Computational Optimization and Applications (2025) 90:151–192
https://doi.org/10.1007/s10589-024-00626-z

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

Ke-wei Ding[1] · **Jörg Fliege**[2] · **Phan Tu Vuong**[2]

Received: 3 April 2024 / Accepted: 5 November 2024 / Published online: 18 November 2024

Abstract

This paper studies the convergence rate of a second-order dynamical system associated with a nonsmooth bilinearly coupled convex-concave saddle point problem, as well as the convergence rate of its corresponding discretizations. We derive the convergence rate of the primal-dual gap for the second-order dynamical system with asymptotically vanishing damping term. Based on an implicit discretization scheme, we propose a primal-dual algorithm and provide a non-ergodic convergence rate under a general setting for the inertial parameters when one objective function is continuously differentiable and convex and the other is a proper, convex and lower semicontinuous function. For this algorithm we derive a $O\left(1/k^2\right)$ convergence rate under three classical rules proposed by Nesterov, Chambolle-Dossal and Attouch-Cabot without assuming strong convexity, which is compatible with the results of the continuous-time dynamic system. For the case when both objective functions are continuously differentiable and convex, we further present a primal-dual algorithm based on an explicit discretization. We provide a corresponding non-ergodic convergence rate for this algorithm and show that the sequence of iterates generated weakly converges to a primal-dual optimal solution. Finally, we present numerical experiments that indicate the superior numerical performance of both algorithms.

Keywords Saddle point problem · Primal-dual dynamical system · Convergence rate · Numerical algorithm · Nesterov's accelerated gradient method · Iterates convergence

✉ Jörg Fliege
J.Fliege@soton.ac.uk

Ke-wei Ding
bluedkw@163.com

Phan Tu Vuong
T.V.Phan@soton.ac.uk

1 School of Mathematics, Southwest Minzu University, Chengdu 610041, Sichuan, People's Republic of China

2 Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, UK

Mathematics Subject Classification 37N40 · 49J35 · 65K10 · 68Q25 · 90C25

1 Introduction

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

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

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

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

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

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

This saddle point problem is naturally associated with the convex optimization problem

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

where $f:\mathcal{X}\to\mathbb{R}$ is a continuously differentiable convex function, $g:\mathcal{Y}\to\mathbb{R}\cup\{+\infty\}$ 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. As such, saddle point problems arise regularly in determining primal-dual pairs of constrained convex optimisation problems. Recently such saddle points have also been studied widely due to their occurrence in many relevant and challenging applications in the field of imaging processing [14, 17], reinforcement learning [19, 36] and generative adversarial networks [9, 13].

We denote by $\mathbb{S}$ the set of saddle points of problem (1.1). We assume that problem (1.1) has at least one solution (x^*,y^*) which also satisfies the 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. Now define the operator $\mathcal{T}_L:\mathcal{X}\times\mathcal{Y}\to\mathcal{X}\times\mathcal{Y}$ by

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

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 maximal monotone, see Corollary 20.28 in [7]. Thus, $\mathbb{S}$ can be interpreted as the set of zeros of the maximal monotone operator $\mathcal{T}_L$ and so $\mathbb{S}$ is closed and convex.

Let us recall some significant developments regarding primal-dual algorithms for the saddle point problem (1.1). Consider first the case of f and g being proper, convex, and lower semicontinuous. With their celebrated first-order primal-dual algorithm, Chambolle and Pock [14] provided an ergodic convergence rate of $O(1/k)$ for the primal-dual gap of problem (1.1). They also showed that their algorithm has strong connections with other well-known methods, such as the extra-gradient method [27], the Douglas-Rachford splitting method [30] and the preconditioned ADMM method [20]. When either f or g is strongly convex (the partially strongly convex case), they also proved an ergodic convergence rate of $O\left(1/k^2\right)$ for an accelerated version of their primal-dual algorithm. In addition, an ergodic linear convergence rate has been provided when both f and g are strongly convex (the strongly convex case). By employing the Bregman distance, Chambolle and Pock [15] later established 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 [14], He et al. [21] proposed a generalized primal-dual algorithm whilst relaxing the condition for ensuring convergence, and obtained a convergence rate of $O(1/k)$ in both the ergodic and pointwise sense. Under the assumption that f is a convex and Fréchet differentiable function with L_f-Lipschitz continuous gradient and g being proper, convex and lower semicontinuous, Chen et al. [17] provided 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. [26] 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. [24] showed a non-ergodic convergence rate of $O(1/k)$ 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. [28] 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. [36] developed a lifted primal-dual first order algorithm and showed a lower complexity bound under the assumption that f and g are both strongly convex smooth functions. Further results regarding (1.1) can be found in [14, 15, 17, 18, 21] 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 solution algorithms for various optimization problems. A decisive step was taken by Su et al. in [35], where, for the minimization of a continuously differentiable convex function $\Phi : \mathcal{X} \to \mathbb{R}$, the authors considered the following second-order inertial dynamic with asymptotic vanishing viscous damping:

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

For the case $\alpha = 3$, the authors successfully link the inertial dynamic (1.4) with the accelerated gradient method of Nesterov [8, 33]. 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 [31] 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 an 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 the development of the corresponding accelerated alternating direction method of multipliers (ADMM) via temporal discretization. For the same type of problem, Boţ and Nguyen [10] discussed the convergence behavior of the primal-dual gap, the feasibility measure, the objective function value and the trajectory of a second-order dynamical system with asymptotically vanishing damping term. Likewise, Boţ et al. [11] recently presented a corresponding numerical optimization algorithm originating from the second-order dynamical system described in [10]. 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 is thus natural to employ a dynamical system framework to study bilinearly coupled convex-concave saddle point problems. Li et al. [29] 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. Zeng et al. [38] presented convergence rates for an inertial primal-dual dynamical system with asymptotic vanishing damping. Recently, He et al. [25] provided convergence rates for a general inertial primal-dual dynamical system with damping, scaling and extrapolation coefficients. Motivated by the works described above, we consider here 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 $(x_0, y_0), (\dot{x}_0, \dot{y}_0) \in \mathcal{X} \times \mathcal{Y}$. Compared with the dynamical systems mentioned in [25, 38], the second line of system (1.5) is a differential inclusion problem due to the nonsmoothness of g. If g is differentiable, we can make minor adjustments accordingly and replace 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):

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

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:

- We provide 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 implicit discretization, for the case of f being a continuously differentiable convex function with Lipschitz continuous gradient and g being a proper, convex and lower semicontinuous function. We consider a general setting for the inertial parameters which covers three classical rules proposed by Nesterov [33], Chambolle-Dossal [16] 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 $O\left(L_f/k^2 + \|A\|/k\right)$ rate derived in [17]. In contrast to [15, 24, 26], we obtain the rate $O\left(1/k^2\right)$ without the assumption of strong convexity.
- We develop a primal-dual algorithm based on explicit discretization, for the case 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 [32]. 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 [16] 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. [11]. 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 a primal-dual algorithm for the case when when both f and g are smooth, and show the convergence of the sequence of iterates. In Section 5, we test and compare our algorithms with other relevant algorithms from the literature, before we summarize our results in Section 6.

2 The primal-dual dynamical system

In this section, we assume that 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, 25, 34, 35, 38], and choosing an appropriate one is crucial. Motivated by the energy function introduced in Attouch et al. [3] and Boţ and Nguyen [10], we define the 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 the dynamical system (1.5) and $(x^*, y^*) \in \mathbb{S}$. Suppose $\alpha \geq 3$ and $1/(\alpha - 1) \leq \theta \leq 1/2$. Then we have*

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

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

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

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

Proof By differentiating $\mathcal{E}_i(t)$ with respect to t and employing a similar discussion to Remark 4.7 in [5], we have

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

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

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

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

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

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

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

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

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

holds, which yields $\|\dot{x}(t)\| = O(1/t)$ under the appropriate assumption. Similarly we have $\|\dot{y}(t)\| = O(1/t)$. □

We have thus 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 corresponding $O\left(1/k^2\right)$ convergence rates, which is compatible with the results in the continuous case.

3 A fast primal-dual algorithm based on implicit discretization

In this section, we assume that f is a continuously differentiable convex function with L_f-Lipschitz continuous gradient, and that g is a proper, convex and lower semicontinuous function. We will investigate the convergence properties of a numerical algorithm which is derived from the implicit discretization of the dynamical system (1.5), i. e. the nonergodic convergence rate for the primal dual gap, and the convergence rate of some infinite series of iterates.

3.1 The fast scheme: from continuous to discrete

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. [11], and He et al. [22]. Let

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

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:

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

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 provide

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

where the choice of z_k is made as follows. 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:

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

We note that g is nonsmooth, for every $k \geq 1$ and time $t := k$ it follows that $y(t) = y_k$, $v(t) = v_k$ and $v^\gamma(t) = v_k^\gamma$. We now consider the following discretization scheme for the last three lines in (3.1):

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

where we replace v_{k+1} with a suitable term $\tilde{v}_{k+1}$ to obtain an executable iterative scheme, which is explained below. This approach also has been taken by Boţ et al. [11], where 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_{k+1}^{γ}, while here we want to choose a suitable $\tilde{v}_{k+1}$ such that $\tilde{v}_{k+1} - v_{k+1} \to 0$ as $k \to +\infty$. For this purpose, 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 $k \geq 1$. We will see from Proposition 3.1 below that $v_{k+1} - v_k$ is bounded under some mild conditions. With this option, we can reformulate the first line in (3.9) to

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

Following Attouch and Cabot [1] and Boţ et al. [11], 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, by combining (3.8), (3.9), (3.10) and the definition of t_k, we arrive at the following discretization scheme of the 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

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

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

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

After rearranging the order in which these sequences are updated, we are finally in a position to propose our fast primal-dual algorithm for smooth f and nonsmooth g.

3.2 A fast algorithm and its convergence analysis

In this section we present the following fast primal-dual algorithm (for short, FPDA1) based on system (3.11). We will prove that it exhibits an $O\left(1/k^2\right)$ convergence rate for the primal-dual gap under three classical rules for the choice of time steps t_k proposed by Nesterov, Chambolle-Dossal and Attouch-Cabot, without any assumption on strong convexity.

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

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

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 or the corresponding accelerated FISTA scheme [6, 8]. We note that we obtain a simplified version of Algorithm 1 by setting $\gamma := 1$, and by the analysis below we will see that this version enjoys the same convergence rate as for the case $\gamma < 1$. In what follows we consider the general case $\gamma \le 1$, since some results for $\gamma \le 1$ will be crucial in the analysis of Algorithm 3 described in the following section.

Remark 3.1 When $A = 0$ and $f = 0$, consider $y_0 = y_1$ and a nondecreasing sequence $\{t_k\}_{k\ge1}$ which satisfies (3.15) for every $k \ge 1$. 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 $A = 0$ and $g = 0$, suppose $x_0 = x_1$ and consider a nondecreasing sequence $\{t_k\}_{k\ge1}$ which satisfies (3.15) for every $k \ge 1$. We can then reformulate Algorithm 1 as the following accelerated gradient scheme:

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

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 $\{(x_k, y_k)\}_{k\ge0}$ be the sequence generated by Algorithm 1 and let $(x^*, y^*) \in \mathbb{S}$. Then, for every $k \ge 1$ the following two inequalities hold:*

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

and

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

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

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

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

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

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

$$f(x_{k+1}) + \langle A(x_{k+1} - x_k), y^*\rangle - f(x_k) \leq -\frac{1}{\gamma}\left\langle A^*\left(v_{k+1}^{\gamma} - \gamma y^*\right), x_{k+1} - x_k\right\rangle$$

$$+\frac{1}{\sigma}\langle z_k - x_{k+1}, x_{k+1} - x_k\rangle + \frac{L_f}{2}\|x_{k+1} - z_k\|^2 - \frac{1}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2. \tag{3.22}$$

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 now the following energy function:

$$E(k) := t_{k+1}(t_{k+1} - 1)\left(L(x_k, y^*) - L(x^*, y_k)\right) + E_1(k) + E_2(k), \tag{3.23}$$

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 $k \geq 1$. To some extent, energy function (3.23) can be viewed as a discretization of (2.1). Next we show two important inequalities for $E_1(k)$ and $E_2(k)$ which will play a significant role in the analysis that follows.

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

$$\begin{aligned}
&E_1(k+1) - E_1(k) \\
&\leq -\gamma t_{k+1}\left(L(x_{k+1}, y^*) - L(x^*, y^*)\right) - t_{k+1}(t_{k+1} - 1)\left(L(x_{k+1}, y^*) - L(x_k, y^*)\right) \\
&\quad - \frac{t_{k+1}}{\gamma}\left\langle A^*\left(v_{k+1}^\gamma - \gamma y^*\right), u_{k+1}^\gamma - \gamma x^*\right\rangle - \frac{(1-\gamma)}{\sigma}(t_{k+1} - 1)\|x_{k+1} - x_k\|^2 \\
&\quad - \frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 \\
&\quad - \frac{t_{k+1}}{2\sigma}\left((\gamma - L_f\sigma)t_{k+1} + (1-\gamma)L_f\sigma\right)\|x_{k+1} - z_k\|^2 \\
&\quad - \frac{t_{k+1}(t_{k+1} - 1)}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2,
\end{aligned} \tag{3.24}$$

and

$$\begin{aligned}
&E_2(k+1) - E_2(k) \\
&\leq \gamma t_{k+1}(L(x^*, y_{k+1}) - L(x^*, y^*)) + t_{k+1}(t_{k+1} - 1)(L(x^*, y_{k+1}) - L(x^*, y_k)) \\
&\quad + \frac{t_{k+1}}{\gamma}\left\langle A\left(u_{k+1}^\gamma - \gamma x^*\right), v_{k+1}^\gamma - \gamma y^*\right\rangle - (1-\gamma)(t_{k+1} - 1 + \frac{\gamma}{2})\|y_{k+1} - y_k\|^2 \\
&\quad - \frac{1}{2}\|v_{k+1}^\gamma - v_k^\gamma\|^2.
\end{aligned} \tag{3.25}$$

Proof To better understand the inequalities we have to prove, we first 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

$$\begin{aligned}
&\frac{1}{2}\|v^{\gamma}_{k+1} - \gamma y^*\|^2 - \frac{1}{2}\|v^{\gamma}_{k} - \gamma y^*\|^2 \\
&= \langle v^{\gamma}_{k+1} - v^{\gamma}_{k}, v^{\gamma}_{k+1} - \gamma y^*\rangle - \frac{1}{2}\|v^{\gamma}_{k+1} - v^{\gamma}_{k}\|^2 \\
&= \langle v_{k+1} - v_k + (\gamma - 1)(y_{k+1} - y_k), \gamma(y_{k+1} - y^*) + (t_{k+1} - 1)(y_{k+1} - y_k)\rangle \\
&\quad -\frac{1}{2}\|v^{\gamma}_{k+1} - v^{\gamma}_{k}\|^2 \\
&= t_{k+1}\langle \frac{1}{\gamma} Au^{\gamma}_{k+1} - \eta_{k+1}, \gamma(y_{k+1} - y^*) + (t_{k+1} - 1)(y_{k+1} - y_k)\rangle \\
&\quad +(\gamma - 1)\gamma\langle y_{k+1} - y_k, y_{k+1} - y^*\rangle + (\gamma - 1)(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 - \frac{1}{2}\|v^{\gamma}_{k+1} - v^{\gamma}_{k}\|^2 \\
&= -\gamma t_{k+1}\langle \eta_{k+1} - Ax^*, y_{k+1} - y^*\rangle - t_{k+1}(t_{k+1} - 1)\langle \eta_{k+1} - Ax^*, y_{k+1} - y_k\rangle \\
&\quad +\frac{t_{k+1}}{\gamma}\langle A\left(u^{\gamma}_{k+1} - \gamma x^*\right), \\
&\quad v^{\gamma}_{k+1} - \gamma y^*\rangle + (\gamma - 1)\gamma\langle y_{k+1} - y_k, y_{k+1} - y^*\rangle + (\gamma - 1)(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 \\
&\quad -\frac{1}{2}\|v^{\gamma}_{k+1} - v^{\gamma}_{k}\|^2,
\end{aligned} \tag{3.26}$$

as well as

$$\langle y_{k+1} - y_k, y_{k+1} - y^*\rangle = -\frac{1}{2}\left(\|y_k - y^*\|^2 - \|y_{k+1} - y^*\|^2 - \|y_{k+1} - y_k\|^2\right), \tag{3.27}$$

and

$$\begin{aligned}
&-\gamma t_{k+1}\langle \eta_{k+1} - Ax^*, y_{k+1} - y^*\rangle - t_{k+1}(t_{k+1} - 1)\langle \eta_{k+1} - Ax^*, y_{k+1} - y_k\rangle \\
&\le -\gamma t_{k+1}\left(g(y_{k+1}) - g(y^*) - \langle Ax^*, y_{k+1} - y^*\rangle\right) - t_{k+1}(t_{k+1} - 1)\left(g(y_{k+1})\right. \\
&\left. - g(y_k) - \langle Ax^*, y_{k+1} - y_k\rangle\right) \\
&= \gamma t_{k+1}(L(x^*, y_{k+1}) - L(x^*, y^*)) \\
&\quad +t_{k+1}(t_{k+1} - 1)(L(x^*, y_{k+1}) - L(x^*, y_k)),
\end{aligned} \tag{3.28}$$

where the inequality comes from the convexity of the function $g(\cdot) - \langle Ax^*, \cdot\rangle$. Combining (3.26), (3.27) and (3.28), we arrive at

$$\begin{aligned}
&E_2(k+1) - E_2(k) \\
&= \frac{1}{2}\|v^{\gamma}_{k+1} - \gamma y^*\|^2 - \frac{1}{2}\|v^{\gamma}_{k} - \gamma y^*\|^2 + \frac{\gamma(1-\gamma)}{2}\left(\|y_{k+1} - y^*\|^2 - \|y_k - y^*\|^2\right) \\
&\le \gamma t_{k+1}(L(x^*, y_{k+1}) - L(x^*, y^*)) + t_{k+1}(t_{k+1} - 1)(L(x^*, y_{k+1}) - L(x^*, y_k)) \\
&\quad +\frac{t_{k+1}}{\gamma}\langle A\left(u^{\gamma}_{k+1} - \gamma x^*\right), v^{\gamma}_{k+1} - \gamma y^*\rangle - (1-\gamma)(t_{k+1} - 1 + \frac{\gamma}{2})\|y_{k+1} - y_k\|^2
\end{aligned}$$

$$-\frac{1}{2}\|v_{k+1}^{\gamma}-v_k^{\gamma}\|^2,$$

which is nothing else than (3.25). Next, in accordance with the coefficients of the primal-dual gap in (3.25), by multiplying (3.18) with γt_{k+1} and (3.19) with $t_{k+1}(t_{k+1}-1)$, we arrive at

$$\begin{aligned}
&\gamma t_{k+1}\left(L(x_{k+1},y^*)-L(x^*,y^*)\right)+t_{k+1}(t_{k+1}-1)\left(L(x_{k+1},y^*)-L(x_k,y^*)\right)\\
&\le -\frac{t_{k+1}}{\gamma}\langle A^*(v_{k+1}^{\gamma}-\gamma y^*),\gamma(x_{k+1}-x^*)+(t_{k+1}-1)(x_{k+1}-x_k)\rangle\\
&+\frac{t_{k+1}}{\sigma}\langle z_k-x_{k+1},\gamma(x_{k+1}-x^*)+(t_{k+1}-1)(x_{k+1}-x_k)\rangle\\
&+\frac{L_f\,(t_{k+1}-1+\gamma)\,t_{k+1}}{2}\|x_{k+1}-z_k\|^2-\frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k)-\nabla f(x^*)\|^2\\
&-\frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k)-\nabla f(x_k)\|^2\\
&= -\frac{t_{k+1}}{\gamma}\langle A^*\left(v_{k+1}^{\gamma}-\gamma y^*\right),u_{k+1}^{\gamma}-\gamma x^*\rangle+\frac{t_{k+1}}{\sigma}\langle z_k-x_{k+1},u_{k+1}^{\gamma}-\gamma x^*\rangle\\
&\quad+\frac{L_f\,(t_{k+1}-1+\gamma)\,t_{k+1}}{2}\|x_{k+1}-z_k\|^2-\frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k)-\nabla f(x^*)\|^2\\
&\quad-\frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k)-\nabla f(x_k)\|^2.
\end{aligned}\tag{3.29}$$

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

$$\begin{aligned}
&\frac{t_{k+1}}{\sigma}\langle z_k-x_{k+1},u_{k+1}^{\gamma}-\gamma x^*\rangle\\
&=\frac{1}{\sigma}\Big(\langle u_k^{\gamma}-u_{k+1}^{\gamma},u_{k+1}^{\gamma}-\gamma x^*\rangle-(1-\gamma)(t_{k+1}-1)\|x_{k+1}-x_k\|^2\\
&\quad+(1-\gamma)\gamma\langle(x_k-x_{k+1}),(x_{k+1}-x^*)\rangle\Big)\\
&=-\frac{1}{2\sigma}\|u_k^{\gamma}-u_{k+1}^{\gamma}\|^2-\frac{1}{2\sigma}\|u_{k+1}^{\gamma}-\gamma x^*\|^2+\frac{1}{2\sigma}\|u_k^{\gamma}-\gamma x^*\|^2\\
&\quad-\frac{(1-\gamma)(t_{k+1}-1)}{\sigma}\|x_{k+1}-x_k\|^2-\frac{(1-\gamma)\gamma}{2\sigma}\|x_k-x_{k+1}\|^2\\
&\quad-\frac{(1-\gamma)\gamma}{2\sigma}\|x_{k+1}-x^*\|^2+\frac{(1-\gamma)\gamma}{2\sigma}\|x_k-x^*\|^2
\end{aligned}\tag{3.30}$$

holds. Note that all summands of $E_1(k+1)-E_1(k)$ can be found in (3.30). Combining this obervation with (3.29), we obtain

$$\begin{aligned}
&E_1(k+1) - E_1(k)\\
=& \frac{1}{2\sigma}\|u^\gamma_{k+1} - \gamma x^*\|^2 - \frac{1}{2\sigma}\|u^\gamma_k - \gamma x^*\|^2 + \frac{\gamma(1-\gamma)}{2\sigma}\|x_{k+1} - x^*\|^2\\
&-\frac{\gamma(1-\gamma)}{2\sigma}\|x_k - x^*\|^2\\
=& -\frac{t_{k+1}}{\sigma}\langle z_k - x_{k+1}, u^\gamma_{k+1} - \gamma x^*\rangle - \frac{(1-\gamma)}{\sigma}\left(t_{k+1} - 1 + \frac{\gamma}{2}\right)\|x_{k+1} - x_k\|^2\\
-& \frac{1}{2\sigma}\|u^\gamma_k - u^\gamma_{k+1}\|^2\\
\le& -\gamma t_{k+1}\left(L(x_{k+1}, y^*) - L(x^*, y^*)\right) - t_{k+1}(t_{k+1} - 1)\left(L(x_{k+1}, y^*) - L(x_k, y^*)\right)\\
&-\frac{t_{k+1}}{\gamma}\langle A^*\left(v^\gamma_{k+1} - \gamma y^*\right), u^\gamma_{k+1} - \gamma x^*\rangle - \frac{(1-\gamma)}{\sigma}\left(t_{k+1} - 1 + \frac{\gamma}{2}\right)\|x_{k+1} - x_k\|^2\\
&-\frac{1}{2\sigma}\|u^\gamma_k - u^\gamma_{k+1}\|^2\\
&+\frac{L_f\left(t_{k+1} - 1 + \gamma\right)t_{k+1}}{2}\|x_{k+1} - z_k\|^2 - \frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2\\
&-\frac{t_{k+1}(t_{k+1} - 1)}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2. \qquad (3.31)
\end{aligned}$$

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

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

And finally, by substituting (3.32) into (3.31), we obtain (3.24). □

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

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

$$\sum_{k\ge 1} t_{k+1}\big((\gamma - L_f\sigma)t_{k+1} + (1-\gamma)L_f\sigma\big)\|x_{k+1} - z_k\|^2$$

$$
\begin{aligned}
&< +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1}-1)\|x_{k+1}-x_k\|^2 < +\infty,\\
&\sum_{k\geq 1}\|v_{k+1}^{\gamma}-v_k^{\gamma}\|^2 < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1}-1)\|y_{k+1}-y_k\|^2 < +\infty,\\
&\sum_{k\geq 1} t_{k+1}\|\nabla f(z_k)-\nabla f(x^*)\|^2\\
&< +\infty,\ \sum_{k\geq 1} t_{k+1}(t_{k+1}-1)\|\nabla f(z_k)-\nabla f(x_k)\|^2 < +\infty. \qquad (3.33)
\end{aligned}
$$

Proof From Lemma 3.2 we conclude that

$$
\begin{aligned}
&E(k+1)-E(k)\\
=\ &\big(t_{k+2}(t_{k+2}-1)-t_{k+1}(t_{k+1}-1)\big)\big(L(x_{k+1},y^*)-L(x^*,y_{k+1})\big)+E_1(k+1)-E_1(k)\\
&+t_{k+1}(t_{k+1}-1)\big((L(x_{k+1},y^*)-L(x_k,y^*))-(L(x^*,y_{k+1})-L(x^*,y_k))\big)\\
&+E_2(k+1)-E_2(k)\\
\leq\ &\Big(t_{k+2}^2-t_{k+1}^2-t_{k+2}+(1-\gamma)t_{k+1}\Big)\big(L(x_{k+1},y^*)-L(x^*,y_{k+1})\big)\\
&-\frac{(1-\gamma)}{\sigma}\big(t_{k+1}-1\big)\|x_{k+1}-x_k\|^2\\
&-\frac{t_{k+1}}{2\sigma}\big((\gamma-L_f\sigma)t_{k+1}+(1-\gamma)L_f\sigma\big)\|x_{k+1}-z_k\|^2-\frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k)-\nabla f(x^*)\|^2\\
&-\frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k)-\nabla f(x_k)\|^2-(1-\gamma)(t_{k+1}-1+\frac{\gamma}{2})\|y_{k+1}-y_k\|^2\\
&-\frac{1}{2}\|v_{k+1}^{\gamma}-v_k^{\gamma}\|^2. \qquad (3.34)
\end{aligned}
$$

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

$$
\begin{aligned}
t_{k+2}^2-t_{k+1}^2-t_{k+2}+(1-\gamma)t_{k+1} &\leq (m-1)t_{k+2}\\
+(1-\gamma)t_{k+1} &\leq (m-\gamma)t_{k+1}\leq 0. \qquad (3.35)
\end{aligned}
$$

Thus, all the coefficients in the right-hand side of (3.34) are nonpositive, and it follows that the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing for every $k \geq 1$. We complete the proof of (3.33) via Lemma A. 4. □

Remark 3.2 By Theorem 3.1, if $\gamma < 1$, we have

$$
\sum_{k\geq 1}(t_{k+1}-1)(\|x_{k+1}-x_k\|^2+\|y_{k+1}-y_k\|^2) < +\infty. \qquad (3.36)
$$

It is obvious that $\gamma < 1$ is equivalent to $\theta\alpha - \theta - 1 > 0$. In the continuous case, if $\theta\alpha - \theta - 1 > 0$, then (2.4) can be rewritten as

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

which can be seen as the continuous counterpart of (3.36). When t_k is computed by the Chambolle-Dossal [16] or the Attouch-Cabot rule [1], (2.3) can likewise be seen as the continuous counterpart of the first line of (3.33).

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 $O(1/t_k)$ convergence rate for the sequences $\|x_k - x_{k-1}\|$ and $\|y_k - y_{k-1}\|$. This is compatible with the results $\|\dot{x}(t)\| = O(1/t)$ and $\|\dot{y}(t)\| = O(1/t)$ for the continuous-time dynamic system, as described in Theorem 2.1.

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

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

Then, the sequences $\{x_k\}_{k \geq 0}$, $\{y_k\}_{k \geq 0}$, $\{t_k(x_k - x_{k-1})\}_{k \geq 1}$ and $\{t_k(y_k - y_{k-1})\}_{k \geq 1}$ are bounded. Moreover, the sequence $\{v_{k+1} - v_k\}_{k \geq 0}$ is bounded.

Proof Since $\{E(k)\}_{k \geq 1}$ is nonincreasing we have $E(k) \leq E(1)$. Consequently, by the definition of $\{E(k)\}_{k \geq 1}$, the sequences $\{u_k^\gamma\}_{k \geq 1}$ and $\{v_k^\gamma\}_{k \geq 1}$ are bounded. Now 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

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

holds. Similarly, we have

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

and so the energy function can be rewritten as

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

$$+\frac{\gamma}{2}t_k\|h_k-h^*\|^2_{\mathcal{W}}-\frac{\gamma}{2}(t_k-1)\|h_{k-1}-h^*\|^2_{\mathcal{W}}$$
$$+\frac{1}{2}(t_k-1+\gamma)(t_k-1)\|h_k-h_{k-1}\|^2_{\mathcal{W}}. \tag{3.38}$$

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

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

It follows that

$$\begin{aligned}\frac{\gamma}{2}t_k\|h_k-h^*\|^2_{\mathcal{W}}&\leq\frac{\gamma}{2}(t_k-1)\|h_{k-1}-h^*\|^2_{\mathcal{W}}+E(1)\\&\leq\frac{\gamma}{2}t_{k-1}\|h_{k-1}-h^*\|^2_{\mathcal{W}}+E(1),\end{aligned}\tag{3.39}$$

where the second inequality follows from the fact $t_{k+1}-t_k<1$ in Lemma A. 1. After summing up (3.39) from 1 to k, we have for every $k \geq 1$ that

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

and so

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

With this, we conclude that both sequences $\{x_k\}_{k\geq 0}$ and $\{y_k\}_{k\geq 0}$ 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. Moreover, by the definition of v_k, it is obvious that the sequence $\{v_{k+1}-v_k\}_{k\geq 0}$ is bounded. □

Note that condition (3.37) in Proposition 3.1, which is crucial for the following analysis of weak convergence of sequence of iterates, has also been proposed in [6, 11]. We would like to emphasize here that we can show the boundedness of the sequences considered in Proposition 3.1 by assuming $\gamma < 1$. This can be seen as follows. By Theorem 3.1, the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing, by which $E(k) \leq E(1)$ follows. This yields

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

We notice that the sequences $\{x_k\}_{k\geq 0}$ and $\{y_k\}_{k\geq 0}$ are bounded when $\gamma < 1$. Combining with the boundedness of $\{u_k^\gamma\}_{k\geq 1}$ and $\{v_k^\gamma\}_{k\geq 1}$, we also present that $\{t_k(x_k - x_{k-1})\}_{k\geq 1}$, $\{t_k(y_k - y_{k-1})\}_{k\geq 1}$ are bounded, in other words, $\|x_k - x_{k-1}\| = O(1/t_k)$ and $\|y_k - y_{k-1}\| = O(1/t_k)$.

3.3 Fast convergence

This section has two aims. First, we show fast convergence of the primal-dual gap of the bilinearly coupled convex-concave saddle point problem (1.1) when Algorithm 1 is deployed to solve it. Second, we consider a special case of (1.1) equivalent to a nonsmooth convex optimization problem with linear constraints and simplify Algorithm 1 accordingly. We then show fast convergence of the primal-dual gap, the feasibility measure and the objective function value of this nonsmooth problem when this algorithm is deployed to solve it.

Note again that the sequence $\{E(k)\}_{k\geq 1}$ is nonincreasing, which yields $E(k) \leq E(1)$. Thus we have

$$L(x_k, y^*) - L(x^*, y_k) \leq \frac{E(1)}{t_{k+1}(t_{k+1} - 1)}. \tag{3.40}$$

Boţ et al. [11] presented several prominent choices for the sequence $\{t_k\}_{k\geq 1}$, i. e. Nesterov's rule [33], the Chambolle-Dossal rule [16], and the Attouch-Cabot rule [1] (which $k \geq [\alpha] + 1$). These rules all satisfy the condition (3.15) in Algorithm 1. We now consider what convergence rates we can achieve under these classical construction of $\{t_k\}_{k\geq 1}$.

First, we consider Nesterov's rule as proposed in Nesterov [33]:

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

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 [8]). 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

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

Second, the Chambolle-Dossal rule [16] 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 $\tau = \frac{1}{\alpha-1}$ (see, for instance, Example 3.15 in [11]). 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 as above 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:

$$\begin{aligned} &\min_{y\in\mathcal{Y}} \; g(y), \\ &\text{s.t. } A^*y = b. \end{aligned} \tag{3.41}$$

where g is a proper, convex and lower semicontinuous function. Recently, He et al. [23] 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 $t_{k+1}^2 - mt_{k+1} - t_k^2 \leq 0$, $\forall k \geq 1$. choose $x_0 = x_1$ and $y_0 = y_1$. For every $k \geq 1$, set

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

Note that by choosing $t_k = 1 + \frac{k-2}{\alpha-1}$ in Algorithm 2 we arrive at an algorithm similar but different to Algorithm 1 in [23]. Next, we will show fast convergence of the feasibility measure and the objective function value when Algorithm 2 is used to solve problem (3.41).

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

$$\|A^*y_k - b\| \leq \frac{t_1^2\|A^*y_1 - b\| + 2C}{t_k^2}, \tag{3.42}$$

$$|g(y_k) - g(y^*)| \leq \frac{(t_1^2\|A^*y_1 - b\| + 2C)\|x^*\|}{t_k^2} + \frac{E(1)}{t_{k+1}(t_{k+1}-1)}. \quad (3.43)$$

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

$$x_{k+1} - x_k - \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1}) = -t_{k+1}(A^*y_{k+1} - b) + (t_{k+1}-1)(A^*y_k - b). \quad (3.44)$$

Reformulating (3.44) yields

$$-t_{k+1}(x_{k+1} - x_k) + t_k(x_k - x_{k-1}) - (x_k - x_{k-1}) = \delta_{k+1} - (1-a_k)\delta_k, \quad (3.45)$$

where $\delta_k := t_k^2(A^*y_k - b)$ 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 (3.45), we arrive at

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

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

$$\|A^*y_k - b\| \leq \frac{t_1^2\|A^*y_1 - b\| + 2C}{t_k^2},$$

which yields (3.42). Finally, by (3.40) and $|g(y_k) - g(y^*)| \leq \|L(x_k, y^*) - L(x^*, y_k)\| + \|x^*\|\|(A^*y_k - b)\|$, we arrive at (3.43), which completes the proof. □

We can now consider some special cases of Algorithm 2. With Nesterov's rule [33], the Chambolle-Dossal rule [16], 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 $o(1/k)$ derived in [12]. However, note that their $o(1/k)$ convergence rate for $\|y_k - y_{k-1}\|$ is better than our convergence rate of $O(1/k)$.

4 A fast primal-dual algorithm based on explicit discretization

In this section, we assume 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 obtain the same convergence properties of the dynamical system (1.5) in a way similar to the proof of Theorem 2.1. In what follows, we will investigate a numerical algorithm that is derived directly from an explicit discretization of the dynamical system (1.5).

4.1 A fast algorithm and its convergence rate of the primal-dual gap

Fast gradient algorithms originating from various second order dynamical systems in the spirit of Nesterov's accelerated gradient method have been proposed in [3, 11, 35]. 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):

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

where λ_k can be obtained through a discussion analogous to that for z_k. By the relations given in (4.2), we get

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

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

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

Now we are in a position to present our fast primal-dual algorithm (for short, FPDA3) for the smooth case:

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

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

Choose $\{t_k\}_{k \ge 1}$ as a nondecreasing sequence with

$$t_1 \ge 1 \text{ and } t_{k+1}^2 - m t_{k+1} - t_k^2 \le 0, \ \forall k \ge 1. \tag{4.5}$$

Choose $x_0 = x_1$ and $y_0 = y_1$. For every $k \ge 1$, set

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

Compared to Algorithm 1 which has been designed for the nonsmooth case, the subproblem in Algorithm 3 does not rely on the structure of f or g. Moreover, although a choice of $\gamma = 1$ provides for a simplified version of Algorithm 3 without affecting the fast convergence rate, we will see that the condition $\gamma < 1$ is indispensable for showing weak convergence of the iterates (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 $\alpha \ge 3$, 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:

$$\mathcal{E}(k) := t_{k+1}(t_{k+1} - 1)\left(L(x_k, y^*) - L(x^*, y_k)\right) + \mathcal{E}_1(k) + \mathcal{E}_2(k), \tag{4.6}$$

where

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

It is obvious that $\mathcal{E}(k) \geq 0$ for every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$.

Theorem 4.1 *Let $\{(x_k, y_k)\}_{k\geq 0}$ 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:*

$$\begin{aligned}
&(\gamma - m)\sum_{k\geq 1} t_k\left(L(x_k, y^*) - L(x^*, y_k)\right) < +\infty,\\
&\sum_{k\geq 1} t_{k+1}\big((\gamma - L_f\sigma)t_{k+1} + (1-\gamma)L_f\sigma\big)\|x_{k+1} - z_k\|^2\\
&\quad < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1} - 1)\|x_{k+1} - x_k\|^2 < +\infty,\\
&\sum_{k\geq 1} t_{k+1}\big((\gamma - L_g\rho)t_{k+1} + (1-\gamma)L_g\rho\big)\|y_{k+1} - \lambda_k\|^2\\
&\quad < +\infty,\ (1-\gamma)\sum_{k\geq 1}(t_{k+1} - 1)\|y_{k+1} - y_k\|^2 < +\infty,\\
&\sum_{k\geq 1} t_{k+1}\|\nabla f(z_k) - \nabla f(x^*)\|^2 < +\infty,\\
&\quad\sum_{k\geq 1} t_{k+1}(t_{k+1} - 1)\|\nabla f(z_k) - \nabla f(x_k)\|^2 < +\infty,\\
&\sum_{k\geq 1} t_{k+1}\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 < +\infty,\\
&\quad\sum_{k\geq 1} t_{k+1}(t_{k+1} - 1)\|\nabla g(\lambda_k) - \nabla g(y_k)\|^2\\
&\quad < +\infty.
\end{aligned} \tag{4.7}$$

Proof In a way similar to the proof of (3.24), we can show that

$$\mathcal{E}_2(k+1) - \mathcal{E}_2(k)$$

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

holds. Combining this with (3.24), we arrive at

$$
\begin{aligned}
&\mathcal{E}(k+1) - \mathcal{E}(k) \\
&= \left(t_{k+2}(t_{k+2}-1) - t_{k+1}(t_{k+1}-1)\right)\left(L(x_{k+1}, y^*) - L(x^*, y_{k+1})\right) + \mathcal{E}_1(k+1) - \mathcal{E}_1(k) \\
&\quad + t_{k+1}(t_{k+1}-1)\left((L(x_{k+1}, y^*) - L(x_k, y^*)) - (L(x^*, y_{k+1}) - L(x^*, y_k))\right) \\
&\quad + \mathcal{E}_2(k+1) - \mathcal{E}_2(k) \\
&\le \left(t_{k+2}^2 - t_{k+1}^2 - t_{k+2} + (1-\gamma)t_{k+1}\right)\left(L(x_{k+1}, y^*) - L(x^*, y_{k+1})\right) \\
&\quad - \frac{(1-\gamma)}{\sigma}\left(t_{k+1}-1\right)\|x_{k+1}-x_k\|^2 \\
&\quad - \frac{t_{k+1}}{2\sigma}\left((\gamma - L_f\sigma)t_{k+1} + (1-\gamma)L_f\sigma\right)\|x_{k+1}-z_k\|^2 - \frac{\gamma t_{k+1}}{2L_f}\|\nabla f(z_k) - \nabla f(x^*)\|^2 \\
&\quad - \frac{t_{k+1}(t_{k+1}-1)}{2L_f}\|\nabla f(z_k) - \nabla f(x_k)\|^2 \\
&\quad - \frac{(1-\gamma)}{\rho}\left(t_{k+1}-1\right)\|y_{k+1}-y_k\|^2 - \frac{\gamma t_{k+1}}{2L_g}\|\nabla g(\lambda_k) - \nabla g(y^*)\|^2 \\
&\quad - \frac{t_{k+1}}{2\rho}\left((\gamma - L_g\rho)t_{k+1} + (1-\gamma)L_g\rho\right)\|y_{k+1}-\lambda_k\|^2 \\
&\quad - \frac{t_{k+1}(t_{k+1}-1)}{2L_g}\|\nabla g(\lambda_k) - \nabla g(y_k)\|^2.
\end{aligned} \tag{4.8}
$$

By the assumptions (4.4), (4.5), and inequality (3.35), all the coefficients in the right-hand side of (4.8) are nonpositive and so $\mathcal{E}(k+1) - \mathcal{E}(k) \le 0$. We complete the proof of (4.7) via Lemma A. 4. □

Remark 4.1 Compared to the energy function in [11], which contains 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 [11], where v_{k+1}^{γ} is replaced by $\tilde{v}_{k+1}^{\gamma} := v_{k+1}^{\gamma} + (1-\gamma)(y_{k+1} - y_k)$. Only in the nonsmooth case we replaced v_{k+1} 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 indeed replace v_{k+1}^{γ} with $\tilde{v}_{k+1}^{\gamma} := v_{k+1}^{\gamma} + (1-\gamma)(y_{k+1} - y_k)$ in the first line of explicit discretization scheme (4.2) and then introduce the following energy function to analyze the behaviour of the thus modified algorithm:

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

for every $(x^*, y^*) \in \mathbb{S}$ and every $k \geq 1$. By an analysis similar to Theorem 4.1 and Proposition 3.9 in [11], we then obtain that the sequence $\{\mathcal{E}_k'\}_{k\geq 1}$ is nonincreasing and we have the following statements:

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

Since $\{\mathcal{E}(k)\}_{k\geq 1}$ is nonincreasing for every $k \geq 1$, we again arrive at $L(x_k, y^*) - L(x^*, y_k) \leq \frac{\mathcal{E}(1)}{t_{k+1}(t_{k+1}-1)}$. We notice that the classical three schemes for t_k, i. e. Nesterov's rule [33], the Chambolle-Dossal rule [16], 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.

4.2 Convergence of the iterates

In this section, we will prove that the sequence generated by Algorithm 3 weakly converges to a primal-dual optimal solution of the bilinearly coupled saddle point problem (1.1) for the smooth case. We start by providing some useful estimates.

Proposition 4.1 *Let* $\{(x_k, y_k)\}_{k\geq 0}$ *be the sequence generated by Algorithm 3 and let* $(x^*, y^*) \in \mathbb{S}$. *Assume that* $0 < m < \gamma < 1$ *holds. Then we have the following statements:*

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

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

$$\sum_{k\geq 1} t_k \left\| A \left(x_k - x^*\right) \right\|^2 < +\infty, \tag{4.11}$$

$$\sum_{k\geq 1} t_{k+1}(t_{k+1} - 1)^2 \left\| A \left(x_{k+1} - x_k\right) \right\|^2 < +\infty. \tag{4.12}$$

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

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

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

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

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

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

But by (4.5), we see that

$$\begin{aligned}\frac{t_{k+1}(t_{k+1}-1)}{\gamma} - t_k\left(1+\frac{t_k-1}{\gamma}\right) &= \frac{1}{\gamma}\left(t_{k+1}^2 - t_{k+1} - t_k^2 + t_k - \gamma t_k\right)\\ &\le \frac{1}{\gamma}\left((m-1)t_{k+1} - (\gamma-1)t_k\right)\\ &\le \left(\frac{m}{\gamma}-1\right)t_k, \end{aligned} \tag{4.13}$$

where the last inequality follows from the fact that $m < 1$ and $\{t_k\}$ is nondecreasing. Therefore,

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

Let us set

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

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

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

and the sequence $\left\{t_k \left(1 + \frac{t_k - 1}{\gamma}\right) \|A^* (y_k - y^*)\|^2\right\}$ is convergent and bounded. Similarly, we obtain (4.11), (4.12) and the fact that $t_k \left(1 + \frac{1}{\gamma}(t_k - 1)\right) \|A^* (x_k - x^*)\|^2$ is convergent. Since $t_k \leq \left(1 + \frac{1}{\gamma}(t_k - 1)\right)$ 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 $M > 0$. Thus, $\|A^* (y_k - y^*)\| \leq \frac{M}{t_k}$. Similarly, we obtain $\|A (x_k - x^*)\| \leq \frac{M}{t_k}$. □

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, let $(x^*, y^*) \in \mathbb{S}$ and suppose that $0 < m < \gamma < 1$ holds. Then, the limit $\lim_{k\to+\infty} \frac{1}{\sigma}\|x_k - x^*\|^2 + \frac{1}{\rho}\|y_k - y^*\|^2$ exists.*

Proof Set $a_k := \frac{\gamma}{2\sigma}\|x_k - x^*\|^2 + \frac{\gamma}{2\rho}\|y_k - y^*\|^2$ for all $k \geq 0$. By considering a reformulation of $\mathcal{E}(k)$ similar to (3.38) and the fact that $\mathcal{E}(k+1) \leq \mathcal{E}(k)$ for every $k \geq 1$, we obtain

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

and thus

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

$$
\begin{aligned}
&+\frac{1}{2}(t_{k+1}-1+\gamma)\left(\frac{1}{\sigma}\|x_{k+1}-x_k\|^2+\frac{1}{\rho}\|y_{k+1}-y_k\|^2\right)\\
&\leq (t_k-1)\left(\frac{t_{k+1}(t_{k+1}-1)}{t_k}\left(L(x_k,y^*)-L(x^*,y_k)\right)+\frac{1}{2}(t_k-1+\gamma)\left(\frac{1}{\sigma}\|x_k-x_{k-1}\|^2\right.\right.\\
&\left.\left.+\frac{1}{\rho}\|y_k-y_{k-1}\|^2\right)\right)+(t_k-1)\left(a_k-a_{k-1}\right)\\
&+(t_k+1)\left(L(x_k,y^*)-L(x^*,y_k)\right)\\
&+\frac{1}{2}(t_{k+1}-1+\gamma)\left(\frac{1}{\sigma}\|x_{k+1}-x_k\|^2+\frac{1}{\rho}\|y_{k+1}-y_k\|^2\right),
\end{aligned}
\tag{4.15}
$$

where the last inequality follows from $t_{k+1}-1 \leq t_k$, which in turn holds due to Lemma A. 1. We further define the sequences

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

From these definitions, it is obvious that $a_{k+1} \leq a_k + \beta_{k+1}$. By (4.15) we arrive at $t_{k+1}\beta_{k+1} \leq (t_k-1)\beta_k + d_k$. In addition, from Theorem 4.1, we note 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 4.2 *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3 and let $(x^*, y^*) \in \mathbb{S}$. Assume further that $\{t_k\}_{k\geq 1}$ is chosen to satisfy (3.37) 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,

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

Proof From the results of Theorem 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 (3.37), 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 therefore

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

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 (3.37) and (4.9), we obtain $\|A^*(y_k - y^*)\| = o\left(1/\sqrt{k}\right)$ which yields $\|\nabla_x L(x_k, y_k)\| = o\left(1/\sqrt{k}\right)$. Similarly, we have $\|\nabla_y L(x_k, y_k)\| = o\left(1/\sqrt{k}\right)$. This completes the proof. □

Theorem 4.3 *Let $\{(x_k, y_k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3 and let $(x^*, y^*) \in \mathbb{S}$. Assume further that $\{t_k\}_{k\geq 1}$ is chosen to satisfy* (3.37) *and that $0 < m < \gamma < 1$ holds. Then, the sequence $\{(x_k, y_k)\}_{k\geq 0}$ 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 the sequence $\{(x_k, y_k)\}_{k\geq 0}$. Thus, there exists a sequence $\{(x_{k_n}, y_{k_n})\}_{n\geq 0}$ such that $(x_{k_n}, y_{k_n}) \rightharpoonup (\bar{x}, \bar{y})$ as $n \to +\infty$. By Theorem 4.2, we get

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

Since the graph of the operator $\mathcal{T}_L$ in (1.3) is sequentially closed (see Proposition 20.38 from [7]), we conclude that

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

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 Suppose we choose the Chambolle-Dossal rule or the Attouch-Cabot rule for the sequence $\{t_k\}_{k\geq 1}$ with $\alpha > 3$, $m = \frac{1}{\alpha-2} < \gamma < 1$, $\sigma \leq \frac{\gamma}{L_f}$ and $\rho \leq \frac{\gamma}{L_g}$. Then, by Theorem 4.3, 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 fast convergence rate of the primal-dual gap still holds, we can not obtain the convergence of the sequence of primal-dual iterates since the conditions in Theorem 4.3 require $m < \gamma < 1$.

5 Numerical experiments

In this section we compare our proposed algorithms with other algorithms from the literature on some examples. We illustrate that the theoretical convergence rates we obtained are closely matched by numerical results. In our first example, we consider the convergence of the primal-dual gap and the term $\|(x_k, y_k) - (x_{k-1}, y_{k-1})\|$ for a nonsmooth bilinearly coupled saddle point problem. In a second example we consider the convergence behavior of the primal-dual gap, the gradient and the term $\|(x_k, y_k) - (x_{k-1}, y_{k-1})\|$ for a smooth bilinearly coupled saddle point problem. In order to investigate the stability and efficiency of our algorithms, we test two groups of problems with different dimensions and run our algorithms ten times for each group. Then, we report the average numerical performances of the primal-dual gap, the gradient and the term $\|(x_k, y_k) - (x_{k-1}, y_{k-1})\|$ in the following examples.

5.1 Nonsmooth saddle-point problems

First, let us consider the following family of nonsmooth saddle point problems:

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m} \frac{1}{2}\|Qx - q\|^2 + \langle Ax, y\rangle - \left(\mu\|y\|_1 + \frac{\kappa}{2}\|y\|^2\right), \tag{5.1}$$

where $Q \in \mathbb{R}^{n\times n}$, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $\mu, \kappa \geq 0$. We note that several classical problems admit this formulation, for examples Lasso models and regression problem with elastic net regularizer.

Set $\mu = 1$ and $\kappa = 0.1$. We generate Q, A and q by choosing each entry of these matrices and vector independently from each other by drawing from the standard Gaussian distribution. We solve the subproblem occuring in Algorithm 1 by the fast iterative shrinkage-thresholding algorithm (FISTA, [8]) with the stopping condition

$$\frac{\|z'_k - z'_{k-1}\|}{\max\{\|z'_{k-1}\|, 1\}} \leq \delta$$

or the number of iterations exceeds 100. Here, the z'_k are the iterates generated by FISTA, and we use $\delta := 10^{-10}$.

We compare the performance of Algorithm 1 (FPDA1) with the primal-dual algorithm (PDA) [[14], Algorithm 1] and the accelerated primal-dual algorithm (APDA) [[24], Algorithm 1 with Option 2] which fully utilizes the strong convexity of $g(y)$. For the parameters occuring in the various algorithms, we have configured them in such a way that each group of parameters fulfills its respective convergence conditions, and set them as follows:

- FPDA1: $\sigma = 10^{-4}, \theta = 10^{-1}$. We consider three different choices for α, namely $\alpha = 30$, $\alpha = 50$ and $\alpha = 70$.
- PDA: $\alpha = 2 \times 10^{-4}, \tau = 2 \times 10^{-3}, \theta = 1$;
- APDA: $\alpha = 0.2/\|A\|^2, \beta = 4$.

To provide an estimate for the value of the primal-dual gap, we approximate the unknown solution (x^*, y^*) by

$$(\tilde{x}^*, \tilde{y}^*) := (x_k, y_k),$$

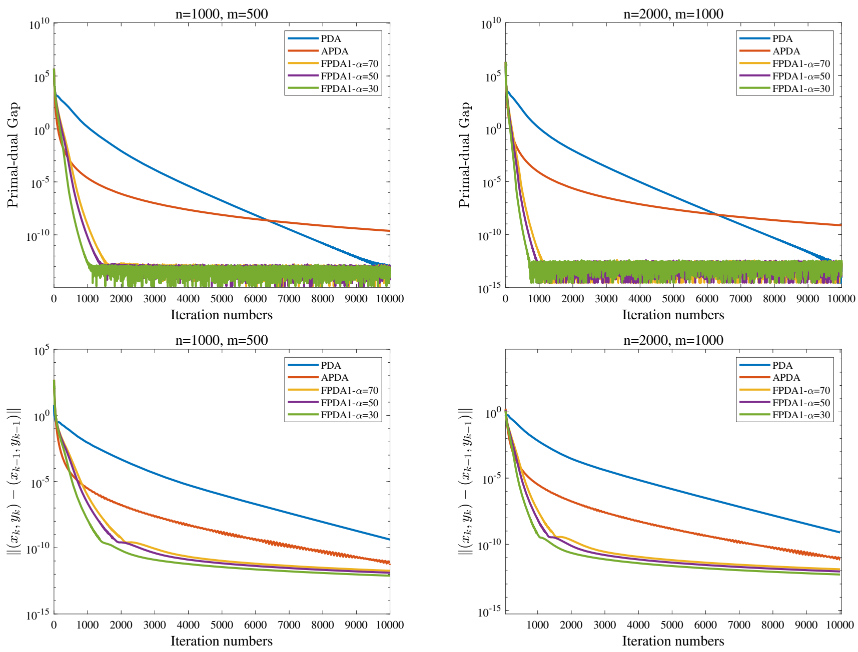

Fig. 1 Average performance of the primal-dual gap and the sequence of iterates for the nonsmooth test case

where k is the smallest index for which

$$\|(x_k, y_k) - (x_{k-1}, y_{k-1})\|_2 \leq \epsilon_1$$

and

$$\|L(x_k, y_{k-1}) - L(x_{k-1}, y_k)\|_2 \leq \epsilon_2$$

holds. Here we have used $\epsilon_1 := 10^{-11}$ and $\epsilon_2 := 10^{-14}$.

Figure 1 illustrates the average convergence results of primal-dual gap and sequence against the number of iterations. The results are very similar for the different problem dimensions $n = 1000, m = 500$ and $n = 2000, m = 1000$ considered. As can be seen in Fig 1, our Algorithm FPDA1 exhibits superior performance compared to the other algorithms, for all choices of the parameter α. In all cases, the number of iterations needed for Algorithm FPDA1 to achieve convergence is smaller than any other algorithms. We also observe that the smaller the parameter α is, the better Algorithm FPDA1 performs.

Next, we consider the non-strongly convex scenario for problem (5.1), where we require only that both f and g are merely convex functions. For the function f, we

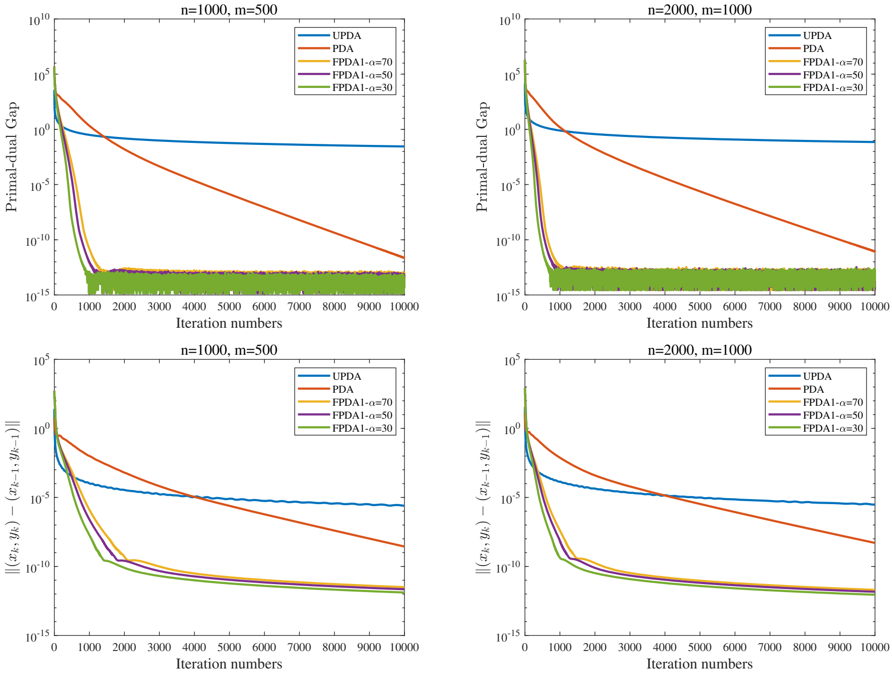

Fig. 2 Average performance of the primal-dual gap and the sequence of iterates for the non-strongly convex test case

now consider a matrix Q of the form

$$Q = \begin{pmatrix} Q_1 & 0 \\ 0 & 0 \end{pmatrix},$$

where $Q_1 \in \mathbb{R}^{(n-1)\times(n-1)}$ is generated by choosing each entry independently of the others, with each entry drawn from the standard Gaussian distribution. For g, we set $\kappa = 0$. Then, both f and g are convex but not strongly convex. We compare the performance of Algorithm 1 (FPDA1) with $\alpha = 30$, $\alpha = 50$ and $\alpha = 70$, against the performance of the primal-dual algorithm PDA [[14], Algorithm 1] and the unified primal-dual algorithm UPDA [[37], Algorithm 1]. Here we use the same parameters for FPDA1 and PDA as in the first experiment. For UPDA, we set $\beta_0 = 50$, while τ is chosen as in Theorem 1 from [37]. Figure 2 illustrates the average convergence results of primal-dual gap and sequence against the number of iterations. We notice that Algorithm FPDA1 still exhibits superior numerical performance compared to the other algorithms considered, for all choices of α.

One can also observe that Algorithm FPDA1 exhibits oscillations in the primal-dual gap found, once numerical convergence has been established. This is due to using $(\tilde{x}^*, \tilde{y}^*)$ as an approximation for the actual optimal primal-dual solution (x^*, y^*) to compute an approximation to the primal-dual gap which cannot be computed directly.

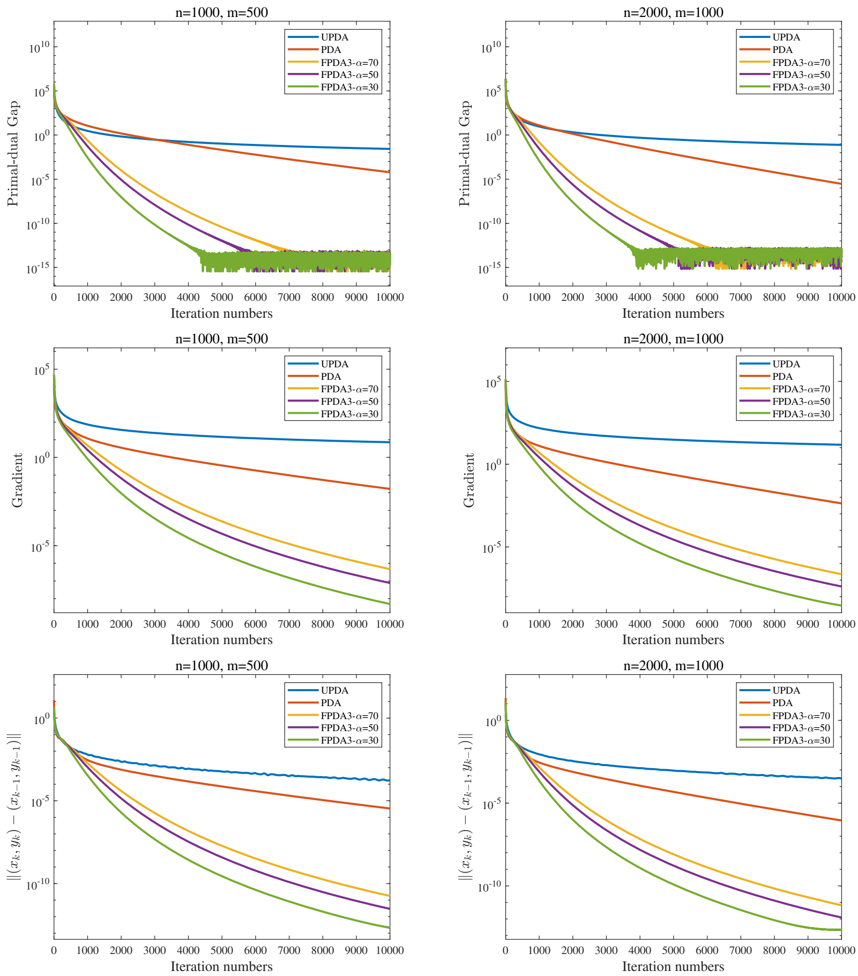

Fig. 3 Average performance of the primal-dual gap, the gradient of the Lagrangian function, and the sequence of iterates for the smooth test case

5.2 Smooth saddle point problems

Here we focus on the following family of smooth convex-concave saddle point problem:

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} \frac{1}{2}\|Qx - q\|^2 + \langle Ax, y\rangle - \frac{1}{2}\|Py - p\|^2,$$

where $Q \in \mathbb{R}^{n\times n}$, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ as well as $P \in \mathbb{R}^{m\times m}$, $p \in \mathbb{R}^m$. Note that quadratic minimax games, linear regression and robust least squares problems admit this formulation. Again, we generate Q, A, P and q, p by choosing each entry of these matrices and vector independently from each other by drawing from the standard Gaussian distribution.

We compare the performance of Algorithm 3 (FPDA3) with $\alpha = 30$, $\alpha = 50$ and $\alpha = 70$, against the performance of the primal-dual algorithm (PDA) [[14], Algorithm 1] and the unified primal-dual algorithm (UPDA) [[37], Algorithm 1].

In [37], the authors introduced three choices for the parameter τ occuring in their algorithm, for the cases where the objective function is convex-concave, strongly convex-concave and strongly convex-strongly concave, respectively. However, since Q and P are randomly generated, we only know the convexity of f and g but can not guarantee their strong convexity. For this reason, we are still using the iteration rule of τ given in Theorem 1 in [37], which does not rely on the strong convexity of f and g. Here we take the same setting of parameters for FPDA3 and PDA as in the first experiment and additionally set $\rho = 10^{-4}$ in FPDA3. For UPDA, we use the parameter $\beta_0 = 0.1$, while τ is chosen as described above.

Figure 3 shows the primal-dual gap, the norm of the gradient of the Lagrangian function, and the sequence of iterates against the number of iterations. As it can be seen, the results are similar for different problem dimensions. Moreover, Algorithm FPDA3 exhibits superior numerical performance compared to the other algorithms considered. We note that the norm of the gradient, $\|(\nabla_x L(x_k, y_k), \nabla_y L(x_k, y_k))\|$, converges faster to zero when Algorithm FPDA3 is used than when any other algorithm is deployed. Also, the smaller α is, the better Algorithm FPDA3 performs.

6 Conclusion and perspectives

Our novel inertial primal-dual dynamics (1.5) allow us to construct two first-order algorithms for bilinearly coupled saddle point problems. 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, as well as global convergence of the iterates to optimal saddle points. 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 a convergence rate of $O\left(1/k^2\right)$ when both f and g are two convex lower semicontinuous and proper functions. 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. These topics are subject to further research.

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. The discrete version of the lemma stated below can be found in Lemma 2.47 of [7].

Lemma A. 2 *Let $\mathcal{X}$ be a Hilbert space, 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 $x^* \in S$, 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 the Hilbert space $\mathcal{X}$ 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,

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

Proof The proof is similar to the one of Lemma 4 in [22]. □

The following lemma can be found as Lemma 1.1 in [11]:

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 $\sum_{k\geq 1} d_k < +\infty$. Suppose 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 [11]:

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* $k \geq 1$ *we have*

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

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

Acknowledgements The authors are grateful to the editor and reviewers whose helpful comments and suggestions have led to much improvement of the paper. The research of Kewei Ding has been supported by Grant No. 2024ZYD0059 from the Central Government Guided Local Science and Technology Development Projects.

Data Availability Availability of supporting data not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval and consent to participate Not applicable.

Consent for publication Not applicable.

Human and animal rights Not applicable.

References

1. Attouch, H., Cabot, A.: Convergence rates of inertial forward-backward algorithms. SIAM J. Optimiz. **28**, 849–874 (2018)
2. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics. J. Optimiz. Theory Appl. **193**, 704–736 (2022)
3. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**, 123–175 (2018)
4. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**, 1824–1834 (2016)
5. Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping. J. Differ. Equations **261**, 5734–5783 (2016)
6. Bauschke, H.H., Bui, M.N., Wang, X.: Applying FISTA to optimization problems (with or) without minimizers. Math. Program. **184**, 349–381 (2020)
7. Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York (2017)

8. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)
9. Böhm, A., Sedlmayer, M., Csetnek, E.R., Boţ, R.I.: Two steps at a time–taking GAN training in stride with Tseng's method. SIAM J. Math. Data Sci. **4**, 750–771 (2022)
10. Boţ, R.I., Nguyen, D.K.: Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping. J. Differ. Eq. **303**, 369–406 (2021)
11. Boţ, R.I., Csetnek, E.R., Nguyen, D.K.: Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates. Math. Program. **200**, 147–197 (2023)
12. Boţ, R.I., Csetnek, E.R., Sedlmayer, M.: An accelerated minimax algorithm for convex-concave saddle point problems with nonsmooth coupling function. Comput. Optim. Appl. **86**, 925–966 (2023)
13. Boţ, R.I., Sedlmayer, M., Vuong, P.T.: A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs. J. Mach. Learn Res. **24**, 1–37 (2023)
14. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
15. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**, 253–287 (2016)
16. Chambolle, A., Dossal, C.: On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**, 968–982 (2015)
17. Chen, Y., Lan, G., Ouyang, Y.: Optimal primal-dual methods for a class of saddle point problems. SIAM J. Optimiz. **24**, 1779–1814 (2014)
18. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**, 460–479 (2013)
19. Du, S.S., Chen, J., Li, L., Xiao, L., Zhou, D.: Stochastic variance reduction methods for policy evaluation. In: International conference on machine learning pp. 1049-1058 (2017)
20. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal- dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**, 1015–1046 (2010)
21. He, B., Ma, F., Xu, S., Yuan, X.: A generalized primal-dual algorithm with improved convergence condition for saddle point problems. SIAM J. Imaging Sci. **15**, 1157–1183 (2022)
22. He, X., Hu, R., Fang, Y.P.: Fast primal-dual algorithm via dynamical system for a linearly constrained convex optimization problem. Automatica **146**, 110547 (2022)
23. He, X., Hu, R., Fang, Y.P.: Inertial accelerated primal-dual methods for linear equality constrained convex optimization problems. Numer. Algorithms **90**, 1669–1690 (2022)
24. He, X., Huang, N., Fang, Y.P.: Non-ergodic convergence rates of first-order primal-dual algorithms for saddle point problems. (2023) arXiv:2311.11274
25. He, X., Hu, R., Fang, Y.P.: A second order primal-dual dynamical system for a convex-concave bilinear saddle point problem. Appl. Math. Opt. **89**, 1–22 (2024)
26. Jiang, F., Cai, X., Wu, Z., Han, D.: Approximate first-order primal-dual algorithms for saddle point problems. Math. Comput. **90**, 1227–1262 (2021)
27. Korpelevich, G.M.: The extragradient method for finding saddle points and other problems. Ekon. Mat. Metody **12**, 747–756 (1976)
28. Kovalev, D., Gasnikov, A., Richtárik, P.: Accelerated primal-dual gradient method for smooth and convex-concave saddle-point problems with bilinear coupling. (2021) arXiv:2112.15199
29. Li, C.J., Yuan, A., Gidel, G., Jordan, MI.: Nesterov meets optimism: rate-optimal optimistic-gradient-based method for stochastic bilinearly-coupled minimax optimization. (2022) arXiv:2210.17550
30. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
31. May, R.: Asymptotic for a second-order evolution equation with convex potential and vanishing damping term, Turk. J. Math. **41**, 681–685 (2017)
32. Mokhtari, A., Ozdaglar, A.E., Pattathil, S.: Convergence rate of $O(1/k)$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. SIAM J. Optimiz. **30**, 3230–3251 (2020)
33. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(\frac{1}{k^2})$. Soviet Math. Dokl. **27**, 372–376 (1983)
34. Sebbouh, O., Dossal, C., Rondepierre, A.: Convergence rates of damped inertial dynamics under geometric conditions and perturbations. SIAM J. Optimiz. **30**, 1850–1877 (2020)
35. Su, W., Boyd, S., Candés, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. J. Mach. Learn. Res. **17**, 1–43 (2016)

36. Thekumparampil, K.K., He, N., Oh, S.: Lifted primal-dual method for bilinearly coupled smooth minimax optimization. In: International conference on artificial intelligence and statistics, pp. 4281-4308 (2022)
37. Tran-Dinh, Q.: A unified convergence rate analysis of the accelerated smoothed gap reduction algorithm. Optim. Lett. **16**, 1235–1257 (2022)
38. Zeng, X., Lei, J., Chen, J.: Accelerated first-order continuous-time algorithm for solving convex-concave bilinear saddle point problem. IFAC-PapersOnLine **53**, 7362–7367 (2020)