Fast convergence of the primal-dual dynamical system and algorithms for a nonsmooth bilinearly coupled saddle point problem

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

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

[c] Mathematical Sciences, 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 discretizations associated with a nonsmooth bilinearly coupled convex-concave saddle point problem. We derive the convergence rate of the primal-dual gap for the second-order dynamical system with asymptotically vanishing damping term. Based on the implicit discretization, we propose a primal-dual algorithm and show the non-ergodic convergence rate under a general setting for the inertial parameters when one objective function is a continuously differentiable convex function and another one is a proper, convex and lower semicontinuous function. We present the $O\left(1/k^2\right)$ convergence rate under three classical rules proposed by Nesterov, Chambolle-Dossal and Attouch-Cabot without the strong convexity, which is compatible with the results of the continuous-time dynamic system. We further present a primal-dual algorithm based on the explicit discretization when both objective functions are continuously differentiable convex functions. We show the corresponding non-ergodic convergence rate and prove that the sequence of iterates generated by the primal-dual algorithm weakly converges to a primal-dual optimal solution.

Key Words: Saddle point problem; Primal-dual dynamical system; Convergence rate; 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

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

*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

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

which is naturally associated with the following convex optimization problem

$$\min_{x\in X} f(x) + g^*(Ax),$$

where $f : \mathcal{X} \to \mathbb{R}$ is a continuously differentiable convex function and $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function, and $g^* : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is the Fenchel conjugate of g. Here we call $\langle Ax, y\rangle$ the bilinear coupling term. Saddle point problems arise regularly in determining primal-dual pairs of constrained convex optimisation problems, recently they have been studied widely due to many relevant and challenging applications in the field of imaging processing [13, 16], reinforcement learning [18, 34] and generative adversarial networks [8, 12].

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

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

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

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

It is obvious that the optimality condition (1.2) can be reformulated as $\mathcal{T}_L(x_*, y_*) \ni 0$ and $\mathbb{S}$ can be viewed as the set of zeros of the operator $\mathcal{T}_L$. Since f (resp. g) is convex and continuously differentiable (resp. convex and lower semicontinuous) and A (resp. A^*) is a linear operator, it is obvious that $\mathcal{T}_L(x, y)$ is maximally monotone (see Corollary 20.28 in [6]). Moreover, we note 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 [13] showed an ergodic convergence rate $O(1/k)$ of their celebrated first-order primal-dual algorithm for the primal-dual gap of problem (1.1) with f and g are proper, convex and lower semicontinuous. The authors showed that their primal-dual algorithm has strong connections with other well-known methods, such as the extra-gradient method [25], the Douglas-Rachford splitting method [28] and the preconditioned ADMM method [19]. When either f or g is strongly convex (the partially strongly convex case), they proved the ergodic convergence rate $O\left(1/k^2\right)$ of an accelerated primal-dual algorithm. Additionally, an ergodic linear convergence rate has been shown when both f and g are strongly convex (the strongly convex case). By employing the Bregman distance, Chambolle and Pock [14] established the ergodic convergence rates with simpler proofs for a more general case in which f has a nonsmooth plus smooth composite structure. Based on the primal-dual algorithm described in [13], He et al. [20] proposed 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. With the assumption that f is a convex and Fréchet differentiable function with L_f-Lipschitz continuous gradient and g is proper, convex and lower semicontinuous, Chen et al. [16] presented an ergodic convergence rate of $O\left(L_f/k^2 + \|A\|/k\right)$ for the primal-dual gap of problem (1.1). Jiang et al. [24] provided an accelerated $O\left(1/k^2\right)$ rate and linear convergence for the strongly convex case and partially strongly convex case, respectively. When both f and g exhibit a nonsmooth plus smooth composite structure, He et al. [23] showed a non-ergodic convergence rate $O\left(1/k\right)$ under convexity assumptions, a non-ergodic convergence rate of $O\left(1/k^2\right)$ for the partially strongly convex case, and an ergodic linear convergence rate for the strongly convex case. Under the assumption that both f and g are smooth, Kovalev et al. [26] proposed an accelerated primal-dual gradient method for solving the saddle point problem and showed linear convergence when the objective function is strongly convex-concave, convex-strongly concave, or even just convex-concave. Thekumparampil et al. [34] 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 [13, 14, 16, 17, 20] 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 optimization problems. A decisive step was taken by Su et al. in [33], 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 [7, 31] in the case $\alpha = 3$. Moreover, Attouch et al. [3] showed that any trajectory of (1.4) weakly converges to a minimizer of Φ when $\alpha > 3$ and established strong convergence properties for various practical settings. In addition, Attouch and Peypouquet [4] and May [29] showed that the asymptotic convergence rate of (1.4) is $o\left(1/k^2\right)$ when $\alpha > 3$.

Subsequently, the inertial dynamic method has been generalized to linear equality constrained convex optimization problems by employing the augmented Lagrangian approach. Attouch et al. [2] 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 [9] 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. [10] presented the corresponding numerical optimization algorithm originating from the second-order dynamical system in [9]. They were the first to provide convergence results regarding the sequence of iterates generated by a fast primal-dual 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. [27] 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. Motivated by the works described above, we consider the following second order primal-dual dynamical system with asymptotically vanishing viscous damping:

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

where $t_0>0, \alpha>0, \theta>0$ and $(x_0, y_0), (\dot{x}_0, \dot{y}_0)\in\mathcal{X}\times\mathcal{Y}$. If g is differentiable, we have only made minor adjustments by replacing the symbols "∂" and "$\ni$" with "∇" and "$=$" in the second line of (1.5) respectively. By unfolding the expressions of the gradients of $L(\cdot,\cdot)$ in the dynamical system (1.5), we have the following reformulation of system (1.5):

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

In this paper, we design two numerical algorithms based on the discretization of the second-order dynamical system (1.5) to solve problem (1.1). Our main contributions are as follows:

(a) We show a convergence rate of $O\left(1/t^2\right)$ for the primal-dual dynamical system (1.5) with asymptotically vanishing viscous damping term and present the corresponding inertial algorithm based on the implicit discretization when f is a continuously differentiable convex function with Lipschitz continuous gradient and g is a proper, convex and lower semicontinuous function. A general setting has been considered for the inertial parameters which covers three classical rules proposed by Nesterov [31], Chambolle-Dossal [15] and Attouch-Cabot [1]. We obtain a non-ergodic convergence rate of $O\left(1/k^2\right)$ for the primal-dual gap under these rules which improves the ergodic convergence rate $O\left(L_f/k^2+L_A/k\right)$ rate derived in [16]. In contrast to [14, 23, 24], we obtain the rate $O\left(1/k^2\right)$ without the assumption of strongly convexity.

(b) We also develop a primal-dual algorithm based on explicit discretization when both f and g are two continuously differentiable convex functions with Lipschitz continuous gradients. For smooth bilinearly coupled convex-concave saddle point problems, our non-ergodic $O\left(1/k^2\right)$ convergence rate of the primal-dual gap under the three classical rules of the inertial parameters improves the ergodic $O\left(1/k\right)$ rate for general smooth saddle problems described in [30]. In addition, we show that the sequence of iterates generated by our algorithm weakly converges to a primal-dual solution in a general setting which covers the rules of Chambolle-Dossal [15] and Attouch-Cabot [1]. This algorithm, based on the discretization of a continuous energy function, is different from the one descried in Boţ et al. [10]. Our main result can be seen as an extension of their result for linear equality constrained convex optimization problems.

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. In Section 3, we present a primal-dual algorithm derived from the implicit discretization of the dynamical system and derive the convergence rate of the primal-dual gap within a general setting for the inertial parameters. In Section 4, we present the numerical primal-dual algorithm and the convergence of the sequence of iterates when both f and g are smooth, before we summarize the results in Section 5.

2 The primal-dual dynamical system

In this section, we suppose f is a convex continuously differentiable function with L_f-Lipschitz continuous gradient and g is a proper, convex and lower semicontinuous function. To derive the asymptotic behavior of the dynamical system (1.5), we note that the standard way to analyse such systems is based on energy (Lyapunov) functions. Many energy functions have been proposed to study dynamcial systems with various damping terms and time scaling terms, see e. g. [2, 3, 32, 33], and choosing an appropriate one is crucial. Motivated by the one introduced in Attouch et al. [3] and Boţ and Nguyen [9], we define the following energy function $\mathcal{E}_{\alpha,\theta} : [t_0, +\infty) \to \mathbb{R}$ as

$$\mathcal{E}_{\alpha,\theta}(t) = \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{E}_2 \tag{2.1}$$

with

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

where $\xi = \theta\alpha - \theta - 1 \geq 0$.

Theorem 2.1. *Let $(x(t), y(t))$ be a solution of dynamical system (1.5) and $(x^*, y^*) \in \mathbb{S}$. Suppose $\alpha \geq 3$ and $\frac{1}{\alpha-1} \leq \theta \leq \frac{1}{2}$, it holds*

$$L(x(t), y^*) - L(x^*, y(t)) \leq \frac{\mathcal{E}_{\alpha,\theta}(t_0)}{\theta^2 t^2}, \tag{2.2}$$

$$(1 - 2\theta)\int_{t_0}^{+\infty} t(L(x(t), y^*) - L(x^*, y(t)))dt \leq \frac{\mathcal{E}_{\alpha,\theta}(t_0)}{\theta} < +\infty, \tag{2.3}$$

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

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

Proof. Differentiating $\mathcal{E}_i(t)$ with respect to t, $i = 0, 1, 2$, we have

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

where $\eta \in \partial g(y)$. Combining these terms, we arrive at

$$\begin{aligned}
\dot{\mathcal{E}}_{\alpha,\theta}(t) &= 2\theta^2 t(L(x, y^*) - L(x^*, y)) + \theta(1 + \theta - \theta\alpha)t\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right)\\
&\quad -\theta t\left(\langle x - x^*, \nabla f(x)\rangle - \langle x^*, A^*y\rangle + \langle y - y^*, \eta\rangle + \langle Ax, y^*\rangle\right)\\
&\leq 2\theta^2 t(L(x, y^*) - L(x^*, y)) + \theta(1 + \theta - \theta\alpha)t\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right)\\
&\quad +\theta t\left(f(x^*) - f(x) + \langle x^*, A^*y\rangle + g(y^*) - g(y) - \langle Ax, y^*\rangle\right)\\
&= \theta(2\theta - 1)t(L(x, y^*) - L(x^*, y)) + \theta(1 + \theta - \theta\alpha)t\left(\|\dot{x}\|^2 + \|\dot{y}\|^2\right),
\end{aligned} \tag{2.5}$$

where the inequality follows from the convexity of f and g. From the assumption, we obtain $\dot{\mathcal{E}}_{\alpha,\theta}(t) \leq 0$ and so $\mathcal{E}_{\alpha,\theta}(t)$ is nonincreasing on $[t_0, +\infty)$. For every $t \geq t_0$, it holds that

$$\begin{aligned}
\mathcal{E}_{\alpha,\theta}(t) &= \theta^2 t^2 (L(x(t), y^*) - L(x^*, y(t))) + \frac{1}{2}\|(x(t) - x^*) + \theta t \dot{x}(t)\|^2 + \frac{\theta\alpha - \theta - 1}{2}\|x(t) - x^*\|^2 \\
&\quad + \frac{1}{2}\|(y(t) - y^*) + \theta t \dot{y}(t)\|^2 + \frac{\theta\alpha - \theta - 1}{2}\|y(t) - y^*\|^2 \\
&\leq \mathcal{E}_{\alpha,\theta}(t_0),
\end{aligned} \tag{2.6}$$

which yields (2.2). For every $t \geq t_0$, by integrating (2.5) from t_0 to t, we have

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

All items inside the integrals are nonnegative. Thus, we arrive at (2.3) and (2.4) by passing $t \to +\infty$. From (2.6), we arrive at

$$\|\dot{x}(t)\| \leq \frac{1}{\theta t}\left(\|(x(t) - x^*) + \theta t \dot{x}(t)\| + \|x(t) - x^*\|\right) \leq \frac{1}{\theta t}\left(1 + \frac{1}{\theta\alpha - \theta - 1}\right)\sqrt{2\mathcal{E}_{\alpha,\theta}(t_0)},$$

which yields $\|\dot{x}(t)\| = O(\frac{1}{t})$. Similarly we have $\|\dot{y}(t)\| = O(\frac{1}{t})$. □

We have shown a $O\left(1/t^2\right)$ convergence rate of the primal-dual gap for the dynamic system (1.5). Moreover, it is not difficult to prove that the primal-dual trajectory of the second-order dynamical system (1.5) asymptotically weakly converges to a primal-dual optimal solution of the original saddle point problem (1.1) when $\alpha > 3$ and both f and g are continuously differentiable convex functions with Lipschitz continuous gradient. In this paper we mainly focus on the convergence rates of numerical algorithms that are derived from discretizations of the dynamic system (1.5). Next, we will describe two primal-dual algorithms that also exhibit the corresponding $O\left(1/k^2\right)$ convergence rates, which is compatible with the results in the continuous case. For this, in the following we always suppose $\alpha \geq 3$ and $\frac{1}{\alpha-1} \leq \theta \leq \frac{1}{2}$ to achieve fast convergence rates.

3 A primal-dual algorithm based on the implicit discretization

In order to provide a reasonable time discretization of the dynamical system (1.5), we follow the techniques described in Attouch et al. [3], Boţ et al. [10], and He et al. [21]. Let

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

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

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

Since f and g in system (3.1) do not necessarily have the same degree of smoothness, we consider two different time steps for them respectively. Let $\sigma > 0$. For x we consider the time step

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

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 $\sqrt{\sigma}k$. The first three lines of (3.1) at time $t := \sqrt{\sigma_k}k$ for x, u, u^γ then give

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

where z_k is chosen by the construction of the proximal operator. To be specific, the second line of (3.3) yields

$$x_{k+1} = \frac{\alpha - 1}{k + \alpha - 1}u_{k+1} + \frac{k}{k + \alpha - 1}x_k, \tag{3.4}$$

and consequently we take the following choice for z_k:

$$z_k = \frac{\alpha - 1}{k + \alpha - 1}u_k + \frac{k}{k + \alpha - 1}x_k. \tag{3.5}$$

Employing the second line of (3.3) again, we get

$$z_k = \frac{\alpha - 1}{k + \alpha - 1}\left(x_k + \frac{k-1}{\alpha - 1}(x_k - x_{k-1})\right) + \frac{k}{k + \alpha - 1}x_k = x_k + \frac{k-1}{k + \alpha - 1}(x_k - x_{k-1}). \tag{3.6}$$

In addition, by (3.4) and (3.5), we arrive at

$$u_{k+1} - u_k = \frac{k + \alpha - 1}{\alpha - 1}(x_{k+1} - z_k). \tag{3.7}$$

Consequently, (3.3) can be reformulated as follows:

$$\begin{cases} x_{k+1} = z_k - \sigma\nabla f(z_k) - \frac{\sigma}{\gamma}A^* v^\gamma_{k+1}, \\ z_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}), \\ u^\gamma_{k+1} = \gamma x_{k+1} + \frac{k}{\alpha-1}(x_{k+1} - x_k). \end{cases} \tag{3.8}$$

Now, for every $k \geq 1$, set $\bar{t}_k = k$, $y_k = y(\bar{t}_k)$, $\lambda_k = v(\bar{t}_k)$ and $v^\gamma_k = v^\gamma(\bar{t}_k)$. We then take the follow discretization scheme for the last three lines in (3.1),

$$\begin{cases} \tilde{v}_{k+1} - v_k \in \frac{k}{\gamma(\alpha-1)}Au^\gamma_{k+1} - \frac{k}{\alpha-1}\partial g(y_{k+1}), \\ v_{k+1} = y_{k+1} + \frac{k}{\alpha-1}(y_{k+1} - y_k), \\ v^\gamma_{k+1} = \gamma y_{k+1} + \frac{k}{\alpha-1}(y_{k+1} - y_k), \end{cases} \tag{3.9}$$

where we replace v_{k+1} with a suitable term $\tilde{v}_{k+1}$ to obtain a reasonably executable iterative scheme. This approach also has been taken by Boţ et al. [10] in which the authors strive to derive an easily implementable numerical algorithm from their discretization of a second order dynamical system for a linear equality constrained convex optimization problem. They focus on an improvement on the dual variable term v^γ_{k+1}, while here we want to choose a suitable $\tilde{v}_{k+1}$ such that $\tilde{v}_{k+1} - v_{k+1} \to 0$ as $k \to +\infty$. To be specific, we choose $\tilde{v}_{k+1} = v_{k+1} - \frac{\alpha-1}{k+\alpha-1}(v_{k+1} - v_k)$. We then have $\tilde{v}_{k+1} - v_{k+1} \to 0$ when $k \to +\infty$ as long as $v_{k+1} - v_k$ is bounded for every $k \geq 1$. Actually, from the following Proposition 3.1, $v_{k+1} - v_k$ is bounded under some mild conditions. Then, we can reformulate the first line in (3.9) as

$$v_{k+1} - v_k \in \frac{k+\alpha-1}{\gamma(\alpha-1)} A u_{k+1}^{\gamma} - \frac{k+\alpha-1}{\alpha-1} \partial g(y_{k+1}). \tag{3.10}$$

Following Attouch and Cabot [1] and Boţ et al. [10] for general iterative schemes, we use the following change of variables for every $k \geq 1$:

$$t_k = 1 + \frac{k-1}{\alpha-1} = \frac{k+\alpha-2}{\alpha-1},$$

which yields $t_{k+1} - 1 = \frac{k}{\alpha-1}$ and $\frac{t_k-1}{t_{k+1}} = \frac{\frac{k-1}{\alpha-1}}{1+\frac{k}{\alpha-1}} = \frac{k-1}{k+\alpha-1}$. Therefore, combing (3.8), (3.9), (3.10) and the definition of t_k, we arrive at the following discretization scheme of dynamical system (1.5):

$$\begin{cases} x_{k+1} = z_k - \sigma \nabla f(z_k) - \frac{\sigma}{\gamma} A^* v_{k+1}^{\gamma}, \\ z_k = x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}), \\ u_{k+1}^{\gamma} = \gamma x_{k+1} + (t_{k+1} - 1)(x_{k+1} - x_k), \\ v_{k+1} - v_k \in \frac{t_{k+1}}{\gamma} A u_{k+1}^{\gamma} - t_{k+1} \partial g(y_{k+1}), \\ v_{k+1} = y_{k+1} + (t_{k+1} - 1)(y_{k+1} - y_k), \\ v_{k+1}^{\gamma} = \gamma y_{k+1} + (t_{k+1} - 1)(y_{k+1} - y_k). \end{cases} \tag{3.11}$$

By the relations given in (3.11), we have

$$v_{k+1} - v_k = t_{k+1}(y_{k+1} - y_k) - (t_k - 1)(y_k - y_{k-1}) \tag{3.12}$$

and

$$\begin{aligned} A u_{k+1}^{\gamma} &= (t_{k+1} + \gamma - 1) A x_{k+1} - (t_{k+1} - 1) A x_k \\ &= (t_{k+1} + \gamma - 1) A (z_k - \sigma \nabla f(z_k)) - (t_{k+1} - 1) A x_k - \frac{\sigma}{\gamma} (t_{k+1} + \gamma - 1) A A^* v_{k+1}^{\gamma} \\ &= \xi_k - \frac{\sigma}{\gamma} (t_{k+1} + \gamma - 1)^2 A A^* \left(y_{k+1} - \frac{t_{k+1} - 1}{t_{k+1} + \gamma - 1} y_k \right), \end{aligned} \tag{3.13}$$

where $\xi_k = (t_{k+1} + \gamma - 1) A (z_k - \sigma \nabla f(z_k)) - (t_{k+1} - 1) A x_k$. Substituting (3.12) and (3.13) into the fourth line of (3.11), we arrive at

$$0 \in \partial g(y_{k+1}) + y_{k+1} - y_k - \frac{(t_k - 1)}{t_{k+1}}(y_k - y_{k-1}) - \frac{1}{\gamma}\xi_k + \frac{\sigma}{\gamma^2} (t_{k+1} + \gamma - 1)^2 A A^* \left(y_{k+1} - \frac{t_{k+1} - 1}{t_{k+1} + \gamma - 1} y_k \right).$$

Now we are in a position to present our first algorithm:

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

$$0 < \max\{m, \sigma L_f\} \leq \gamma \leq 1. \tag{3.14}$$

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

$$t_1 \geq 1 \text{ and } t_{k+1}^2 - m t_{k+1} - t_k^2 \leq 0, \ \forall\, k \geq 1. \tag{3.15}$$

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

$$
\begin{aligned}
z_k &:= x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}),\\
\xi_k &:= (t_{k+1} + \gamma - 1)\, A\,(z_k - \sigma \nabla f(z_k)) - (t_{k+1} - 1)\, A x_k,\\
\bar{y}_k &:= y_k + \frac{(t_k - 1)}{t_{k+1}}(y_k - y_{k-1}),\\
s_{k+1} &:= \frac{\sigma}{\gamma^2}\,(t_{k+1} + \gamma - 1)^2,\\
\zeta_k &:= \frac{t_{k+1} - 1}{t_{k+1} + \gamma - 1} y_k,\\
y_{k+1} &:= \arg\min_{y \in \mathcal{Y}} \left\{ g(y) + \frac{1}{2}\|y - \bar{y}_k\|^2 + \frac{s_{k+1}}{2}\|A^*(y - \zeta_k)\|^2 + \frac{1}{\gamma}\langle \eta_k, y\rangle \right\},\\
v_{k+1}^{\gamma} &:= \gamma y_{k+1} + (t_{k+1} - 1)(y_{k+1} - y_k),\\
x_{k+1} &:= z_k - \sigma \nabla f(z_k) - \frac{\sigma}{\gamma} A^* v_{k+1}^{\gamma}.
\end{aligned}
$$

The subproblem in Algorithm 1 determining y_{k+1} has a special splitting structure which can be solved by some classical splitting methods such as the proximal method and the corresponding accelerated scheme of FISTA [5, 7]. We notice that we would obtain a simplified version of Algorithm 1 when $\gamma = 1$, and this version enjoys the same convergence rate as the case $\gamma < 1$. In what follows we still consider the general case $\gamma \leq 1$, since some results for $\gamma \leq 1$ will be crucial in the analysis of Algorithm 3 in the following section.

3.1 Convergence analysis of Algorithm 1

Before discussing the convergence properties of Algorithm 1, we first introduce the following equations which will be used repeatedly

$$
2\langle a, b\rangle = \|a + b\|^2 - \|a\|^2 - \|b\|^2, \tag{3.16}
$$

$$
\frac{1}{s+t}\|sa + tb\|^2 = s\|a\|^2 + t\|b\|^2 - \frac{st}{s+t}\|a - b\|^2. \tag{3.17}
$$

where a, b lie in a Hilbert space and $s, t \in \mathbb{R}$ such that $s + t \neq 0$. Next, we provide some useful inequalities.

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

$$
\begin{aligned}
&L(x_{k+1}, y^*) - L(x^*, y^*) \qquad (3.18)\\
&\leq -\frac{1}{\gamma}\left\langle A^*\left(v_{k+1}^{\gamma} - \gamma y^*\right), x_{k+1} - x^*\right\rangle + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x^*\rangle + \frac{L_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&L(x_{k+1}, y^*) - L(x_k, y^*) \qquad (3.19)\\
&\leq -\frac{1}{\gamma}\left\langle A^*\left(v_{k+1}^{\gamma} - \gamma y^*\right), x_{k+1} - x_k\right\rangle + \frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x_k\rangle + \frac{L_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2.
\end{aligned}
$$

Proof. Since $f(x)$ is a convex continuously differentiable function with L_f-Lipschitz continuous gradient, 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

$$\begin{aligned} & f(x_{k+1}) - f(x) \\ \leq \ & \langle \nabla f(z_k), x_{k+1} - x \rangle + \frac{L_f}{2} \|x_{k+1} - z_k\|^2 - \frac{1}{2L_f} \|\nabla f(z_k) - \nabla f(x)\|^2 \\ = \ & -\frac{1}{\gamma} \langle A^* v_{k+1}^\gamma, x_{k+1} - x \rangle + \frac{1}{\sigma} \langle z_k - x_{k+1}, x_{k+1} - x \rangle + \frac{L_f}{2} \|x_{k+1} - z_k\|^2 - \frac{1}{2L_f} \|\nabla f(z_k) - \nabla f(x)\|^2, \end{aligned} \tag{3.20}$$

where the last equation follows from the first line of (3.11). By taking inequality (3.20) with $x := x^*$ and adding $\langle x_{k+1} - x^*, A^* y^* \rangle$ on both sides, we obtain

$$\begin{aligned} & f(x_{k+1}) + \langle A(x_{k+1} - x^*), y^* \rangle - f(x^*) \\ \leq & -\frac{1}{\gamma} \left\langle A^* \left(v_{k+1}^\gamma - \gamma y^* \right), x_{k+1} - x^* \right\rangle + \frac{1}{\sigma} \langle z_k - x_{k+1}, x_{k+1} - x^* \rangle + \frac{L_f}{2} \|x_{k+1} - z_k\|^2 - \frac{1}{2L_f} \|\nabla f(z_k) - \nabla f(x^*)\|^2. \end{aligned} \tag{3.21}$$

Similarly, by taking inequality (3.20) with $x := x_k$ and adding $\langle x_{k+1} - x_k, A^* y^* \rangle$ on both sides, we have

$$\begin{aligned} & f(x_{k+1}) + \langle A(x_{k+1} - x_k), y^* \rangle - f(x_k) \\ \leq & -\frac{1}{\gamma} \left\langle A^* \left(v_{k+1}^\gamma - \gamma y^* \right), x_{k+1} - x_k \right\rangle + \frac{1}{\sigma} \langle z_k - x_{k+1}, x_{k+1} - x_k \rangle + \frac{L_f}{2} \|x_{k+1} - z_k\|^2 - \frac{1}{2L_f} \|\nabla f(z_k) - \nabla f(x_k)\|^2. \end{aligned} \tag{3.22}$$

By recalling the definition of $L(x, y)$, we complete the proof. □

For every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$, we introduce the following energy function:

$$E(k) := t_{k+1}(t_{k+1} - 1) \left(L(x_k, y^*) - L(x^*, y_k) \right) + E_1(k) + E_2(k),$$

where

$$E_1(k) := \frac{1}{2\sigma} \|u_k^\gamma - \gamma x^*\|^2 + \frac{\gamma(1-\gamma)}{2\sigma} \|x_k - x^*\|^2 \text{ and } E_2(k) := \frac{1}{2} \|v_k^\gamma - \gamma y^*\|^2 + \frac{\gamma(1-\gamma)}{2} \|y_k - y^*\|^2.$$

It is obvious that $E(k) \geq 0$ for every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$. Next we show two important inequalities for $E_1(k)$ and $E_2(k)$ which will play a significant role in the following analysis.

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

$$\begin{aligned} & E_1(k+1) - E_1(k) \\ \leq \ & -\gamma t_{k+1} \left(L(x_{k+1}, y^*) - L(x^*, y^*) \right) - t_{k+1}(t_{k+1} - 1) \left(L(x_{k+1}, y^*) - L(x_k, y^*) \right) \\ & -\frac{t_{k+1}}{\gamma} \left\langle A^* \left(v_{k+1}^\gamma - \gamma y^* \right), u_{k+1}^\gamma - \gamma x^* \right\rangle - \frac{(1-\gamma)}{\sigma} (t_{k+1} - 1) \|x_{k+1} - x_k\|^2 - \frac{\gamma t_{k+1}}{2L_f} \|\nabla f(z_k) - \nabla f(x^*)\|^2 \\ & -\frac{t_{k+1}}{2\sigma} \big((\gamma - L_f \sigma) t_{k+1} + (1-\gamma) L_f \sigma \big) \|x_{k+1} - z_k\|^2 - \frac{t_{k+1}(t_{k+1} - 1)}{2L_f} \|\nabla f(z_k) - \nabla f(x_k)\|^2, \end{aligned} \tag{3.23}$$

and

$$\begin{aligned} & E_2(k+1) - E_2(k) \\ \leq \ & \gamma t_{k+1} (L(x^*, y_{k+1}) - L(x^*, y^*)) + t_{k+1}(t_{k+1} - 1)(L(x^*, y_{k+1}) - L(x^*, y_k)) \\ & + \frac{t_{k+1}}{\gamma} \left\langle A \left(u_{k+1}^\gamma - \gamma x^* \right), v_{k+1}^\gamma - \gamma y^* \right\rangle - (1-\gamma)(t_{k+1} - 1 + \frac{\gamma}{2}) \|y_{k+1} - y_k\|^2 - \frac{1}{2} \|v_{k+1}^\gamma - v_k^\gamma\|^2. \end{aligned} \tag{3.24}$$

Proof. For a better understanding of the constructed inequalities, we firstly deal with the inequality associated with g. Suppose $\eta_{k+1} \in \partial g(y_{k+1})$. According to the last three lines of (3.11) and (3.16), we have

$$\begin{aligned}
&\frac{1}{2}\|v_{k+1}^{\gamma} - \gamma y^*\|^2 - \frac{1}{2}\|v_k^{\gamma} - \gamma y^*\|^2 \qquad (3.25)\\
= \;& \langle v_{k+1}^{\gamma} - v_k^{\gamma}, v_{k+1}^{\gamma} - \gamma y^*\rangle - \frac{1}{2}\|v_{k+1}^{\gamma} - v_k^{\gamma}\|^2\\
= \;& \langle v_{k+1} - v_k + (\gamma-1)(y_{k+1} - y_k), \gamma(y_{k+1} - y^*) + (t_{k+1} - 1)(y_{k+1} - y_k)\rangle - \frac{1}{2}\|v_{k+1}^{\gamma} - v_k^{\gamma}\|^2\\
= \;& t_{k+1}\langle \frac{1}{\gamma} A u_{k+1}^{\gamma} - \eta_{k+1}, \gamma(y_{k+1} - y^*) + (t_{k+1} - 1)(y_{k+1} - y_k)\rangle\\
& + (\gamma - 1)\gamma\langle y_{k+1} - y_k, y_{k+1} - y^*\rangle + (\gamma - 1)(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 - \frac{1}{2}\|v_{k+1}^{\gamma} - v_k^{\gamma}\|^2\\
= \;& -\gamma t_{k+1}\langle \eta_{k+1} - Ax^*, y_{k+1} - y^*\rangle - t_{k+1}(t_{k+1} - 1)\langle \eta_{k+1} - Ax^*, y_{k+1} - y_k\rangle + \frac{t_{k+1}}{\gamma}\langle A\left(u_{k+1}^{\gamma} - \gamma x^*\right),\\
& v_{k+1}^{\gamma} - \gamma y^*\rangle + (\gamma - 1)\gamma\langle y_{k+1} - y_k, y_{k+1} - y^*\rangle + (\gamma - 1)(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 - \frac{1}{2}\|v_{k+1}^{\gamma} - v_k^{\gamma}\|^2,
\end{aligned}$$

in which,

$$\langle y_{k+1} - y_k, y_{k+1} - y^*\rangle = -\frac{1}{2}\left(\|y_k - y^*\|^2 - \|y_{k+1} - y^*\|^2 - \|y_{k+1} - y_k\|^2\right), \qquad (3.26)$$

and

$$\begin{aligned}
& -\gamma t_{k+1}\langle \eta_{k+1} - Ax^*, y_{k+1} - y^*\rangle - t_{k+1}(t_{k+1} - 1)\langle \eta_{k+1} - Ax^*, y_{k+1} - y_k\rangle\\
\leq \;& -\gamma t_{k+1}\left(g(y_{k+1}) - g(y^*) - \langle Ax^*, y_{k+1} - y^*\rangle\right) - t_{k+1}(t_{k+1} - 1)\left(g(y_{k+1}) - g(y_k) - \langle Ax^*, y_{k+1} - y_k\rangle\right)\\
= \;& \gamma t_{k+1}(L(x^*, y_{k+1}) - L(x^*, y^*)) + t_{k+1}(t_{k+1} - 1)(L(x^*, y_{k+1}) - L(x^*, y_k)), \qquad (3.27)
\end{aligned}$$

where the inequality comes from the convexity of the function $g(\cdot) - \langle Ax^*, \cdot\rangle$. Combining (3.25), (3.26) and (3.27), we have

$$\begin{aligned}
& E_2(k+1) - E_2(k)\\
= \;& \frac{1}{2}\|v_{k+1}^{\gamma} - \gamma y^*\|^2 - \frac{1}{2}\|\lambda_k^{\gamma} - \gamma y^*\|^2 + \frac{\gamma(1-\gamma)}{2}\left(\|y_{k+1} - y^*\|^2 - \|y_k - y^*\|^2\right)\\
\leq \;& \gamma t_{k+1}(L(x^*, y_{k+1}) - L(x^*, y^*)) + t_{k+1}(t_{k+1} - 1)(L(x^*, y_{k+1}) - L(x^*, y_k))\\
& + \frac{t_{k+1}}{\gamma}\langle A\left(u_{k+1}^{\gamma} - \gamma x^*\right), v_{k+1}^{\gamma} - \gamma y^*\rangle - (1-\gamma)(t_{k+1} - 1 + \frac{\gamma}{2})\|y_{k+1} - y_k\|^2 - \frac{1}{2}\|v_{k+1}^{\gamma} - \lambda_k^{\gamma}\|^2,
\end{aligned}$$

which is nothing else than (3.24). On the other hand, in accordance with the coefficients of the primal-dual gap in (3.24), by multiplying (3.18) with γt_{k+1} and (3.19) with $t_{k+1}(t_{k+1} - 1)$, we arrive at

$$\begin{aligned}
& \gamma t_{k+1}\left(L(x_{k+1}, y^*) - L(x^*, y^*)\right) + t_{k+1}(t_{k+1} - 1)\left(L(x_{k+1}, y^*) - L(x_k, y^*)\right) \qquad (3.28)\\
\leq \;& -\frac{t_{k+1}}{\gamma}\langle A^* v_{k+1}^{\gamma} - \gamma y^*, \gamma(x_{k+1} - x^*) + (t_{k+1} - 1)(x_{k+1} - x_k)\rangle\\
& + \frac{t_{k+1}}{\sigma}\langle z_k - x_{k+1}, \gamma(x_{k+1} - x^*) + (t_{k+1} - 1)(x_{k+1} - x_k)\rangle\\
& + \frac{L_f\left(t_{k+1} - 1 + \gamma\right)t_{k+1}}{2}\|x_{k+1} - z_k\|^2 - \frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 - \frac{t_{k+1}(t_{k+1} - 1)}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2\\
= \;& -\frac{t_{k+1}}{\gamma}\langle A^*\left(v_{k+1}^{\gamma} - \gamma y^*\right), u_{k+1}^{\gamma} - \gamma x^*\rangle + \frac{t_{k+1}}{\sigma}\langle z_k - x_{k+1}, u_{k+1}^{\gamma} - \gamma x^*\rangle\\
& + \frac{L_f\left(t_{k+1} - 1 + \gamma\right)t_{k+1}}{2}\|x_{k+1} - z_k\|^2 - \frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 - \frac{t_{k+1}(t_{k+1} - 1)}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2.
\end{aligned}$$

We notice that

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

which we combine with the third line of (3.11) and (3.16) to see that

$$\begin{aligned} & \frac{t_{k+1}}{\sigma}\langle z_k - x_{k+1}, u_{k+1}^\gamma - \gamma x^* \rangle \\ = \ & \frac{1}{\sigma}\left(\langle u_k^\gamma - u_{k+1}^\gamma, u_{k+1}^\gamma - \gamma x^*\rangle - (1-\gamma)(t_{k+1} - 1)\|x_{k+1} - x_k\|^2 + (1-\gamma)\gamma\langle (x_k - x_{k+1}), (x_{k+1} - x^*)\rangle\right) \\ = \ & -\frac{1}{2\sigma}\|u_k^\gamma - u_{k+1}^\gamma\|^2 - \frac{1}{2\sigma}\|u_{k+1}^\gamma - \gamma x^*\|^2 + \frac{1}{2\sigma}\|u_k^\gamma - \gamma x^*\|^2 - \frac{(1-\gamma)(t_{k+1}-1)}{\sigma}\|x_{k+1} - x_k\|^2 \\ & -\frac{(1-\gamma)\gamma}{2\sigma}\|x_k - x_{k+1}\|^2 - \frac{(1-\gamma)\gamma}{2\sigma}\|x_{k+1} - x^*\|^2 + \frac{(1-\gamma)\gamma}{2\sigma}\|x_k - x^*\|^2 \end{aligned} \tag{3.29}$$

holds. We notice that all summands of $E_1(k+1) - E_1(k)$ can be found in (3.29). Combining this obervation with (3.28), we obtain

$$\begin{aligned} & E_1(k+1) - E_1(k) \\ = \ & \frac{1}{2\sigma}\|u_{k+1}^\gamma - \gamma x^*\|^2 - \frac{1}{2\sigma}\|u_k^\gamma - \gamma x^*\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_{k+1} - x^*\|^2 - \frac{\gamma(1-\gamma)}{2\sigma}\|x_k - x^*\|^2 \\ = \ & -\frac{t_{k+1}}{\sigma}\langle z_k - x_{k+1}, u_{k+1}^\gamma - \gamma x^*\rangle - \frac{(1-\gamma)}{\sigma}\left(t_{k+1} - 1 + \frac{\gamma}{2}\right)\|x_{k+1} - x_k\|^2 - \frac{1}{2\sigma}\|u_k^\gamma - u_{k+1}^\gamma\|^2 \\ \leq \ & -\gamma t_{k+1}\left(L(x_{k+1}, y^*) - L(x^*, y^*)\right) - t_{k+1}(t_{k+1} - 1)\left(L(x_{k+1}, y^*) - L(x_k, y^*)\right) \\ & -\frac{t_{k+1}}{\gamma}\langle A^*\left(v_{k+1}^\gamma - \gamma y^*\right), u_{k+1}^\gamma - \gamma x^*\rangle - \frac{(1-\gamma)}{\sigma}\left(t_{k+1} - 1 + \frac{\gamma}{2}\right)\|x_{k+1} - x_k\|^2 - \frac{1}{2\sigma}\|u_k^\gamma - u_{k+1}^\gamma\|^2 \\ & + \frac{L_f\left(t_{k+1} - 1 + \gamma\right)t_{k+1}}{2}\|x_{k+1} - z_k\|^2 - \frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 - \frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2. \end{aligned} \tag{3.30}$$

By (3.7), the third line of (3.11), and (3.17), we deduce

$$\begin{aligned} -\frac{1}{2\sigma}\|u_{k+1}^\gamma - u_k^\gamma\|^2 &= -\frac{1}{2\sigma}\|u_{k+1} - u_k + (\gamma - 1)(x_{k+1} - x_k)\|^2 \\ &= -\frac{\gamma}{2\sigma}\|u_{k+1} - u_k\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_{k+1} - x_k\|^2 - \frac{1-\gamma}{2\sigma}\|u_k - u_{k+1} - x_{k+1} + x_k\|^2 \\ &\leq -\frac{\gamma t_{k+1}^2}{2\sigma}\|x_{k+1} - z_k\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_{k+1} - x_k\|^2. \end{aligned} \tag{3.31}$$

And finally, by substituting (3.31) into (3.30), we obtain (3.23). □

Theorem 3.1. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1 and let $(x^*, y^*) \in \mathbb{S}$. Then the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing and the following inequalities hold*

$$(\gamma - m)\sum_{k\geq 1} t_{k+1}\left(L(x_{k+1}, y^*) - L(x^*, y_{k+1})\right) < +\infty, \tag{3.32}$$

$$\sum_{k\geq 1} t_{k+1}\big((\gamma - L_f\sigma)t_{k+1} + (1-\gamma)L_f\sigma\big)\|x_{k+1} - z_k\|^2 < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1} - 1)\|x_{k+1} - x_k\|^2 < +\infty,$$

$$\sum_{k\geq 1}\|v_{k+1}^\gamma - v_k^\gamma\|^2 < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 < +\infty,$$

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

Proof. From Lemma 3.2, we conclude that

$$
\begin{aligned}
& E(k+1)-E(k) \\
= \;& \left(t_{k+2}(t_{k+2}-1)-t_{k+1}(t_{k+1}-1)\right)\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})\right)+E_1(k+1)-E_1(k) \\
& +t_{k+1}(t_{k+1}-1)\left(\left(L(x_{k+1},y^*)-L(x_k,y^*)\right)-\left(L(x^*,y_{k+1})-L(x^*,y_k)\right)\right)+E_2(k+1)-E_2(k) \\
\leq \;& \left(t_{k+2}^2-t_{k+1}^2-t_{k+2}+(1-\gamma)t_{k+1}\right)\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})\right)-\frac{(1-\gamma)}{\sigma}\left(t_{k+1}-1\right)\|x_{k+1}-x_k\|^2 \\
& -\frac{t_{k+1}}{2\sigma}\left((\gamma-L_f\sigma)t_{k+1}+(1-\gamma)L_f\sigma\right)\|x_{k+1}-z_k\|^2-\frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k)-\nabla f(x^*)\|^2 \\
& -\frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k)-\nabla f(x_k)\|^2-(1-\gamma)(t_{k+1}-1+\frac{\gamma}{2})\|y_{k+1}-y_k\|^2-\frac{1}{2}\|v_{k+1}^\gamma-v_k^\gamma\|^2. \quad (3.33)
\end{aligned}
$$

Due to $\gamma-L_f\sigma\geq 0$ and $0<\gamma\leq 1$ in (3.14), it follows that $(\gamma-L_f\sigma)t_{k+1}+(1-\gamma)L_f\sigma\geq 0$. According to (3.15), we have

$$
t_{k+2}^2-t_{k+1}^2-t_{k+2}+(1-\gamma)t_{k+1}\leq (m-1)t_{k+2}+(1-\gamma)t_{k+1}\leq (m-\gamma)t_{k+1}\leq 0. \quad (3.34)
$$

Thus, all the coefficients in the right-hand side of (3.33) are nonpositive, it follows that the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing for every $k\geq 1$. We complete the proof of (3.32) via Lemma A.4. □

For every $h:=(x,y),h':=(x',y')\in\mathcal{X}\times\mathcal{Y}$, we define the inner product $\langle h,h'\rangle_{\mathcal{W}}=\langle (x,y),(x',y')\rangle_{\mathcal{W}}=\frac{1}{\sigma}\langle x,x'\rangle_{\mathcal{X}}+\langle y,y'\rangle_{\mathcal{Y}}$ and the corresponding norm $\|h\|_{\mathcal{W}}=\sqrt{\frac{1}{\sigma}\|x\|^2+\|y\|^2}$ for all $h:=(x,y)\in\mathcal{X}\times\mathcal{Y}$. Next, we show the boundedness of the sequence generated by Algorithm 1 and an $O\left(1/t_k\right)$ convergence rate for the sequences $\|x_k-x_{k-1}\|$ and $\|y_k-y_{k-1}\|$.

Proposition 3.1. *Let $\{(x_k,y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 1. Suppose that*

$$
\tau:=\inf_{k\geq 1}\frac{t_k}{k}>0. \quad (3.35)
$$

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 the third line of (3.11) and (3.17), 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

$$
\begin{aligned}
E(k) \;=\;& t_{k+1}(t_{k+1}-1)\left(L(x_k,y^*)-L(x^*,y_k)\right)+\frac{\gamma}{2}t_k\|h_k-h^*\|_{\mathcal{W}}^2-\frac{\gamma}{2}(t_k-1)\|h_{k-1}-h^*\|_{\mathcal{W}}^2 \\
& +\frac{1}{2}(t_k-1+\gamma)(t_k-1)\|h_k-h_{k-1}\|_{\mathcal{W}}^2. \quad (3.36)
\end{aligned}
$$

From the fact that $E(k)$ is nonincreasing, for every $k \geq 1$ we get

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

It follows that

$$\frac{\gamma}{2} t_k \|h_k - h^*\|_{\mathcal{W}}^2 \leq \frac{\gamma}{2}(t_k - 1)\|h_{k-1} - h^*\|_{\mathcal{W}}^2 + E(1) \leq \frac{\gamma}{2} t_{k-1}\|h_{k-1} - h^*\|_{\mathcal{W}}^2 + E(1), \tag{3.37}$$

where the second inequality follows from the fact $t_{k+1} - t_k < 1$ in Lemma A.1. After summing up (3.37) from 1 to k, we have for every $k \geq 1$

$$\frac{\gamma}{2} t_k \|h_k - h^*\|_{\mathcal{W}}^2 \leq kE(1) + \frac{\gamma t_0}{2}\|h_0 - h^*\|_{\mathcal{W}}^2,$$

and so

$$\|h_k - h^*\|_{\mathcal{W}}^2 \leq \frac{2k}{\gamma t_k} E(1) + \frac{t_0}{t_k}\|h_0 - h^*\|_{\mathcal{W}}^2 \leq \frac{2}{\gamma\tau}\mathcal{E}_1(x^*, y^*) + \|h_0 - h^*\|_{\mathcal{W}}^2 < +\infty.$$

With this, we conclude that $\{x_k\}_{k\geq 1}$, $\{y_k\}_{k\geq 1}$ are bounded. In addition, by the definitions of u_k^γ and v_k^γ in (3.11), 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 yields that the sequences $\{t_k(x_k - x_{k-1})\}_{k\geq 1}$ and $\{t_k(y_k - y_{k-1})\}_{k\geq 1}$ are also bounded. $\square$

By the definition of v_k, it is obvious that $v_{k+1} - v_k$ is bounded. Note that condition (3.35) in Proposition 3.1, which is crucial for the following analysis of weak convergence of sequence of iterates, has also been proposed in [5, 10]. Actually, we can show that the boundedness of the sequences considered in Proposition 3.1 can also be guaranteed with $\gamma < 1$. By Theorem 3.1, the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing and so $E(k) \leq E(1)$, which yields

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

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 and consequently $\{t_k(x_k - x_{k-1})\}_{k\geq 1}$, $\{t_k(y_k - y_{k-1})\}_{k\geq 1}$ are bounded, in other words, $\|x_k - x_{k-1}\| = O(1/t_k)$ and $\|y_k - y_{k-1}\| = O(1/t_k)$.

Remark 3.1. *If $A = 0$ and $f = 0$ or $g = 0$, then problem (1.1) becomes an unconstrained optimization problem. On the one hand, when $f = 0$, consider $y_0 = y_1$ and a nondecreasing sequence $\{t_k\}_{k\geq 1}$ which satisfies (3.15) for every $k \geq 1$. Then Algorithm 1 reduces to the following proximal scheme:*

$$\begin{aligned} \bar{y}_k &:= y_k + \frac{t_k - 1}{t_{k+1}}(y_k - y_{k-1}), \\ y_{k+1} &:= \arg\min_{y\in\mathcal{Y}} \left\{ g(y) + \frac{1}{2}\|y - \bar{y}_k\|^2 \right\} = prox_g(\bar{y}_k). \end{aligned}$$

On the other hand, if $g = 0$, suppose $x_0 = x_1$ and consider a nondecreasing sequence $\{t_k\}_{k\geq 1}$ which satisfies (3.15) for every $k \geq 1$. We can then reformulate Algorithm 1 as the following accelerated gradient scheme:

$$\begin{aligned} z_k &:= x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}), \\ x_{k+1} &:= z_k - \sigma \nabla f(z_k). \end{aligned}$$

3.2 Fast convergence rate

The aim of this section is to show a fast convergence rate of the primal-dual gap of the bilinearly coupled convex-concave saddle point problem (1.1) as well as a fast convergence rate of the primal-dual gap, the feasibility measure and the objective function value for nonsmooth convex optimization problems with linear constraints.

For every $k \geq 1$, the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing, which yields $E(k) \leq E(1)$. This we have

$$L(x_k, y^*) - L(x^*, y_k) \leq \frac{E(1)}{t_{k+1}(t_{k+1}-1)}. \tag{3.38}$$

Boţ et al. [10] presented several prominent choices for the sequence $\{t_k\}_{k\geq 1}$, i. e. Nesterov's rule [31], the Chambolle-Dossal rule [15], and the Attouch-Cabot rule [1] (here this rule requires $k \geq [\alpha]+1$). These rules all satisfy the conditions (4.5) in Algorithm 1. Next, we consider convergence rates under several classical construction of $\{t_k\}_{k\geq 1}$. First, we consider Nesterov's rule as proposed in Nesterov [31]:

$$t_1 := 1 \text{ and } t_{k+1} := \frac{1+\sqrt{1+4t_k^2}}{2}, \forall k \ \geq 1.$$

This sequence $\{t_k\}_{k\geq 1}$ is strictly increasing. In our case, from (3.15) we have $\frac{m+\sqrt{m^2+4t_k^2}}{2} \geq t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$,, and we recover Nesterov's rule by setting $m := 1$. In addition, $t_k \geq \frac{k+1}{2}$ holds for every $k \geq 1$ and so $\tau \geq \frac{1}{2}$ (see, for instance, Lemma 4.3 in [7]). Since $t_{k+1}(t_{k+1}-1) \geq \frac{(k+2)k}{4} \geq \frac{k^2}{4}$, we arrive at a convergence rate for the primal-dual gap of

$$L(x_k, y^*) - L(x^*, y_k) \leq O\left(\frac{1}{k^2}\right).$$

Second, the Chambolle-Dossal rule [15] is given by

$$t_1 := 1 \text{ and } t_k := 1 + \frac{k-1}{\alpha-1}, \forall k \geq 1,$$

where $\alpha \geq 3$. Let us set $m := \frac{2}{\alpha-1}$, with which we arrive at $\tau = \frac{2}{\alpha-1}$ (see, for instance, Example 3.15 in [10]). With $t_{k+1}(t_{k+1}-1) = \left(1+\frac{k}{\alpha-1}\right)\frac{k}{\alpha-1} \geq \frac{k^2}{(\alpha-1)^2}$, we see that

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

holds. Finally, we have $t_k := \frac{k-1}{\alpha-1}$ in the Attouch-Cabot rule for every $k \geq [\alpha]+1$. We can thus obtain the same convergence rate by a similar analysis.

Consider now the case $f(x) = -\langle x, b\rangle$ with $b \in \mathcal{X}$ fixed. Then problem (1.1) can be reformulated as $-\min_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} -L(x,y) \equiv g(y) - \langle x, A^*y - b\rangle$ and is therefore equivalent to the following linear equality constrained optimization problem:

$$\begin{aligned} &\min_{y\in Y} g(y), \\ &\text{s.t. } A^*y = b. \end{aligned} \tag{3.39}$$

where g is a proper, convex and lower semicontinuous function. Recently, He et al. [22] obtained a convergence rate of $O\left(1/k^2\right)$ for the primal-dual gap, feasibility measure and the objective function value for this type of nonsmooth case. In our case, by $\nabla f = -b$ we can choose $L_f = \gamma$ and $\sigma = 1$ to satisfy (3.14) and obtain the following simplified Algorithm 2 with $\gamma = 1$ that also achieves a convergence rate of $O\left(1/k^2\right)$ for the three general choices of t_k discussed above.

Algorithm 2 Choose $0 < m \leq 1$ and $\{t_k\}_{k\geq 1}$ as a nondecreasing sequence such that $t_1 > 1$ and $t_{k+1}^2 - mt_{k+1} - t_k^2 \leq 0$, $\forall\, k \geq 1$. Given $x_0 = x_1, y_0 = y_1$. For every $k \geq 1$, we set

$$\begin{aligned}
&\xi_k := t_k A x_k - (t_k - 1)\, A x_{k-1} + t_{k+1} A b,\\
&\bar{y}_k := y_k + \frac{(t_k - 1)}{t_{k+1}}(y_k - y_{k-1}),\\
&y_{k+1} := \arg\min_{y\in\mathcal{Y}} \left\{ g(y) + \frac{1}{2}\|y - \bar{y}_k\|^2 + \frac{t_{k+1}^2}{2}\left\| A^*\left(y - \frac{t_{k+1}-1}{t_{k+1}} y_k\right)\right\|^2 + \langle \xi_k, y\rangle \right\},\\
&v_{k+1}^{\gamma} := y_{k+1} + (t_{k+1} - 1)(y_{k+1} - y_k),\\
&x_{k+1} := x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}) - (A^* y_{k+1} - b) - (t_{k+1} - 1)A^*(y_{k+1} - y_k).
\end{aligned}$$

It is not difficult to verify that Algorithm 2 with $t_k = 1 + \frac{k-2}{\alpha-1}$ is different from the Algorithm 1 in [22]. Next, we will show a fast convergence rate of the feasibility measure and the objective function value for problem (3.39).

Theorem 3.2. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 2 and $(x^*, y^*) \in \mathbb{S}$. Then, for every $k \geq 1$, we have*

$$\|A^* y_k - b\| \leq \frac{t_2^2\|A^* y_2 - b\| + 2C}{t_k^2}, \tag{3.40}$$

$$|g(y_k) - g(y^*)| \leq \frac{(t_2^2\|A^* y_2 - b\| + 2C)\|x^*\|}{t_k^2} + \frac{E(1)}{t_{k+1}(t_{k+1} - 1)}. \tag{3.41}$$

where $C := t_1^2\|A^* y_1 - b\| + \sup_{k\geq 1}\|t_{k+1}(x_{k+1} - x_k)\| + \|t_1(x_1 - x_0)\| + \sup_{k\geq 1}\|x_k\| + \|x_0\|$.

Proof. From the last line of Algorithm 2, we obtain

$$x_{k+1} - x_k - \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}) = -t_{k+1}(A^* y_{k+1} - b) + (t_{k+1} - 1)(A^* y_k - b). \tag{3.42}$$

Reformulating (3.42) yields

$$-t_{k+1}(x_{k+1} - x_k) + t_k(x_k - x_{k-1}) - (x_k - x_{k-1}) = \delta_{k+1} - (1 - a_k)\delta_k, \tag{3.43}$$

where $\delta_{k+1} = t_{k+1}^2(A^* y_{k+1} - b)$ and $a_k = 1 - \frac{t_{k+1}^2 - t_{k+1}}{t_k^2}$. By $t_k > 1$ and $t_{k+1}^2 - t_{k+1} \leq t_{k+1}^2 - mt_{k+1} \leq t_k^2$, we have $0 \leq a_k < 1$, for every $k \geq 1$. By telescoping (3.43), we arrive at

$$\begin{aligned}
\left\|\delta_{k+1} + \sum_{i=1}^{k} a_i \delta_i\right\| &= \|\delta_1 - t_{k+1}(x_{k+1} - x_k) + t_1(x_1 - x_0) - (x_k - x_0)\|\\
&\leq C,
\end{aligned}$$

where the last inequality follows from the boundedness of x_k and $t_{k+1}(x_{k+1} - x_k)$. By Lemma A.3, for every $k \geq 1$, we have

$$\|A^* y_{k+1} - b\| \leq \frac{t_1^2\|A^* y_1 - b\| + 2C}{t_{k+1}^2},$$

which yields (3.40). Finally, by (3.38) and $|g(y_k) - g(y^*)| \leq \|L(x_k, y^*) - L(x^*, y_k)\| + \|x^*\|\|(A^* y_k - b)\|$, we arrive at (3.41), which completes the proof. □

We can consider some special cases. With Nesterov's rule [31], the Chambolle-Dossal rule [15], or the Attouch-Cabot rule [1] for the sequence $\{t_k\}_{k\geq 1}$ we obtain a convergence rate of $O\left(1/k^2\right)$ for the primal-dual gap, the feasibility measure and the objective function value. This is an improvement to the convergence rate $o\left(1/k\right)$ derived in [11]. However, note that their $o\left(1/k\right)$ convergence rate for $\|y_k - y_{k-1}\|$ is better than our convergence rate of $O\left(1/k\right)$.

4 A primal-dual algorithm based on explicit discretization

In this section, we suppose that both f and g are continuously differentiable convex functions and ∇f, ∇g are L_f-, L_g-Lipschitz continuous, respectively. When $\alpha \geq 3$ and $\frac{1}{\alpha-1} \leq \theta \leq \frac{1}{2}$, we can obtain the same convergence properties of dynamical system (1.5) for smooth f and g in a way similar to the proof of Theorem 2.1. In the following, we will investigate a numerical algorithm that is derived directly from an explicit discretization of the dynamical system (1.5). Fast gradient algorithms originating from various second order dynamical systems in the spirit of Nesterov's accelerated gradient method have been proposed in [3, 10, 33]. In our approach we will use the time step σ_k defined in (3.2) for the variable x. Suppose $\rho > 0$. For y, we then take the time step

$$\rho_k := \rho\left(1 + \frac{\alpha-1}{k}\right) \text{ for every } k \geq 1. \tag{4.1}$$

We have $y(\sqrt{\rho_k}k) \approx y_{k+1}$, $v(\sqrt{\rho_k}k) \approx v_{k+1}$ and $v^\gamma(\sqrt{\rho_k}k) \approx v^\gamma_{k+1}$. By considering the same construction of a smooth function as in (3.11) for f and using a similar approach for g, we obtain the following explicit discretization of the smooth scheme (3.1):

$$\begin{cases} x_{k+1} = z_k - \sigma\nabla f(z_k) - \frac{\sigma}{\gamma}A^* v^\gamma_{k+1}, \\ z_k := x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}), \\ u^\gamma_{k+1} = \gamma x_{k+1} + (t_{k+1}-1)(x_{k+1} - x_k), \\ y_{k+1} = \lambda_k - \rho\nabla g(\lambda_k) + \frac{\rho}{\gamma}A u^\gamma_{k+1}, \\ \lambda_k := y_k + \frac{t_k - 1}{t_{k+1}}(y_k - y_{k-1}), \\ v^\gamma_{k+1} = \gamma y_{k+1} + (t_{k+1}-1)(y_{k+1} - y_k), \end{cases} \tag{4.2}$$

where we can see that z_k and λ_k are obtained by the application of a proximal operator. By the relations given in (4.2),

$$\begin{aligned} A^* v^\gamma_{k+1} &= \left(t_{k+1} + \gamma - 1\right) A^* y_{k+1} - \left(t_{k+1} - 1\right) A^* y_k \\ &= \left(t_{k+1} + \gamma - 1\right) A^* \left(\lambda_k - \rho\nabla g(\lambda_k)\right) - \left(t_{k+1} - 1\right) A^* y_k + \frac{\rho}{\gamma}\left(t_{k+1} + \gamma - 1\right) A^* A u^\gamma_{k+1} \\ &= \bar{\xi}_k + \frac{\rho}{\gamma}\left(t_{k+1} + \gamma - 1\right)^2 A^* A \left(x_{k+1} - \frac{t_{k+1} - 1}{t_{k+1} + \gamma - 1} x_k\right), \end{aligned} \tag{4.3}$$

where $\bar{\xi}_k = \left(t_{k+1} + \gamma - 1\right) A^* \left(\lambda_k - \rho\nabla g(\lambda_k)\right) - \left(t_{k+1} - 1\right) A^* y_k$. Substituting (4.3) into the first line of (4.2), we arrive at

$$0 = \frac{1}{\sigma}(x_{k+1} - z_k) + \nabla f(z_k) + \frac{1}{\gamma}\bar{\xi}_k + \frac{\rho}{\gamma^2}\left(t_{k+1} + \gamma - 1\right)^2 A^* A \left(x_{k+1} - \frac{t_{k+1} - 1}{t_{k+1} + \gamma - 1} x_k\right).$$

Now we are in a position to present the following algorithm for the smooth case:

Algorithm 3 Choose $\gamma, \sigma, \rho, m > 0$ be such that

$$0 < \max\{m, \sigma L_f, \rho L_g\} \leq \gamma \leq 1. \tag{4.4}$$

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

$$t_1 \geq 1 \text{ and } t_{k+1}^2 - mt_{k+1} - t_k^2 \leq 0, \ \forall\, k \geq 1. \tag{4.5}$$

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

$$\begin{aligned}
z_k &:= x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}),\\
\lambda_k &:= y_k + \frac{t_k - 1}{t_{k+1}}(y_k - y_{k-1}),\\
\bar{\xi}_k &= (t_{k+1} + \gamma - 1)\, A^*\,(\lambda_k - \rho\nabla g(\lambda_k)) - (t_{k+1} - 1)\, A^* y_k,\\
\bar{s}_{k+1} &:= \frac{\rho}{\gamma^2}\,(t_{k+1} + \gamma - 1)^2\,,\\
\bar{x}_k &:= \frac{t_{k+1} - 1}{t_{k+1} + \gamma - 1} x_k,\\
x_{k+1} &:= \arg\min_{x\in X}\left\{\frac{1}{2\sigma}\|x - z_k\|^2 + \frac{s_{k+1}}{2}\|A(x - \bar{x}_k)\|^2 + \langle \nabla f(z_k), x\rangle + \frac{1}{\gamma}\langle \bar{\xi}_k, x\rangle\right\},\\
u_{k+1}^{\gamma} &:= \gamma x_{k+1} + (t_{k+1} - 1)(x_{k+1} - x_k),\\
y_{k+1} &:= \lambda_k - \rho\nabla g(\lambda_k) + \frac{\rho}{\gamma} A u_{k+1}^{\gamma}.
\end{aligned}$$

Compared with Algorithm 1 for the nonsmooth case, the subproblem in Algorithm 3 does not rely on the structure of f or g. Although the choice of $\gamma = 1$ gives a simplified version of Algorithm 3 without affecting the fast convergence rates, we will see that the condition of $\gamma < 1$ is an indispensable part of the weak convergence of iterate (x_k, y_k) to a primal-dual optimal solution. This phenomenon can also be found in corresponding continuous and discrete schemes for unconstrained optimization problems. Fast convergence can be shown for $\alpha \geq 3$, while the weak convergence of the trajectory or the sequence of iterate holds only for $\alpha > 3$. By recalling the definition of γ, it is obvious that $\gamma < 1$ holds only for $\alpha > 3$.

For every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$, we introduce the following energy function:

$$\mathcal{E}(k) = t_{k+1}(t_{k+1} - 1)\,(L(x_k, y^*) - L(x^*, y_k)) + \mathcal{E}_1(k) + \mathcal{E}_2(k),$$

where

$$\mathcal{E}_1(k) = \frac{1}{2\sigma}\|u_k^{\gamma} - \gamma x^*\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_k - x^*\|^2 \text{ and } \mathcal{E}_2(k) = \frac{1}{2\rho}\|v_k^{\gamma} - \gamma y^*\|^2 + \frac{\gamma(1-\gamma)}{2\rho}\|y_k - y^*\|^2.$$

Proposition 4.1. *Let $\{(x_k, y_k)\}_{k\geq 1}$ be the sequence generated by Algorithm 3 and $(x^*, y^*) \in \mathbb{S}$. Then, for every $k \geq 1$, the sequence $\{\mathcal{E}(k)\}_{k\geq 1}$ is nonincreasing and we have the following statements:*

$$(\gamma - m)\sum_{k\geq 1} t_{k+1}\,(L(x_k, y^*) - L(x^*, y_k)) < +\infty, \tag{4.6}$$

$$\sum_{k\geq 1} t_{k+1}\big((\gamma - L_f\sigma)t_{k+1} + (1-\gamma)L_f\sigma\big)\|x_{k+1} - z_k\|^2 < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1} - 1)\|x_{k+1} - x_k\|^2 < +\infty,$$

$$\sum_{k\geq 1} t_{k+1}\big((\gamma - L_g\rho)t_{k+1} + (1-\gamma)L_g\rho\big)\|y_{k+1} - \lambda_k\|^2 < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 < +\infty,$$

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

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

Proof. In a way similar to the proof of (3.23), we can show that

$$
\begin{aligned}
& \mathcal{E}_2(k+1)-\mathcal{E}_2(k) \\
\leq\ & \gamma t_{k+1}\left(L(x^*,y_{k+1})-L(x^*,y^*)\right)+t_{k+1}(t_{k+1}-1)\left(L(x^*,y_{k+1})-L(x^*,y_k)\right) \\
& \frac{t_{k+1}}{\gamma}\left\langle A^*\left(v_{k+1}^{\gamma}-\gamma y^*\right),u_{k+1}^{\gamma}-\gamma x^*\right\rangle-\frac{(1-\gamma)}{\rho}\left(t_{k+1}-1\right)\|y_{k+1}-y_k\|^2-\frac{\gamma t_{k+1}}{2L_g}\|\nabla g(\lambda_k)-\nabla g(y^*)\|^2 \\
& -\frac{t_{k+1}}{2\rho}\left((\gamma-L_g\rho)t_{k+1}+(1-\gamma)L_g\rho\right)\|y_{k+1}-\lambda_k\|^2-\frac{t_{k+1}(t_{k+1}-1)}{2L_g}\|\nabla g(\lambda_k)-\nabla g(y_k)\|^2.
\end{aligned}
$$

Combining with (3.23), we arrive at

$$
\begin{aligned}
& \mathcal{E}(k+1)-\mathcal{E}(k) \\
=\ & \left(t_{k+2}(t_{k+2}-1)-t_{k+1}(t_{k+1}-1)\right)\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})\right)+\mathcal{E}_1(k+1)-\mathcal{E}_1(k) \\
& +t_{k+1}(t_{k+1}-1)\left(\left(L(x_{k+1},y^*)-L(x_k,y^*)\right)-\left(L(x^*,y_{k+1})-L(x^*,y_k)\right)\right)+\mathcal{E}_2(k+1)-\mathcal{E}_2(k) \\
\leq\ & \left(t_{k+2}^2-t_{k+1}^2-t_{k+2}+(1-\gamma)t_{k+1}\right)\left(L(x_{k+1},y^*)-L(x^*,y_{k+1})\right)-\frac{(1-\gamma)}{\sigma}\left(t_{k+1}-1\right)\|x_{k+1}-x_k\|^2 \\
& -\frac{t_{k+1}}{2\sigma}\left((\gamma-L_f\sigma)t_{k+1}+(1-\gamma)L_f\sigma\right)\|x_{k+1}-z_k\|^2-\frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k)-\nabla f(x^*)\|^2 \\
& -\frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k)-\nabla f(x_k)\|^2-\frac{(1-\gamma)}{\rho}\left(t_{k+1}-1\right)\|y_{k+1}-y_k\|^2-\frac{\gamma t_{k+1}}{2L_g}\|\nabla g(\lambda_k)-\nabla g(y^*)\|^2 \\
& -\frac{t_{k+1}}{2\rho}\left((\gamma-L_g\rho)t_{k+1}+(1-\gamma)L_g\rho\right)\|y_{k+1}-\lambda_k\|^2-\frac{t_{k+1}(t_{k+1}-1)}{2L_g}\|\nabla g(\lambda_k)-\nabla g(y_k)\|^2. \qquad (4.7)
\end{aligned}
$$

By the assumptions (4.4), (4.5), and inequality (3.34), all the coefficients in the right-hand side of (4.7) are nonpositive and so $\mathcal{E}(k+1)-\mathcal{E}(k)\leq 0$. We complete the proof of (4.6) via Lemma A.4. □

Remark 4.1. *Compared to the energy function in [10], which is equipped with an auxiliary term $\|x_{k+1}-x_k\|^2$, our energy function $\mathcal{E}(k)$ is exactly the discretization of the continuous energy function* (2.1)*. In addition, we do not use any additional update in the discretization process as in [10], where v_{k+1}^{γ} is replaced by $\tilde{v}_{k+1}^{\gamma}:=v_{k+1}^{\gamma}+(1-\gamma)(\lambda_{k+1}-\lambda_k)$. Only in the nonsmooth case we replaced v_{k+1}^{γ} by $\tilde{v}_{k+1}=v_{k+1}-\frac{\alpha-1}{k+\alpha-1}(v_{k+1}-v_k)$ to obtain an easily implementable iterative scheme.*

However, we can replace v_{k+1}^{γ} with $\tilde{v}_{k+1}^{\gamma}:=v_{k+1}^{\gamma}+(1-\gamma)(\lambda_{k+1}-\lambda_k)$ in the first line of explicit discretization scheme (4.2) *and introduce the following energy function:*

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

for every $(x^*,y^*)\in\mathbb{S}$ and every $k\geq 1$. By an analysis similar to Proposition 4.1 and Proposition 3.9 in [10], we then obtain that the sequence $\{\mathcal{E}_k'\}_{k\geq 1}$ is nonincreasing and we have the following statements:

$$
\begin{aligned}
& (\gamma-m)\sum_{k\geq 1}t_k\left(L(x_k,y^*)-L(x^*,y_k)\right)<+\infty, \\
& \sum_{k\geq 1}t_{k+1}\left(t_{k+1}\left(\gamma-L_f\sigma\right)+(1-\gamma)L_f\sigma\right)\|x_{k+1}-z_k\|^2<+\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1}-1)\|y_{k+1}-y_k\|^2<+\infty, \\
& \sum_{k\geq 1}t_{k+1}\left(t_{k+1}\left(\gamma-L_g\rho\right)+(1-\gamma)L_g\rho\right)\|y_{k+1}-\lambda_k\|^2<+\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1}-1)\|x_{k+1}-x_k\|^2<+\infty, \\
& \sum_{k\geq 1}t_{k+1}\|\nabla f(z_k)-\nabla f(x^*)\|^2<+\infty,\ \sum_{k\geq 1}t_{k+1}(t_{k+1}-1)\|\nabla f(z_k)-\nabla f(x_k)\|^2<+\infty, \\
& (\gamma-\rho L_g(1-\gamma))\sum_{k\geq 1}t_{k+1}\|\nabla g(\lambda_k)-\nabla g(y^*)\|^2<+\infty,\ \sum_{k\geq 1}t_{k+1}(t_{k+1}-1)\|\nabla g(\lambda_k)-\nabla g(y_k)\|^2<+\infty.
\end{aligned}
$$

Compared with (4.6), one can see that the coefficient t_k of term $\sum_{k\geq 1} t_k \left(L(x_k, y^*) - L(x^*, y_k)\right)$ is one step behind the corresponding one in (4.6). Moreover, the condition $\sum_{k\geq 1} t_{k+1}\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 < +\infty$ is guaranteed by requiring $\gamma - \rho L_g(1-\gamma) > 0$.

Since $\{\mathcal{E}(k)\}_{k\geq 1}$ is nonincreasing for every $k \geq 1$, we again arrive at $L(x_k, y^*) - L(x^*, y_k) \leq \frac{E(1)}{t_{k+1}(t_{k+1}-1)}$. We notice that the classical three schemes for t_k, i. e. Nesterov's rule [31], the Chambolle-Dossal rule [15], and the Attouch-Cabot rule [1] (with the additional requirement $k \geq [\alpha] + 1$) all still satisfy the conditions (4.5) in Algorithm 3. As such, a convergence rate of $O\left(1/k^2\right)$ for the primal-dual gap follows in the same way as before.

4.1 Convergence of the iterates

In this section, we provide some important estimates which will be useful for the proof of the convergence of the sequence $\{(x_k, y_k)\}_{k\geq 1}$ generated by Algorithm 3 for the smooth case.

Proposition 4.2. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3 and $(x^*, y^*) \in \mathbb{S}$. Assume that $0 < m < \gamma \leq 1$ holds. Then we have the following statements:*

$$\sum_{k\geq 1} t_k \left\|A^*\left(y_k - y^*\right)\right\|^2 < +\infty, \tag{4.8}$$

$$\sum_{k\geq 1} t_{k+1}(t_{k+1}-1)^2 \left\|A^*\left(y_{k+1} - y_k\right)\right\|^2 < +\infty, \tag{4.9}$$

$$\sum_{k\geq 1} t_k \left\|A\left(x_k - x^*\right)\right\|^2 < +\infty, \tag{4.10}$$

$$\sum_{k\geq 1} t_{k+1}(t_{k+1}-1)^2 \left\|A\left(x_{k+1} - x_k\right)\right\|^2 < +\infty. \tag{4.11}$$

Moreover, there exists an $M > 0$ such that

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

Proof. From the first line of (4.2), we have

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

By Proposition 4.1 and $t_k > 0$ for every $k \geq 1$, it follows that

$$\sum_{k\geq 1} t_{k+1}\left\|A^*\left(\frac{1}{\gamma}v_{k+1}^{\gamma} - y^*\right)\right\|^2 \leq \frac{2}{\sigma^2}\sum_{k\geq 1} t_{k+1}\|z_k - x_{k+1}\|^2 + 2\sum_{k\geq 1} t_{k+1}\|\nabla f(z_k) - \nabla f(x^*)\|^2 < +\infty.$$

According to the last line of (4.2) and (3.17), for every $k \geq 1$ we have

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

which yields

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

But by (4.5), we see that

$$
\begin{aligned}
\frac{t_{k+1}(t_{k+1}-1)}{\gamma} - t_k\left(1+\frac{t_k-1}{\gamma}\right) &= \frac{1}{\gamma}\left(t_{k+1}^2 - t_{k+1} - t_k^2 + t_k - \gamma t_k\right) \\
&\leq \frac{1}{\gamma}\left((m-1)t_{k+1} - (\gamma-1)t_k\right) \\
&\leq \left(\frac{m}{\gamma}-1\right)t_k,
\end{aligned} \tag{4.12}
$$

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

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

Let us set

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

for every $k\geq 1$. By employing Lemma A.4 and $m<\gamma$, we obtain

$$
\sum_{k\geq 1} t_k\left\|A^*\left(y_k-y^*\right)\right\|^2 < +\infty, \quad \sum_{k\geq 1} t_{k+1}(t_{k+1}-1)^2\left\|A^*\left(y_{k+1}-y_k\right)\right\|^2 < +\infty,
$$

and the sequence $\left\{t_k\left(1+\frac{t_k-1}{\gamma}\right)\left\|A^*\left(y_k-y^*\right)\right\|^2\right\}$ is convergent and bounded. Similarly, we obtain (4.10), (4.11) and the fact that $t_k\left(1+\frac{1}{\gamma}(t_k-1)\right)\left\|A^*\left(x_k-x^*\right)\right\|^2$ is convergent. Since $t_k \leq \left(1+\frac{1}{\gamma}(t_k-1)\right)$ for every $k\geq 1$, we arrive at

$$
t_k^2\left\|A^*\left(y_k-y^*\right)\right\|^2 \leq t_k\left(1+\frac{1}{\gamma}(t_k-1)\right)\left\|A^*\left(y_k-y^*\right)\right\|^2 \leq M^2,
$$

where $M>0$. Then, $\left\|A^*\left(y_k-y^*\right)\right\| \leq \frac{M}{t_k}$. Similarly, we obtain $\left\|A\left(x_k-x^*\right)\right\| \leq \frac{M}{t_k}$. $\square$

Next, we will show the weak convergence of the sequence $\{(x_k,y_k)\}_{k\geq 0}$.

Lemma 4.1. *Let $\{(x_k,y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3, $(x^*,y^*)\in\mathbb{S}$ and $0<m<\gamma<1$. Then, the limit $\lim_{k\to+\infty}\frac{1}{\sigma}\|x_k-x^*\|^2+\frac{1}{\rho}\|y_k-y^*\|^2$ exists.*

Proof. Denote $a_k := \frac{\gamma}{2\sigma}\|x_k - x^*\|^2 + \frac{\gamma}{2\rho}\|y_k - y^*\|^2$. By considering a reformulation of $\mathcal{E}(k)$ similar to (3.36) and the fact that $\mathcal{E}(k+1) \leq \mathcal{E}(k)$ for every $k \geq 1$, we obtain

$$\begin{aligned} & t_{k+2}(t_{k+2}-1)\left(L(x_{k+1}, y^*) - L(x^*, y_{k+1})\right) + t_{k+1}a_{k+1} - (t_{k+1}-1)a_k \\ & +\frac{1}{2}(t_{k+1} - 1 + \gamma)(t_{k+1}-1)\left(\frac{1}{\sigma}\|x_{k+1} - x_k\|^2 + \frac{1}{\rho}\|y_{k+1} - y_k\|^2\right) \\ \leq\ & t_{k+1}(t_{k+1}-1)\left(L(x_k, y^*) - L(x^*, y_k)\right) + t_k a_k - (t_k - 1)a_{k-1} \\ & +\frac{1}{2}(t_k - 1 + \gamma)(t_k - 1)\left(\frac{1}{\sigma}\|x_k - x_{k-1}\|^2 + \frac{1}{\rho}\|y_k - y_{k-1}\|^2\right) \end{aligned} \tag{4.13}$$

and so

$$\begin{aligned} & t_{k+1}\left(\frac{t_{k+2}(t_{k+2}-1)}{t_{k+1}}\left(L(x_{k+1}, y^*) - L(x^*, y_{k+1})\right) + \frac{1}{2}(t_{k+1} - 1 + \gamma)\left(\frac{1}{\sigma}\|x_{k+1} - x_k\|^2 \right.\right. \\ & \left.\left. +\frac{1}{\rho}\|y_{k+1} - y_k\|^2\right)\right) + t_{k+1}\left(a_{k+1} - a_k\right) \\ \leq\ & (t_k - 1)\left(\frac{t_{k+1}(t_{k+1}-1)}{t_k}\left(L(x_k, y^*) - L(x^*, y_k)\right) + \frac{1}{2}(t_k - 1 + \gamma)\left(\frac{1}{\sigma}\|x_k - x_{k-1}\|^2 \right.\right. \\ & \left.\left. +\frac{1}{\rho}\|y_k - y_{k-1}\|^2\right)\right) + (t_k - 1)\left(a_k - a_{k-1}\right) \\ & +\frac{t_{k+1}(t_{k+1}-1)}{t_k}\left(L(x_k, y^*) - L(x^*, y_k)\right) + \frac{1}{2}(t_{k+1} - 1 + \gamma)\left(\frac{1}{\sigma}\|x_{k+1} - x_k\|^2 + \frac{1}{\rho}\|y_{k+1} - y_k\|^2\right) \\ \leq\ & (t_k - 1)\left(\frac{t_{k+1}(t_{k+1}-1)}{t_k}\left(L(x_k, y^*) - L(x^*, y_k)\right) + \frac{1}{2}(t_k - 1 + \gamma)\left(\frac{1}{\sigma}\|x_k - x_{k-1}\|^2 \right.\right. \\ & \left.\left. +\frac{1}{\rho}\|y_k - y_{k-1}\|^2\right)\right) + (t_k - 1)\left(a_k - a_{k-1}\right) \\ & +t_{k+1}\left(L(x_k, y^*) - L(x^*, y_k)\right) + \frac{1}{2}(t_{k+1} - 1 + \gamma)\left(\frac{1}{\sigma}\|x_{k+1} - x_k\|^2 + \frac{1}{\rho}\|y_{k+1} - y_k\|^2\right), \end{aligned} \tag{4.14}$$

where the last inequality follows from $t_{k+1} - 1 \leq t_k$, whicl in turn holds due to Lemma A.1. Denote

$$\begin{aligned} \beta_k &:= \frac{t_{k+1}(t_{k+1}-1)}{t_k}\left(L(x_k, y^*) - L(x^*, y_k)\right) + \frac{1}{2}(t_k - 1 + \gamma)\left(\frac{1}{\sigma}\|x_k - x_{k-1}\|^2 \right. \\ & \quad \left. +\frac{1}{\rho}\|y_k - y_{k-1}\|^2\right) + (a_k - a_{k-1}), \\ d_k &:= t_{k+1}\left(L(x_k, y^*) - L(x^*, y_k)\right) + \frac{1}{2}(t_{k+1} - 1 + \gamma)\left(\frac{1}{\sigma}\|x_{k+1} - x_k\|^2 + \frac{1}{\rho}\|y_{k+1} - y_k\|^2\right) \geq 0. \end{aligned}$$

From these definitions, it is obvious that $a_{k+1} \leq a_k + \beta_{k+1}$. By (4.14) we arrive at $t_{k+1}\beta_{k+1} \leq (t_k - 1)\beta_k + d_k$. In addition, from Proposition 4.1, we notice that $\sum_{k\geq 1} d_k < +\infty$ if $0 < m < \gamma < 1$. Thus, by Lemma A.5, we conclude that $\{a_k\}$ is convergent which completes the proof. ☐

Theorem 4.1. *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3 and $(x^*, y^*) \in \mathbb{S}$. Assume further that $\{t_k\}_{k\geq 1}$ is chosen to satisfy* (3.35) *and $0 < m < \gamma < 1$ holds. Then, we have*

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

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 4.1, 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 (3.35), 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

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

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 (3.35) and (4.8), we obtain $\|A^*(y_k-y^*)\|=o\left(1/\sqrt{k}\right)$ which yields $\|\nabla_x L(x,y)\|=o\left(1/\sqrt{k}\right)$. Similarly, we have $\|\nabla_y L(x,y)\|=o\left(1/\sqrt{k}\right)$. This completes the proof. □

Theorem 4.2. *Let $\{(x_k,y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3 and $(x^*,y^*)\in\mathbb{S}$. Assume further that $\{t_k\}_{k\geq 1}$ is chosen to satisfy (3.35) and that $0<m<\gamma<1$ holds. Then, the sequence $\{(x_k,y_k)\}_{k\geq 1}$ weakly converges to a primal-dual optimal solution of the bilinearly coupled saddle point problem (1.1).*

Proof. Suppose $(\bar{x},\bar{y})$ is an arbitrary weak sequential cluster point of (x_k,y_k) as $k\to+\infty$, and so there exists a sequence (x_{k_n},y_{k_n}) such that $(x_{k_n},y_{k_n})\to(\bar{x},\bar{y})$ as $n\to+\infty$. By Theorem 4.1, we get

$$\nabla f(x_{k_n})+A^*y_{k_n}\to 0 \text{ and } \nabla g(y_{k_n})-Ax_{k_n}\to 0, \text{ as } n\to+\infty,$$

respectively. Since the graph of the operator $\mathcal{T}_L$ in (1.3) is sequentially closed (see Proposition 20.38 from [6]), we conclude that

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

which means that $(\bar{x},\bar{y})\in\mathbb{S}$. From Lemma 4.1 we notice that the limit $\lim_{k\to+\infty}\frac{1}{\sigma}\|x_k-x^*\|^2+\frac{1}{\rho}\|y_k-y^*\|^2$ exists for every $(x^*,y^*)\in\mathbb{S}$. With this, we complete the proof via Opial's Lemma as given in Lemma A.2. □

Remark 4.2. *When we chose the Chambolle-Dossal rule or the Attouch-Cabot rule for the sequence $\{t_k\}_{k\geq 1}$ with $\alpha>3$, $m=\frac{1}{\alpha-2}<\gamma<1$, $\sigma\leq\frac{1}{L_f}$ and $\rho\leq\frac{1}{L_g}$, then by Theorem 4.2, the sequence $\{(x_k,y_k)\}_{k\geq 0}$ generated by Algorithm 3 converges weakly to a primal-dual optimal solution of problem (1.1). If the sequence $\{t_k\}_{k\geq 1}$ is chosen to take the Nesterov rule with $m=\gamma=1$, although the fast convergence rate still holds, however, we can not obtain the convergence of the sequence since the conditions in Theorem 4.2 require $m<\gamma<1$.*

5 Conclusion and perspectives

As a brief review of the main result, the inertial primal-dual dynamics (1.5) allow us to construct two first-order algorithms for a bilinearly coupled saddle point problem. These algorithms not only maintain the fast convergence rate for primal-dual values as found in several classical accelerated algorithms, but also possess additional exciting properties, such as the convergence of gradients towards zero, and global convergence of the iterates to optimal saddle points. By recalling the main ideas of the proof of (2.6), we 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 both f and g are two convex lower semicontinuous and proper function. Additionally, it would be worth considering (1.5) in a more general context, which includes situations involving general viscous damping, Hessian-driven damping, and temporal rescaling.

Appendix

Lemma A.1. Let $0 < m \leq 1$ and $\{t_k\}_{k\geq 1}$ a nondecreasing sequence fulfilling

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

Then for every $k \geq 1$ we have that $t_{k+1} - t_k < m \leq 1$ holds.

Proof. Let $k \geq 1$. From the assumption, we get

$$t_{k+1} \leq \frac{m + \sqrt{m^2 + 4t_k^2}}{2} = \frac{m + \sqrt{(m + 2t_k)^2 - 4mt_k}}{2} < m + t_k,$$

and so $t_{k+1} - t_k < m \leq 1$. □

Opial's Lemma which is used for the proof of the weak convergence of the trajectory of dynamical system to a primal-dual solution of the original optimization problem has received much popularity recently. A discrete version of the lemma can be found in Theorem 5.5 of [6].

Lemma A.2. Let S be a nonempty subset of $\mathcal{X}$ and $\{x_k\}_{k\geq 1}$ be a sequence in $\mathcal{X}$. Assume that

(i) for every $y^* \in S$, the limit $\lim_{k\to+\infty} \|x_k - x^*\|$ exists;

(ii) every weak sequential cluster point of the trajectory $\{x_k\}_{k\geq 1}$ as $k \to +\infty$ belongs to S.

Then $\{x_k\}_{k\geq 1}$ converges weakly to a point in S as $k \to +\infty$.

Lemma A.3. Let $\{g_k\}_{k\geq k_0}$ be a sequence in $\mathcal{X}$ and $\{a_k\}_{k\geq k_0}$ be a sequence in $[0, 1)$, where $k_0 \geq 1$. For every $k \geq k_0$, assume

$$\left\| g_{k+1} + \sum_{j=k_0}^{k} a_j g_j \right\| \leq C,$$

then,

$$\sup_{k\geq k_0} \|g_k\| \leq \|g_{k_0}\| + 2C.$$

Proof. The proof is similar to the one of Lemma 4 in [21], so we omit it here. □

The following lemma can be found as Lemma 1.1 of [10]:

Lemma A.4. Let $\{a_k\}$, $\{b_k\}$ and $\{d_k\}$ be sequences of real numbers for every $k \geq 1$. Assume that $\{a_k\}$ is bounded from below, and $\{b_k\}$ and $\{d_k\}$ are nonnegative such that $\sum_{k\geq 1} d_k < +\infty$. Suppoose further that for every $k \geq 1$ it holds

$$a_{k+1} \leq a_k - b_k + d_k.$$

Then the following statements are true

(1) the sequence $\{b_k\}$ is summable, namely $\sum_{k\geq 1} b_k < +\infty$;

(2) the sequence $\{a_k\}$ is convergent.

The following lemma can be founded as Lemma 4.1 of [10]:

Lemma A.5. Let $\{\theta_k\}_{k\geq 1}$, $\{a_k\}_{k\geq 1}$ and $\{t_k\}_{k\geq 1}$ be real sequences such that $\{a_k\}_{k\geq 1}$ is bounded from below and $\{t_k\}_{k\geq 1}$ is nondecreasing and bounded from below by 1. Let $\{d_k\}_{k\geq 1}$ be a nonnegative sequence such that for every $k \geq 1$

$$
\begin{aligned}
&a_{k+1} \leq a_k + \theta_k, \\
&t_{k+1}\theta_{k+1} \leq (t_k - 1)\theta_k + d_k.
\end{aligned}
$$

If $\sum_{k\geq 1} d_k < +\infty$, then the sequence $\{a_k\}_{k\geq 1}$ is convergent.

References

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

[2] 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.

[3] 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.

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

[6] H.H. Bauschke and P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer, New York, 2017.

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

[8] A. Böhm, M. Sedlmayer, E.R. Csetnek and R.I. Boţ, Two steps at a time—taking GAN training in stride with Tseng's method, SIAM J. Math. Data Sci., 4(2022), pp. 750-771.

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

[10] R.I. Boţ, E.R. Csetnek, and D.K. Nguyen, Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates, Math. Program., 200(2023), pp. 147-197.

[11] R.I. Boţ, E.R. Csetnek and M. Sedlmayer, An accelerated minimax algorithm for convex-concave saddle point problems with nonsmooth coupling function, Comput. Optim. Appl., 86(2023), pp. 925–966.

[12] R.I. Boţ, M. Sedlmayer and P.T. Vuong, A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs, J. Mach. Learn Res., 24(2023), pp. 1-37.

[13] 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.

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

[15] 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.

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

[17] 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.

[18] 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.

[19] E. Esser, X. Zhang, and T. F. Chan, A general framework for a class of first order primal- dual algorithms for convex optimization in imaging science, SIAM J. Imaging Sci., 3 (2010), pp. 1015–1046.

[20] 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.

[21] 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), p. 110547.

[22] X. He, R. Hu, and Y.P. Fang, Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems, Numer. Algorithms, 90(2022), pp. 1669-1690.

[23] X. He, N. Huang, and Y.P. Fang, Non-ergodic convergence rates of first-order primal-dual algorithms for saddle point problems, arXiv:2311.11274, (2023).

[24] F. Jiang, X. Cai, Z. Wu and D. Han, Approximate first-order primal-dual algorithms for saddle point problems, Math. Comput., 90(2021), pp. 1227–1262.

[25] G. M. Korpelevich, The extragradient method for finding saddle points and other problems, Ekon. Mat. Metody, 12 (1976), pp. 747–756.

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

[27] 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:2210.17550, (2022).

[28] P.L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.

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

[30] A. Mokhtari, A.E. Ozdaglar and S. Pattathil, Convergence rate of $O(1/k)$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems, SIAM J. Optimiz., 30(2020), pp. 3230-3251.

[31] 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.

[32] 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.

[33] 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.

[34] 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.