2-\frac{(t_{k+1}-1)^3}{\gamma^2}\|A^*(y_{k+1}-y_k)\|^2.\end{aligned}$$

By the assumption (3.37), we notice that $\left(1-\frac{m}{\gamma}\right)t_{k+1}-1$ will be nonnegative when $k\ge k_1$, where $k_1=\left[\frac{\gamma}{(\gamma-m)\tau}\right]$. For every $k\ge k_1$ let us set

$$\begin{aligned}a_k &:= (t_k-1)\left(1+\frac{t_k-1}{\gamma}\right)\|A^*(y_k-y^*)\|^2\ge 0,\\ b_k &:= \left(\left(1-\frac{m}{\gamma}\right)t_{k+1}-1\right)\|A^*(y_k-y^*)\|^2+\frac{(t_{k+1}-1)^3}{\gamma^2}\|A^*(y_{k+1}-y_k)\|^2\ge 0,\\ d_k &:= (t_{k+1}-1)\left\|A^*\left(\frac{1}{\gamma}v_{k+1}^{\gamma}-y^*\right)\right\|^2\ge 0.\end{aligned}$$

By employing Lemma 1.1 in [15] we then obtain that

$$\sum_{k\ge k_1}\left(\left(1-\frac{m}{\gamma}\right)t_{k+1}-1\right)\|A^*(y_k-y^*)\|^2<+\infty,\ \sum_{k\ge k_1}\frac{(t_{k+1}-1)^3}{\gamma^2}\|A^*(y_{k+1}-y_k)\|^2<+\infty,$$

and the sequence $\left\{(t_k-1)\left(1+\frac{t_k-1}{\gamma}\right)\|A^*(y_k-y^*)\|^2\right\}$ is convergent and bounded. Thus, there exists $M>0$ such that $(t_k-1)^2\|A^*(y_k-y^*)\|^2\le a_k\le M^2$, for every $k\ge 1$. We notice that the convergence or divergence of series will not be affected by its first few terms, and we thus arrive at (3.41) and (3.42). Similarly, we obtain (3.43), (3.44) and the fact that $t_k\left(1+\frac{1}{\gamma}(t_k-1)\right)\|A^*(x_k-x^*)\|^2$ is convergent. Since $t_k\le\left(1+\frac{1}{\gamma}(t_k-1)\right)$, we also have

$$t_k^2\|A^*(x_k-x^*)\|^2\le t_k\left(1+\frac{1}{\gamma}(t_k-1)\right)\|A^*(x_k-x^*)\|^2\le M^2,$$

and so $\|A^*(x_k-x^*)\|\le\frac{M}{t_k}$. This completes the proof. ☐

Proposition 3.5. *Let $\{(x_k,y_k)\}_{k\ge 0}$ be the sequence generated by Algorithm and let $0<m<\gamma<1$. Then, for every $(x^*,y^*)\in\mathbb{S}$, the limit $\lim_{k\to+\infty}\|(x_k,y_k)-(x^*,y^*)\|_{\mathcal{W}}$ exists.*

Proof. By Proposition 3.2, we notice that $\mathcal{E}_{k+1}(x^*,y^*) \le \mathcal{E}_k(x^*,y^*)$. By the reformulation (3.38) of the energy function $\mathcal{E}_k(x^*,y^*)$, for every $k \ge 1$ we have

$$\begin{aligned}
& t_{k+1}(t_{k+1}-1+\gamma)\left((L(x_{k+1},y)-L(x,y_{k+1}))+\frac{1}{2}\|h_{k+1}-h_k\|_{\mathcal{W}}^2\right) \\
& +\frac{\gamma}{2}t_{k+1}\left(\|h_{k+1}-h^*\|_{\mathcal{W}}^2-\|h_k-h^*\|_{\mathcal{W}}^2\right)+\frac{(1-\gamma)}{2\rho}t_{k+1}\|y_{k+1}-y_k\|^2 \\
\le\ & (t_k-1)(t_k-1+\gamma)\left((L(x_k,y)-L(x,y_k))+\frac{1}{2}\|h_k-h_{k-1}\|_{\mathcal{W}}^2\right) \\
& +\frac{\gamma}{2}(t_k-1)\left(\|h_k-h^*\|_{\mathcal{W}}^2-\|h_{k-1}-h^*\|_{\mathcal{W}}^2\right)+\frac{(1-\gamma)(t_k-1)}{2\rho}\|y_k-y_{k-1}\|^2 \\
& (t_k-1+\gamma)\left(L(x_k,y)-L(x,y_k)\right)+\frac{1}{2}(t_{k+1}-1+\gamma)\|h_{k+1}-h_k\|_{\mathcal{W}}^2+\frac{1-\gamma}{2\rho}\|y_{k+1}-y_k\|^2. \quad (3.45)
\end{aligned}$$

Denote

$$\begin{aligned}
& a_k := \|h_k-h^*\|_{\mathcal{W}}^2 \ge 0, \\
& \beta_k := (t_k-1+\gamma)\left((L(x_k,y)-L(x,y_k))+\frac{1}{2}\|h_k-h_{k-1}\|_{\mathcal{W}}^2\right)+(a_k-a_{k-1})+\frac{(1-\gamma)}{2\rho}\|y_k-y_{k-1}\|^2, \\
& l_k := (t_k-1+\gamma)\left(L(x_k,y)-L(x,y_k)\right)+\frac{1}{2}(t_{k+1}-1+\gamma)\|h_{k+1}-h_k\|_{\mathcal{W}}^2+\frac{1-\gamma}{2\rho}\|y_{k+1}-y_k\|^2 \ge 0.
\end{aligned}$$

From these definitions, it is obvious that $a_{k+1} \le a_k + \beta_{k+1}$. By (3.45) we have $t_{k+1}\beta_{k+1} \le (t_k-1)\beta_k + d_k$. In addition, from Proposition 3.2, we notice that $\sum_{k\ge 1} l_k < +\infty$ if $0 < m < \gamma < 1$. Thus, by Lemma 4.1 in [15], we conclude that $\{a_k\}$ is convergent which completes the proof. □

Theorem 3.1. *Let $\{(x_k,y_k)\}_{k\ge 0}$ be the sequence generated by Algorithm 1. Assume further that the sequence $\{t_k\}_{k\ge 1}$ is chosen to satisfy* (3.37)*, that $0 < \max\{m, l_f\sigma, l_g\rho\} < \gamma < 1$ holds and that $(x^*,y^*) \in \mathbb{S}$. Then,*

$$\begin{aligned}
& \|\nabla f(x_k)-\nabla f(x^*)\| = o\left(1/\sqrt{k}\right), \quad \|\nabla g(y_k)-\nabla g(y^*)\| = o\left(1/\sqrt{k}\right), \\
& \|Ax_k - Ax^*\| = o\left(1/\sqrt{k}\right), \quad \|A^*y_k - A^*y^*\| = o\left(1/\sqrt{k}\right).
\end{aligned}$$

Consequently,

$$\|\nabla_x L(x,y)\| = o\left(1/\sqrt{k}\right),\ \|\nabla_y L(x,y)\| = o\left(1/\sqrt{k}\right).$$

Proof. From the results of Proposition 3.2, we see that

$$\lim_{k\to+\infty} t_{k+1}\|\nabla f(z_k)-\nabla f(x^*)\|^2 = 0, \lim_{k\to+\infty} t_{k+1}(t_{k+1}-1)\|\nabla f(x_k)-\nabla f(z_k)\|^2 = 0$$

holds By (3.37), it follows that

$$\lim_{k\to+\infty}\sqrt{k}\|\nabla f(z_k)-\nabla f(x^*)\| = 0, \lim_{k\to+\infty}\sqrt{k}\|\nabla f(x_k)-\nabla f(z_k)\| = 0$$

and so

$$\lim_{k\to+\infty}\sqrt{k}\|\nabla f(x_k)-\nabla f(x^*)\| \le \lim_{k\to+\infty}\sqrt{k}\|\nabla f(z_k)-\nabla f(x^*)\| + \lim_{k\to+\infty}\sqrt{k}\|\nabla f(x_k)-\nabla f(z_k)\| = 0,$$

which further gives $\|\nabla f(x_k)-\nabla f(x^*)\| = o\left(1/\sqrt{k}\right)$. Similarly, $\|\nabla g(y_k)-\nabla g(y^*)\| = o\left(1/\sqrt{k}\right)$ holds. By (3.41) and (3.37), we obtain $\|A^*y_k - Ay^*\| = o\left(1/\sqrt{k}\right)$ and so $\|\nabla_x L(x,y)\| = o\left(1/\sqrt{k}\right)$. Similarly, $\|\nabla_y L(x,y)\| = o\left(1/\sqrt{k}\right)$. This completes the proof. □

Now we can present the main result and establish the convergence of the sequence of iterates generated by Algorithm 1, which is the discrete counterpart of Theorem 2.4.

Theorem 3.2. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1. Suppose that the sequence $\{t_k\}_{k\geq 1}$ has been chosen such that* (3.37) *holds. Further assume $0 < \max\{m, l_f\sigma, l_g\rho\} < \gamma < 1$ and $(x^*, y^*) \in \mathbb{S}$. Then, the sequence $\{(x_k, y_k)\}_{k\geq 1}$ converges weakly to a primal-dual optimal solution of the bilinearly coupled convex-concave saddle point problem* (1.1).

Proof. This proof is a straightforward result via Opial's Lemma and the fact that the graph of the monotone operator $\mathcal{T}_L$ in (1.3) is sequentially closed. It is similar to the proof of Theorem 2.4, so we omit it here. □

4 Conclusion and perspectives

As a brief review of the main result, the inertial primal-dual dynamics (1.5) introduce a novel class of first-order algorithms for a bilinearly coupled saddle point problem. These algorithms not only maintain the fast convergence rate of the primal-dual values found in several classical accelerated algorithms but also possess additional exciting properties, such as the convergence of gradients towards zero, global convergence of the iterates to optimal saddle points. By recalling the proof process of (2.10), we can obtain the convergence rate $O\left(1/t^2\right)$ of the primal-dual gap for (1.5) without assuming continuous differentiability of all functions. In light of this, it would be interesting to design a new discretization of (1.5) with the objective of achieving the $O\left(1/k^2\right)$ rate when g is a convex lower semicontinuous and proper function. Additionally, it is worth considering (1.5) in a more general context, which includes situations involving general viscous damping, Hessian-driven damping, and temporal rescaling.

References

[1] F. Alvarez, On the minimizing property of a second order dissipative system in Hilbert spaces, SIAM J. Control Optim., 38(2000), pp. 1102-1119.

[2] F. Alvarez, H. Attouch, J. Bolte, and P. Redont, A second-order gradient-like dissipative dynamical system with hessian-driven damping: Application to optimization and mechanics, Journal de mathématiques pures et appliquées, 81(8)(2022), 747-779.

[3] V. Apidopoulos, J. F. Aujol, and C. Dossal, Convergence rate of inertial Forward–Backward algorithm beyond Nesterov's rule, Math. Program., 180(2020), pp. 137-156.

[4] H. Attouch and J. Peypouquet, The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$, SIAM J. Optim., 26(2016), pp. 1824-1834.

[5] H. Attouch, Z. Chbani, and H. Riahi, Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$, ESAIM: Contr. Optim. CA., 25(2019), pp. 1-34.

[6] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi, Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics, J. Optimiz. Theory Appl., 193(2022), pp. 704-736.

[7] H. Attouch and A. Cabot, Convergence rates of inertial forward-backward algorithms. SIAM J. Optimiz., 28(2018), pp. 849-874.

[8] H. Attouch, X. Goudou, and P. Redont, The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system, Commun. Contemp. Math., 2(2000), pp. 1-34.

[9] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, Math. Program. 168(2018), pp. 123–175.

[10] H.H. Bauschke and P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer, New York, 2011.

[11] H.H. Bauschke, M.N. Bui, and X. Wang, Applying FISTA to optimization problems (with or) without minimizers, Math. Program., 184(2020), pp. 349-381.

[12] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci., 2(2009), pp. 183-202.

[13] A. Beck, First-order Methods in Optimization. Society for Industrial and Applied Mathematics, Philadelphia, 2017.

[14] R. I. Boţ and D. Nguyen, Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping, J. Differ. Equations, 303(2021), pp. 369-406.

[15] R.I. Boţ, E.R. Csetnek, and DK. Nguyen, Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates, Math. Program., 2022, https://doi.org/10.1007/s10107-022-01879-4.

[16] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vis., 40(2011), pp. 120–145.

[17] A. Chambolle and T. Pock, On the ergodic convergence rates of a first-order primal–dual algorithm, Math. Program., 2016, 159(1-2): 253-287.

[18] A. Chambolle and C. Dossal, On the Convergence of the Iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm". J. Optim. Theory Appl., 166(2015), pp. 968–982.

[19] Y. Chen, G. Lan, Y. Ouyang, Optimal primal-dual methods for a class of saddle point problems, SIAM J. Optimiz., 24(2014), pp. 1779-1814.

[20] L. Condat, A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, J. Optim. Theory Appl., 158(2013), pp. 460-479.

[21] S.S. Du, J. Chen, L. Li, L. Xiao, and D. Zhou, Stochastic variance reduction methods for policy evaluation, International Conference on Machine Learning, 2017: 1049-1058.

[22] B. He, F. Ma, S. Xu, and X. Yuan, A generalized primal-dual algorithm with improved convergence condition for saddle point problems, SIAM J. Imaging Sci., 15(2022), pp. 1157-1183.

[23] X. He, R. Hu, and Y.P. Fang, Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems, SIAM J. Control Optim., 2021, 59(5): 3278-3301.

[24] X. He, R. Hu, and Y.P. Fang, Fast primal–dual algorithm via dynamical system for a linearly constrained convex optimization problem, Automatica, 146(2022), 110547.

[25] D. Kovalev, A. Gasnikov, and P. Richtárik, Accelerated primal-dual gradient method for smooth and convex-concave saddle-point problems with bilinear coupling, arXiv preprint arXiv:2112.15199, 2021.

[26] C.J. Li, A. Yuan, G. Gidel, and MI. Jordan, Nesterov Meets Optimism: Rate-Optimal Optimistic-Gradient-Based Method for Stochastic Bilinearly-Coupled Minimax Optimization, arXiv preprint arXiv:2210.17550, 2022.

[27] R. May, Asymptotic for a second-order evolution equation with convex potential andvanishing damping term, Turk, J. Math., 41(2017), pp. 681-685.

[28] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(\frac{1}{k^2})$, Soviet Math. Dokl., 27(1983), pp. 372–376.

[29] O. Sebbouh, C. Dossal, and A. Rondepierre, Convergence rates of damped inertial dynamics under geometric conditions and perturbations, SIAM J. Optimiz., 30(2020), pp. 1850-1877.

[30] W. Su, S. Boyd, and E.J. Candes, A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights. J. Mach. Learn. Res., 17(2016), pp. 1-43.

[31] K.K. Thekumparampil, N. He, and S. Oh, Lifted primal-dual method for bilinearly coupled smooth minimax optimization, International Conference on Artificial Intelligence and Statistics, 2022, pp. 4281-4308.

[32] G. Teschl, Ordinary Differential Equations and Dynamical Systems, American Mathematical Society, 2012.