



# On the convergence of a class of inertial dynamical systems with Tikhonov regularization

Bo Xu<sup>1</sup> · Bo Wen<sup>2</sup>

Received: 9 June 2020 / Accepted: 30 October 2020 / Published online: 14 November 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

We consider a class of inertial second order dynamical system with Tikhonov regularization, which can be applied to solving the minimization of a smooth convex function. Based on the appropriate choices of the parameters in the dynamical system, we first show that the function value along the trajectories converges to the optimal value, and prove that the convergence rate can be faster than  $o(1/t^2)$ . Moreover, by constructing proper energy function, we prove that the trajectories strongly converges to a minimizer of the objective function of minimum norm. Finally, some numerical experiments have been conducted to illustrate the theoretical results.

**Keywords** Convex optimization · Inertial gradient system · Tikhonov regularization · Convergence analysis

## 1 Introduction

In recent years, convex optimization problems draw many researchers' attention due to its arisen in a lot of application areas, such as machine learning [10,30], statistics [18], image processing [20,32] and so on. Hence, various algorithms have been proposed for solving different structured convex optimization problems. One simple and often powerful algorithm is Nesterov accelerated gradient algorithm, whose convergence rate can be  $O(1/t^2)$ . Many accelerated algorithms based on Nesterov's accelerated technique has been proposed since then, we refer the readers to [11,19,25,26,29,31] and the reference therein for an overview of these algorithms.

---

✉ Bo Wen  
wenbohit@163.com

Bo Xu  
xubo123jy@163.com

<sup>1</sup> School of Science, Hebei University of Technology, Tianjin, People's Republic of China

<sup>2</sup> Institute of Mathematics, Hebei University of Technology, Tianjin, People's Republic of China

Most literatures consider Nesterov's accelerated method by using different numerical optimization techniques. However, differential equations are also important and efficient tools to study numerical algorithms. Recently, Su, Boyd, Candés [28] propose a class of second order differential equations to study Nesterov's accelerated gradient method, which is

$$\begin{cases} \ddot{x} + \frac{\alpha}{t}\dot{x} + \nabla\Phi(x) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0, \end{cases} \quad (1.1)$$

where  $\Phi$  is convex and differentiable, and  $\nabla\Phi$  is Lipschitz continuous,  $t_0 > 0$ . They show that this system can be seen as the continuous version of Nesterov's accelerated gradient method. In addition, they prove that the convergence rate of the function value along the trajectories of (1.1) is  $O(1/t^2)$ , if  $\alpha$  is chosen as 3, which is the same as the convergence rate of Nesterov's accelerated gradient method. Moreover, they show that 3 is the minimum constant that guarantees the convergence rate of  $O(1/t^2)$ .

Su, Boyd, Candés work [28] motivates subsequent studies on the second order differential equation (1.1), see, for example, [3,8,13–16,24,33]. Particularly, Attouch, Chbani, Peypouquet and Redont [3] establish the weak convergence of the trajectory if  $\alpha > 3$ , and they also show that the convergence rate of the objective function along the trajectory is  $o(1/t^2)$ . Later, Attouch, Chbani and Riahi [5] consider the convergence properties under the condition that  $\alpha < 3$ . They prove that the convergence rate is

$$\Phi(x(t)) - \min \Phi = O\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right).$$

In order to establish the strong convergence of the trajectory, Attouch, Chbani and Riahi [4] propose the following second order dynamical system:

$$\begin{cases} \ddot{x} + \frac{\alpha}{t}\dot{x} + \nabla\Phi(x) + \varepsilon(t)x(t) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0, \end{cases} \quad (1.2)$$

which add a Tikhonov regular term compared with system (1.1). They show that the function value along the trajectory converges to the optimal value fast, if  $\varepsilon(t)$  decreases to 0 rapidly. In addition, they establish the strong convergence of the trajectory  $x(t)$  to the element of minimum norm of  $\arg \min \Phi$ , if  $\varepsilon(t)$  tends slowly to zero. There are many other literatures considering the Tikhonov regular techniques, the readers can result the references [1,2,9,17,23].

In 2019, Attouch, Chbani and Riahi [6,7] study another differential equation:

$$\begin{cases} \ddot{x} + \gamma(t)\dot{x} + \beta(t)\nabla\Phi(x) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0, \end{cases} \quad (1.3)$$

where  $\gamma(t)$  and  $\beta(t)$  are scalar functions. They first consider the convergence properties of (1.3). Then a discretized numerical algorithm for solving structured convex composite optimization problem based on the differential equation has been proposed.

Inspired by the proof of the convergence of the trajectory of (1.1), they establish the convergence and convergence rate of the algorithm. Concretely, they obtain that the convergence rate of  $\Phi(x(t))$  is

$$\Phi(x(t)) - \min \Phi = O\left(\frac{1}{\beta(t)\Gamma(t)^2}\right),$$

where  $\Gamma(t) = p(t) \int_t^{+\infty} \frac{1}{p(u)} du$ ,  $p(t) = e^{\int_{t_0}^t \gamma(u) du}$ . In particular, if  $\gamma(t)$  is chosen as  $\frac{\alpha}{t}$ , the convergence rate becomes

$$\Phi(x(t)) - \min \Phi = O\left(\frac{1}{\beta(t)t^2}\right).$$

According to the above relation, it can be easily seen that the convergence rate of  $\Phi(x(t))$  can be faster than  $O\left(\frac{1}{t^2}\right)$ , if we choose proper  $\beta(t)$ . For the nonsmooth optimization problems, which means the objective function is not differentiable, differential equations can not be applied directly, we recommend the readers to [12,21] to see the details.

From the above literatures, we note that some work consider the strong convergence of the trajectory  $x(t)$ , the other work study the faster convergence rate of objective  $\Phi(x(t))$ . A natural question is that whether we can combine these discussions together. In this work, both the strong convergence property of the trajectory  $x(t)$  and the fast convergence rate of objective  $\Phi(x(t))$  are studied under different choice of the parameters. To this end, this paper mainly considers the following differential equation:

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) (\nabla \Phi(x(t)) + \varepsilon(t)x(t)) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0, \end{cases} \tag{1.4}$$

where  $\Phi$  is convex and differentiable,  $\nabla \Phi$  is Lipschitz continuous,  $u_0, v_0 \in \mathcal{H}, t_0 > 0$ ,  $\alpha$  is a positive parameter,  $\beta(t)$  is a time scaling parameter, and  $\varepsilon(t)x(t)$  is a Tikhonov regularization term. Throughout the whole paper, we assume that

$$H_1 \begin{cases} t_0 > 0, \varepsilon : [t_0, +\infty) \rightarrow \mathbb{R}^+ \text{ is a nonincreasing function ;} \\ \varepsilon(t) \text{ is continuously differentiable and } \lim_{t \rightarrow +\infty} \varepsilon(t) = 0; \\ \beta : [t_0, +\infty) \rightarrow \mathbb{R}^+ \text{ is a non-negative continuous function.} \end{cases}$$

Our main contributions are as follows:

- By constructing proper energy function, we first prove that the existence and uniqueness of the global solution of dynamical system (1.4);
- We establish the fast convergence rate of  $\Phi(x(t))$  and strong convergence of the trajectory  $x(t)$  of system (1.4). In details, under the condition that

$\int_{t_0}^{+\infty} t \beta(t) \varepsilon(t) dt < +\infty$ , we establish the global convergence rate of  $\Phi(x(t))$  which is

$$\Phi(x(t)) - \min \Phi = o\left(\frac{1}{t^2 \beta(t)}\right).$$

Moreover, if  $\int_{t_0}^{+\infty} \frac{\varepsilon(t)\beta(t)}{t} dt = +\infty$ , we show that the global solution  $x(t)$  of (1.4) satisfies the following ergodic convergence result:

$$\lim_{t \rightarrow \infty} \frac{1}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \|x(\tau) - p\|^2 d\tau = 0,$$

where  $p$  is the element of minimal norm of  $\arg \min \Phi$ . In addition, we prove that  $\liminf_{t \rightarrow \infty} \|x(t) - p\| = 0$ .

The rest of the paper is organized as follows: Section 2 presents some basic notation and preliminary materials. In Sect. 3, the global existence and uniqueness result is established for (1.4). In Sect. 4, we first establish the fast convergence rate of  $\Phi(x(t))$  based on the condition  $\int_{t_0}^{+\infty} t \beta(t) \varepsilon(t) dt < +\infty$ , and then show that the trajectory  $x(t)$  of (1.4) converges to a minimizer of the objective function of minimum norm. In Sect. 5, we perform some numerical experiments to illustrate our theoretical results.

## 2 Notation and preliminaries

The problems we consider in this paper are all in Hilbert space  $\mathcal{H}$ , and we denote its inner product by  $\langle \cdot, \cdot \rangle$ , the corresponding norm is denoted as  $\|\cdot\|$ .

For the real valued convex and differentiable function  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$ , the gradient of  $\Phi$  is said to be  $L_\Phi$ -Lipschitz continuous, if

$$\|\nabla \Phi(x) - \nabla \Phi(y)\| \leq L_\Phi \|x - y\|, \forall x, y \in \mathcal{H}.$$

We say that  $\Phi$  is a  $\sigma$ -strongly convex if and only if  $\Phi(\cdot) - \frac{\sigma}{2} \|\cdot\|^2$  is convex,  $\sigma > 0$ . Moreover, if  $\Phi$  is continuously differentiable, then

$$\Phi(y) \geq \Phi(x) + \langle \nabla \Phi(x), y - x \rangle + \frac{\sigma}{2} \|y - x\|^2.$$

A function  $x : [0, \infty) \rightarrow \mathcal{H}$  is called locally absolutely continuous if  $x : [0, \infty) \rightarrow \mathcal{H}$  is absolutely continuous on every compact interval, which means that there exists an integrable function  $y : [t_0, T) \rightarrow \mathcal{H}$  such that

$$x(t) = x(0) + \int_{t_0}^t y(s) ds \quad \forall t \in [t_0, T).$$

For a locally absolutely continuous function, we would like to point out the following property, which will be used in the following sections.

**Remark 2.1** Every locally absolutely continuous function  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  is differentiable almost everywhere and its derivative coincides with its distributional derivative almost everywhere and one can recover the function from its derivative  $\dot{x} = y$  by the integration formula above.

Before ending this section, we state some lemmas which will be used in our convergence analysis.

**Lemma 2.1** *Suppose that  $F : [0, +\infty) \rightarrow \mathbb{R}$  is locally absolutely continuous and bounded below and that there exist  $G \in L^1([0, +\infty))$  such that for almost every  $t \in [0, +\infty)$*

$$\frac{d}{dt} F(t) \leq G(t).$$

Then there exists  $\lim_{t \rightarrow \infty} F(t) \in \mathbb{R}$ .

Now, we will introduce an energy function we used in the paper, which is

$$W(t) = \frac{1}{2} \|\dot{x}(t)\|^2 + \beta(t) (\Phi_t(x(t)) - \min \Phi), \tag{2.1}$$

where  $\Phi_t(x) = \Phi(x) + \frac{\varepsilon(t)}{2} \|x\|^2$ .

Next, we will give two important results, which play important roles in the analysis of asymptotic behavior of system (1.4).

**Lemma 2.2** *Let  $W$  be defined by (2.1), we have*

$$\frac{dW}{dt}(t) \leq -\frac{\alpha}{t} \|\dot{x}(t)\|^2 + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi).$$

**Proof** From the definition of  $\Phi_t(x(t))$ , we immediately have

$$\nabla \Phi_t(x(t)) = \nabla \Phi(x(t)) + \varepsilon(t)x(t). \tag{2.2}$$

On the other hand, by taking the derivative of the energy function (2.1) and using the definition of  $\Phi_t(x) = \Phi(x) + \frac{\varepsilon(t)}{2} \|x\|^2$ , we have

$$\begin{aligned} \dot{W}(t) &= \langle \dot{x}(t), \ddot{x}(t) \rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi) \\ &\quad + \beta(t) \left( \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle + \dot{\varepsilon}(t) \frac{\|x(t)\|^2}{2} + \varepsilon(t) \langle x(t), \dot{x}(t) \rangle \right) \\ &= \langle \dot{x}(t), \ddot{x}(t) \rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi) \\ &\quad + \beta(t) \left( \langle \nabla \Phi(x(t)) + \varepsilon(t)x(t), \dot{x}(t) \rangle + \dot{\varepsilon}(t) \frac{\|x(t)\|^2}{2} \right). \end{aligned}$$

Combining (2.2) with the above relation, we obtain further that

$$\begin{aligned} \dot{W}(t) &= \langle \dot{x}(t), \ddot{x}(t) \rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi) \\ &\quad + \beta(t) \left( \langle \nabla \Phi_t(x(t)), \dot{x}(t) \rangle + \dot{\varepsilon}(t) \frac{\|x(t)\|^2}{2} \right). \end{aligned}$$

By rearranging terms and using the fact that  $\varepsilon(t)$  is continuously differentiable and nonincreasing, we have

$$\begin{aligned} \dot{W}(t) &= \langle \dot{x}(t), \ddot{x}(t) \rangle + \beta(t) \langle \nabla \Phi_t(x(t)), \dot{x}(t) \rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi) \\ &\quad + \beta(t) \dot{\varepsilon}(t) \frac{\|x(t)\|^2}{2} \\ &= \langle \dot{x}(t), \ddot{x}(t) + \beta(t) \nabla \Phi_t(x(t)) \rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi) \\ &\quad + \beta(t) \dot{\varepsilon}(t) \frac{\|x(t)\|^2}{2} \\ &\leq \langle \dot{x}(t), \ddot{x}(t) + \beta(t) \nabla \Phi_t(x(t)) \rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi), \end{aligned} \tag{2.3}$$

where the last inequality is from the fact that  $\dot{\varepsilon}(t) \leq 0$  and  $\beta(t) \geq 0$ . Moreover, according to system (1.4), we have

$$\ddot{x}(t) + \beta(t) \nabla \Phi_t(x(t)) = \ddot{x}(t) + \beta(t) (\nabla \Phi(x(t)) + \varepsilon(t)x(t)) = -\frac{\alpha}{t} \dot{x}(t). \tag{2.4}$$

Combining (2.3) and (2.4) together, we obtain that

$$\begin{aligned} \dot{W}(t) &\leq \left\langle \dot{x}(t), -\frac{\alpha}{t} \dot{x}(t) \right\rangle + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi) \\ &= -\frac{\alpha}{t} \|\dot{x}(t)\|^2 + \dot{\beta}(t) (\Phi_t(x(t)) - \min \Phi), \end{aligned}$$

which implies our desired conclusion immediately.

In the following, we introduce another auxiliary function

$$h_z(t) = \frac{1}{2} \|x(t) - z\|^2, \tag{2.5}$$

where  $z \in \arg \min \Phi$ , then we will give the following property of  $h_z$ .

**Lemma 2.3** *Suppose  $h_z(t)$  is defined as (2.5), then*

- (i)  $\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) \leq \|\dot{x}(t)\|^2 - \beta(t) \left( \Phi_t(x(t)) - \min \Phi - \frac{\varepsilon(t)}{2} \|z\|^2 + \frac{\varepsilon(t)}{2} \|x(t) - z\|^2 \right)$ .
- (ii)  $\sup_{t \geq t_0} \frac{|\dot{h}_z(t)|}{t} < +\infty$  if  $\sup_{t \geq t_0} \|\dot{x}(t)\| < +\infty$ .

**Proof** (i): From the definition of  $h_z(t)$ , we immediately obtain that

$$\dot{h}_z(t) = \langle \dot{x}(t), x(t) - z \rangle, \quad \ddot{h}_z(t) = \|\dot{x}(t)\|^2 + \langle \ddot{x}(t), x(t) - z \rangle. \tag{2.6}$$

Hence

$$\begin{aligned} \ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) &= \|\dot{x}(t)\|^2 + \langle \ddot{x}(t), x(t) - z \rangle + \frac{\alpha}{t} \langle \dot{x}(t), x(t) - z \rangle \\ &= \|\dot{x}(t)\|^2 + \left\langle \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t), x(t) - z \right\rangle. \end{aligned} \tag{2.7}$$

According to (1.4), we have

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) = -\beta(t) (\nabla \Phi(x(t)) + \varepsilon(t)x(t)) = -\beta(t) \nabla \Phi_t(x(t)).$$

Combining this relation with (2.7), we obtain further that

$$\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) = \|\dot{x}(t)\|^2 - \beta(t) \langle \nabla \Phi_t(x(t)), x(t) - z \rangle. \tag{2.8}$$

On the other hand, recall  $\Phi$  is convex function, from this with the definition  $\Phi_t$ , we see that  $\Phi_t$  is  $\varepsilon(t)$ -strongly convex function. Hence, we have

$$\Phi_t(z) - \Phi_t(x(t)) \geq \langle \nabla \Phi_t(x(t)), z - x(t) \rangle + \frac{\varepsilon(t)}{2} \|z - x(t)\|^2,$$

which implies that

$$\langle \nabla \Phi_t(x(t)), x(t) - z \rangle \geq \Phi_t(x(t)) - \Phi_t(z) + \frac{\varepsilon(t)}{2} \|x(t) - z\|^2. \tag{2.9}$$

Since  $z$  is a minimizer of  $\Phi$ , by (2.9) and the definition of  $\Phi_t(z) = \Phi(z) + \frac{\varepsilon(t)}{2} \|z\|^2 = \min \Phi + \frac{\varepsilon(t)}{2} \|z\|^2$ , we obtain further that

$$\langle \nabla \Phi_t(x(t)), x(t) - z \rangle \geq \Phi_t(x(t)) - \min \Phi - \frac{\varepsilon(t)}{2} \|z\|^2 + \frac{\varepsilon(t)}{2} \|x(t) - z\|^2.$$

Using this, the fact  $\beta(t) \geq 0$  from  $H_1$  and (2.8), we have

$$\begin{aligned} &\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) \\ &= \|\dot{x}(t)\|^2 - \beta(t) \langle \nabla \Phi_t(x(t)), x(t) - z \rangle \\ &\leq \|\dot{x}(t)\|^2 - \beta(t) \left( \Phi_t(x(t)) - \min \Phi - \frac{\varepsilon(t)}{2} \|z\|^2 + \frac{\varepsilon(t)}{2} \|x(t) - z\|^2 \right). \end{aligned}$$

This completes the proof of (i).

Now we prove (ii). By the definition of  $\dot{h}_z(t)$  in (2.6), the assumption that  $\sup_{t \geq t_0} \|\dot{x}(t)\| < +\infty$  and Schwartz's inequality, we obtain that

$$|\dot{h}_z(t)| \leq \|\dot{x}(t)\| \|x(t) - z\| \leq \sup_{s \geq t_0} \|\dot{x}(s)\| \|x(t) - z\|. \tag{2.10}$$

In addition,

$$\|x(t) - z\| \leq \|x(t) - x(t_0)\| + \|x(t_0) - z\| \leq (t - t_0) \sup_{s \geq t_0} \|\dot{x}(s)\| + \|x(t_0) - z\|.$$

Combining the above inequality, the assumption  $\sup_{t \geq t_0} \|\dot{x}(t)\| < +\infty$  with (2.10), we immediately deduce that

$$|\dot{h}_z(t)| \leq \sup_{s \geq t_0} \|\dot{x}(s)\| \left( (t - t_0) \sup_{s \geq t_0} \|\dot{x}(s)\| + \|x(t_0) - z\| \right) \leq \tilde{C} (1 + t),$$

where  $\tilde{C} > 0$  is a constant. Thus,  $\sup_{t \geq t_0} \frac{|\dot{h}_z(t)|}{t} < +\infty$ . This completes the proof.

### 3 Existence and uniqueness of the solution of (1.4)

In this section, we will prove the existence and uniqueness of a global solution of dynamical system (1.4). We first give the definition of a strong global solution of (1.4).

**Definition 3.1** We say that  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  is a strong global solution of (1.4), if it satisfies the following properties:

- (a)  $x, \dot{x} : [t_0, +\infty) \rightarrow \mathcal{H}$  are locally absolutely continuous, in other words, absolutely continuous on each interval  $[t_0, T]$  for  $t_0 < T < +\infty$ ;
- (b)  $\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) (\nabla \Phi(x(t)) + \varepsilon(t)x(t)) = 0$  for almost every  $t \geq t_0$ ;
- (c)  $x(t_0) = u_0$  and  $\dot{x}(t_0) = v_0$ .

We are now ready to prove the existence and uniqueness of system (1.4). We mainly use Cauchy-Lipschitz-Picard theorem for absolutely continuous trajectories (see, for example [[22], proposition 6.2.1], [[27], Theorem 54]) to establish the result. The proof is based on the idea that rewriting (1.4) as a particular first order dynamical system in a suitably chosen product space (see also [8,17]).

**Theorem 3.1** *For any initial points  $u_0, v_0 \in \mathcal{H}$ , there exists a unique  $C^2$ -global solution of the dynamical system (1.4).*

**Proof** Define  $X(t) = (x(t), \dot{x}(t))$ , and  $F : [t_0, +\infty) \times \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H}$  as

$$F(t, u, v) = \left( v, -\frac{\alpha}{t}v - \beta(t) (\nabla \Phi(u) + \varepsilon(t)u) \right). \tag{3.1}$$



Hence, from (1.4), (3.1) and the definition of  $X(t)$ , we see that (1.4) can be rewritten as a first order dynamical system, which is

$$\begin{cases} \dot{X}(t) = F(t, X(t)) = F(t, x(t), \dot{x}(t)) \\ X(t_0) = (u_0, v_0). \end{cases} \tag{3.2}$$

We will first show that  $F(t, \cdot, \cdot)$  is  $L(t)$ -Lipschitz continuous for every  $t \geq t_0$ . And the Lipschitz constant is a function of time with the property that  $L(\cdot) \in L^1_{loc}([t_0, +\infty))$ . Concretely, for every  $(u, v), (\bar{u}, \bar{v}) \in \mathcal{H} \times \mathcal{H}$ , by (3.1), we have

$$\begin{aligned} & \|F(t, u, v) - F(t, \bar{u}, \bar{v})\| \\ &= \sqrt{\|v - \bar{v}\|^2 + \left\| \frac{\alpha}{t}(\bar{v} - v) + \beta(t)(\nabla\Phi(\bar{u}) - \nabla\Phi(u) + \varepsilon(t)(\bar{u} - u)) \right\|^2}. \end{aligned}$$

Using the fact that  $(a + b)^2 \leq 2a^2 + 2b^2$  and the above formula, we have

$$\begin{aligned} & \|F(t, u, v) - F(t, \bar{u}, \bar{v})\| \\ & \leq \sqrt{\|v - \bar{v}\|^2 + 2\left\| \frac{\alpha}{t}(v - \bar{v}) \right\|^2 + 2\beta(t)^2\|\nabla\Phi(\bar{u}) - \nabla\Phi(u) + \varepsilon(t)(\bar{u} - u)\|^2} \\ & \leq \sqrt{\left(1 + 2\frac{\alpha^2}{t^2}\right)\|v - \bar{v}\|^2 + 4\beta(t)^2\|\nabla\Phi(u) - \nabla\Phi(\bar{u})\|^2 + 4\beta(t)^2\varepsilon(t)^2\|u - \bar{u}\|^2}. \end{aligned}$$

From this relation and the fact  $\nabla\Phi$  is  $L_\Phi$ -Lipschitz continuous in the assumption, we obtain further that

$$\begin{aligned} & \|F(t, u, v) - F(t, \bar{u}, \bar{v})\| \\ & \leq \sqrt{\left(1 + 2\frac{\alpha^2}{t^2}\right)\|v - \bar{v}\|^2 + (4L_\Phi^2\beta(t)^2 + 4\beta(t)^2\varepsilon(t)^2)\|u - \bar{u}\|^2} \\ & \leq \sqrt{1 + 2\frac{\alpha^2}{t^2} + 4L_\Phi^2\beta(t)^2 + 4\beta(t)^2\varepsilon(t)^2} \sqrt{\|v - \bar{v}\|^2 + \|u - \bar{u}\|^2} \\ & \leq \left(1 + \sqrt{2}\frac{\alpha}{t} + 2L_\Phi\beta(t) + 2\beta(t)\varepsilon(t)\right)\|(u, v) - (\bar{u}, \bar{v})\|, \end{aligned}$$

where the last inequality follows from the fact  $\sqrt{a^2 + b^2} \leq a + b$  if  $a \geq 0, b \geq 0$ , and  $\alpha > 0, \beta(t) \geq 0, \varepsilon(t) \geq 0$ . Define  $L(t) = 1 + \sqrt{2}\frac{\alpha}{t} + 2L_\Phi\beta(t) + 2\beta(t)\varepsilon(t)$ , then we have

$$\|F(t, u, v) - F(t, \bar{u}, \bar{v})\| \leq L(t)\|(u, v) - (\bar{u}, \bar{v})\|. \tag{3.3}$$

Hence  $F(t, \cdot, \cdot)$  is  $L(t)$ -Lipschitz continuous for every  $t \geq t_0$ . Recall that  $\frac{\alpha}{t}, \beta(t)$  and  $\varepsilon(t)$  are continuous for any  $t \geq t_0$ . Thus we see that  $L(t)$  is integrable on  $[t_0, T]$ , consequently,  $L(\cdot) \in L^1_{loc}([t_0, +\infty))$ .

Next, we will show that  $F(\cdot, u, v) \in L^1_{loc}([t_0, +\infty), \mathcal{H} \times \mathcal{H})$  for all  $u, v \in \mathcal{H}$ . Take any  $u, v \in \mathcal{H}$ , by the definition of  $F$ , for  $t_0 < T < +\infty$ , we have

$$\begin{aligned} & \int_{t_0}^T \|F(t, u, v)\| dt \\ &= \int_{t_0}^T \sqrt{\|v\|^2 + \left\| -\frac{\alpha}{t}v - \beta(t)(\nabla\Phi(u) + \varepsilon(t)u) \right\|^2} dt \\ &\leq \int_{t_0}^T \sqrt{\left(1 + \frac{2\alpha^2}{t^2}\right) \|v\|^2 + 4\beta(t)^2\|\nabla\Phi(u)\|^2 + 4\beta(t)^2\varepsilon(t)^2\|u\|^2} dt \\ &\leq \sqrt{\|v\|^2 + \|\nabla\Phi(u)\|^2 + \|u\|^2} \int_{t_0}^T \sqrt{1 + \frac{2\alpha^2}{t^2} + 4\beta(t)^2 + 4\beta(t)^2\varepsilon(t)^2} dt, \end{aligned} \tag{3.4}$$

where the first inequality follows from the fact  $\|a + b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$ , the last inequality follows from that the points  $u, v \in \mathbb{R}^n$  are fixed.

Hence, by (3.4) and the fact  $\frac{\alpha}{t}, \beta(t)$  and  $\varepsilon(t)$  are continuous for any  $t \geq t_0$ , we immediately obtain that

$$F(\cdot, u, v) \in L^1_{loc}([t_0, +\infty), \mathcal{H} \times \mathcal{H}).$$

Combining this relation with (3.3) and the result  $L(\cdot) \in L^1_{loc}([t_0, +\infty))$ , and using the Cauchy-Lipschitz-Picard theorem, we see that there exists a unique global solution of system (3.2), which implies the existence of a unique  $C^2$ -global solution of (1.4) by the Lipschitz continuity of  $\nabla\Phi$  and the continuities of  $\beta(t)$  and  $\varepsilon(t)$ . This completes the proof.

### 4 Convergence analysis of the trajectory of (1.4)

In this section, we will establish the convergence and convergence rate of the trajectory  $x(t)$  of (1.4). The proof of convergence will be casted into the following two cases.

Case 1:  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t) dt < +\infty$ . In this case, we will show in Theorem 4.1 that for any global solution trajectory of (1.4), the function  $\Phi(x(t))$  satisfies the fast convergence property

$$\Phi(x(t)) - \min \Phi = o\left(\frac{1}{t^2\beta(t)}\right).$$

Case 2:  $\int_{t_0}^{+\infty} \frac{\varepsilon(t)\beta(t)}{t} dt = +\infty$ . In this case, we will show in Theorem 4.2 that for any global solution trajectory of (1.4), the following ergodic convergence result holds

$$\lim_{t \rightarrow \infty} \frac{1}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \|x(\tau) - p\|^2 d\tau = 0,$$

where  $p$  is the element of minimal norm of  $\arg \min \Phi$ . Moreover, the strong global convergence of  $x(t)$  will be established, which is

$$\liminf_{t \rightarrow \infty} \|x(t) - p\| = 0.$$

Now we are ready to present the convergence results case by case.

**4.1 Case  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t)dt < +\infty$**

For simplicity, we set  $m = \min \Phi$ . Take any  $z \in \arg \min \Phi$ , we will introduce another auxiliary function for  $\alpha \neq 1$ , which is

$$E(t) = \frac{t^2}{(\alpha - 1)^2} \beta(t) (\Phi_t(x(t)) - m) + \frac{1}{2} \left\| x(t) - z + \frac{t}{\alpha - 1} \dot{x}(t) \right\|^2, \tag{4.1}$$

where  $\Phi_t$  is the same function as defined in (2.1). Let  $g(t) = \frac{t}{\alpha - 1}$ , then by simply computing, we immediately have

$$1 + \dot{g}(t) = \frac{\alpha}{t} g(t). \tag{4.2}$$

From this relation and the definition of  $E(t)$  in (4.1), we can rewrite  $E(t)$  as the following formula

$$E(t) = g(t)^2 \beta(t) (\Phi_t(x(t)) - m) + \frac{1}{2} \|x(t) - z + g(t) \dot{x}(t)\|^2. \tag{4.3}$$

Combining (4.3) with the definition of  $W(t)$  in (2.1),  $h_z$  in (2.5) and  $\dot{h}_z$  in (2.6), we obtain that

$$E(t) = g(t)^2 W(t) + h_z(t) + g(t) \dot{h}_z(t). \tag{4.4}$$

Next, we will prove that the global convergence rate of  $\Phi(x(t))$  is  $o\left(\frac{1}{t^2\beta(t)}\right)$ .

**Theorem 4.1** *Let  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  be a convex continuously differentiable function such that  $\arg \min \Phi$  is nonempty. Assume that  $\varepsilon(t), \beta(t)$  satisfies condition  $(H_1)$ ,  $\alpha > 1$ ,  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t)dt < +\infty$  and there exist  $b > 0$  such that  $t\dot{\beta}(t) \leq (\alpha - 3 - b)\beta(t)$ . Let  $x(\cdot)$  be a classical global solution of (1.4), consider the energy function (4.4), then*

- (i)  $\dot{E}(t) \leq \frac{1}{2(\alpha-1)} t\beta(t)\varepsilon(t) \|z\|^2$ .
- (ii)  $\int_{t_0}^{+\infty} t\beta(t) (\Phi_t(x(t)) - m) dt < +\infty$ .
- (iii) *The following fast convergence of  $\Phi(x(t))$  holds:*

$$\Phi(x(t)) - m = o\left(\frac{1}{\beta(t)t^2}\right).$$

(iv) Moreover, the trajectory  $x(\cdot)$  is bounded on  $[t_0, +\infty)$  and

$$\int_{t_0}^{+\infty} t \beta(t) \varepsilon(t) \|x(t)\|^2 dt < +\infty.$$

**Proof** We first prove (i). From the definition of  $E(t)$  in (4.4), we immediately have

$$\begin{aligned} \dot{E}(t) &= g(t)^2 \dot{W}(t) + 2g(t) \dot{g}(t) W(t) + \dot{h}_z(t) + \dot{g}(t) \dot{h}_z(t) + g(t) \ddot{h}_z(t) \\ &\leq g(t)^2 \left[ -\frac{\alpha}{t} \|\dot{x}(t)\|^2 + \dot{\beta}(t) (\Phi_t(x(t)) - m) \right] \\ &\quad + 2g(t) \dot{g}(t) \left[ \frac{1}{2} \|\dot{x}(t)\|^2 + \beta(t) (\Phi_t(x(t)) - m) \right] \\ &\quad + \dot{h}_z(t) (1 + \dot{g}(t)) + g(t) \ddot{h}_z(t), \end{aligned} \tag{4.5}$$

where the last inequality follows from the Lemma 2.2.

On the other hand, according to (4.2) and Lemma 2.3, we have

$$\begin{aligned} \dot{h}_z(t) (1 + \dot{g}(t)) + g(t) \ddot{h}_z(t) &= g(t) \left( \ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) \right) \\ &\leq g(t) \left[ \|\dot{x}(t)\|^2 - \beta(t) \left( \Phi_t(x(t)) - m - \frac{\varepsilon(t)}{2} \|z\|^2 + \frac{\varepsilon(t)}{2} \|x(t) - z\|^2 \right) \right]. \end{aligned} \tag{4.6}$$

Combining the (4.5) and (4.6), we obtain further that

$$\begin{aligned} \dot{E}(t) &\leq g(t)^2 \left[ -\frac{\alpha}{t} \|\dot{x}(t)\|^2 + \dot{\beta}(t) (\Phi_t(x(t)) - m) \right] \\ &\quad + 2g(t) \dot{g}(t) \left[ \frac{1}{2} \|\dot{x}(t)\|^2 + \beta(t) (\Phi_t(x(t)) - m) \right] \\ &\quad + g(t) \left[ \|\dot{x}(t)\|^2 - \beta(t) \left( \Phi_t(x(t)) - m - \frac{\varepsilon(t)}{2} \|z\|^2 + \frac{\varepsilon(t)}{2} \|x(t) - z\|^2 \right) \right]. \end{aligned}$$

By rearranging terms, we have

$$\begin{aligned} \dot{E}(t) &\leq g(t) \|\dot{x}(t)\|^2 \left( 1 + \dot{g}(t) - \frac{\alpha}{t} g(t) \right) + g(t) \beta(t) \frac{\varepsilon(t)}{2} \|z\|^2 \\ &\quad + (\Phi_t(x(t)) - m) g(t) (g(t) \dot{\beta}(t) + 2\dot{g}(t) \beta(t) - \beta(t)) - g(t) \beta(t) \\ &\quad \frac{\varepsilon(t)}{2} \|x(t) - z\|^2 \\ &= g(t) \beta(t) \frac{\varepsilon(t)}{2} \|z\|^2 + (\Phi_t(x(t)) - m) g(t) (g(t) \dot{\beta}(t) + 2\dot{g}(t) \beta(t) - \beta(t)) \\ &\quad - g(t) \beta(t) \frac{\varepsilon(t)}{2} \|x(t) - z\|^2, \end{aligned} \tag{4.7}$$

where the second equality follows from the fact  $g(t) = \frac{t}{\alpha-1}$  and  $1 + \dot{g}(t) = \frac{\alpha}{t} g(t)$  from (4.2).

Recall that  $g(t) = \frac{t}{\alpha-1}$  and  $1 + \dot{g}(t) = \frac{\alpha}{t}g(t)$  from (4.2), combining this with the assumption  $t\dot{\beta}(t) \leq (\alpha - 3 - b)\beta(t)$  and  $\alpha > 1$ , we obtain that

$$\begin{aligned} g(t)\dot{\beta}(t) + 2\dot{g}(t)\beta(t) - \beta(t) &= \frac{t}{\alpha-1}\dot{\beta}(t) + 2\left(\frac{\alpha}{t}g(t) - 1\right)\beta(t) - \beta(t) \\ &= \frac{t}{\alpha-1}\dot{\beta}(t) + 2\left(\frac{\alpha}{t} \cdot \frac{t}{\alpha-1} - 1\right)\beta(t) - \beta(t) \\ &= \frac{t}{\alpha-1}\dot{\beta}(t) + \frac{3-\alpha}{\alpha-1}\beta(t) \\ &\leq -b\beta(t) \\ &\leq 0. \end{aligned}$$

Thus, according to this relation, the fact that  $g(t) \geq 0, \beta(t) \geq 0, \varepsilon(t) \geq 0, \Phi_t(x(t)) - m \geq 0$  and (4.7), we obtain further that

$$\dot{E}(t) \leq \frac{1}{2}g(t)\beta(t)\varepsilon(t)\|z\|^2 - \frac{1}{2}g(t)\beta(t)\varepsilon(t)\|x(t) - z\|^2. \tag{4.8}$$

Moreover, using the fact that  $g(t), \beta(t), \varepsilon(t) \geq 0$  and  $g(t) = \frac{t}{\alpha-1}$ , we finally have

$$\dot{E}(t) \leq \frac{1}{2}g(t)\beta(t)\varepsilon(t)\|z\|^2 = \frac{1}{2(\alpha-1)}t\beta(t)\varepsilon(t)\|z\|^2, \tag{4.9}$$

which proves (i).

We now prove (ii). From the assumption  $\int_0^{+\infty} t\beta(t)\varepsilon(t) dt < +\infty$  and (4.9), we deduce that the positive part of  $\dot{E}(t)$  belongs to  $L^1(t_0, +\infty)$ . Using this and the fact that  $E$  is bounded from below, we obtain that  $E(t)$  has a limit as  $t \rightarrow +\infty$  due to Lemma 2.1. Hence, there exists  $C_1 > 0$  such that  $|E(t)| \leq C_1$ .

In addition, according to (4.7), we have

$$\begin{aligned} \dot{E}(t) + g(t)(\beta(t) - g(t)\dot{\beta}(t) - 2\dot{g}(t)\beta(t))(\Phi_t(x(t)) - m) \\ \leq \frac{1}{2}g(t)\beta(t)\varepsilon(t)\|z\|^2. \end{aligned}$$

Recall the fact that  $g(t) = \frac{t}{\alpha-1}, \beta(t) \geq 0$  and the assumption that  $t\dot{\beta}(t) \leq (\alpha - 3 - b)\beta(t), \alpha > 1$ , we obtain further that

$$\dot{E}(t) + \frac{b}{(\alpha-1)^2}t\beta(t)(\Phi_t(x(t)) - m) \leq \frac{1}{2(\alpha-1)}t\beta(t)\varepsilon(t)\|z\|^2.$$

Integrating this inequality, and using that  $\int_0^{+\infty} t\beta(t)\varepsilon(t) dt < +\infty$  and the fact  $E(t)$  is bounded from below, we have

$$\int_0^{+\infty} t\beta(t)(\Phi_t(x(t)) - m) dt < +\infty,$$

which proves (ii).

We now prove (iii). According to Lemma 2.2, we have

$$\frac{t^2}{(\alpha - 1)^2} \dot{W}(t) + \frac{t^2}{(\alpha - 1)^2} \frac{\alpha}{t} \|\dot{x}(t)\|^2 \leq \frac{t^2}{(\alpha - 1)^2} \dot{\beta}(t) (\Phi_t(x(t)) - m).$$

After integration by parts on  $(t_0, t)$ , we obtain

$$\begin{aligned} & \frac{t^2}{(\alpha - 1)^2} W(t) - \frac{t_0^2}{(\alpha - 1)^2} W(t_0) - \frac{2}{(\alpha - 1)^2} \int_{t_0}^t s W(s) ds \\ & + \frac{\alpha}{(\alpha - 1)^2} \int_{t_0}^t s \|\dot{x}(s)\|^2 ds \\ & \leq \frac{1}{(\alpha - 1)^2} \int_{t_0}^t s^2 \dot{\beta}(s) (\Phi_s(x(s)) - m) ds. \end{aligned}$$

From the definition of  $W(t)$  in (2.1), we immediately have

$$\begin{aligned} & \frac{t^2}{(\alpha - 1)^2} W(t) + \frac{1}{\alpha - 1} \int_{t_0}^t s \|\dot{x}(s)\|^2 ds \leq \frac{t_0^2}{(\alpha - 1)^2} W(t_0) \\ & + \frac{1}{(\alpha - 1)^2} \int_{t_0}^t s (s \dot{\beta}(s) + 2\beta(s)) (\Phi_s(x(s)) - m) ds. \end{aligned}$$

From the assumption  $t\dot{\beta}(t) \leq (\alpha - 3 - b)\beta(t)$ , we obtain further that

$$\begin{aligned} & \frac{t^2}{(\alpha - 1)^2} W(t) + \frac{1}{\alpha - 1} \int_{t_0}^t s \|\dot{x}(s)\|^2 ds \leq \frac{t_0^2}{(\alpha - 1)^2} W(t_0) \\ & + \frac{\alpha - 1 - b}{(\alpha - 1)^2} \int_{t_0}^t s \beta(s) (\Phi_s(x(s)) - m) ds. \end{aligned}$$

According to (ii), there exists a constant  $\tilde{C}$  such that  $\frac{\alpha - 1 - b}{(\alpha - 1)^2} \int_{t_0}^{+\infty} s \beta(s) (\Phi_s(x(s)) - m) ds < \tilde{C}$ . Then,

$$\frac{1}{\alpha - 1} \int_{t_0}^{+\infty} s \|\dot{x}(s)\|^2 ds < \frac{t_0^2}{(\alpha - 1)^2} W(t_0) + \tilde{C}.$$

From the definition of  $W(t)$  in (2.1), we have

$$\begin{aligned} & \frac{1}{\alpha - 1} \int_{t_0}^{+\infty} t W(t) dt \\ & = \frac{1}{\alpha - 1} \int_{t_0}^{+\infty} \frac{1}{2} t \|\dot{x}(t)\|^2 dt + \frac{1}{\alpha - 1} \int_{t_0}^{+\infty} t \beta(t) (\Phi_t(x(t)) - m) dt \\ & < \frac{t_0^2}{2(\alpha - 1)^2} W(t_0) + \frac{1}{2} \tilde{C} + \frac{\alpha - 1}{\alpha - 1 - b} \tilde{C}, \end{aligned}$$

which implies that

$$\int_{t_0}^{+\infty} t W(t) dt < +\infty. \tag{4.10}$$

In addition, according to the Lemma 2.2, we have

$$\begin{aligned} & \frac{d}{dt} \left( \frac{t^2}{(\alpha - 1)^2} W(t) \right) \\ &= \frac{2t}{(\alpha - 1)^2} W(t) + \frac{t^2}{(\alpha - 1)^2} \dot{W}(t) \\ &\leq \frac{2t}{(\alpha - 1)^2} \left( \frac{1}{2} \|\dot{x}(t)\|^2 + \beta(t) (\Phi_t(x(t)) - m) \right) \\ &\quad + \frac{t^2}{(\alpha - 1)^2} \left( -\frac{\alpha}{t} \|\dot{x}(t)\|^2 + \dot{\beta}(t) (\Phi_t(x(t)) - m) \right) \\ &= \frac{1}{(\alpha - 1)^2} \left( t \|\dot{x}(t)\|^2 + 2t\beta(t) (\Phi_t(x(t)) - m) - \alpha t \|\dot{x}(t)\|^2 \right. \\ &\quad \left. + t^2 \dot{\beta}(t) (\Phi_t(x(t)) - m) \right) \\ &= \frac{1}{(\alpha - 1)^2} \left( (1 - \alpha) t \|\dot{x}(t)\|^2 + t (2\beta(t) + t \dot{\beta}(t)) (\Phi_t(x(t)) - m) \right) \\ &= \frac{1}{1 - \alpha} t \|\dot{x}(t)\|^2 + \frac{1}{(\alpha - 1)^2} t (2\beta(t) + t \dot{\beta}(t)) (\Phi_t(x(t)) - m) \\ &\leq \frac{1}{(\alpha - 1)^2} t (2\beta(t) + t \dot{\beta}(t)) (\Phi_t(x(t)) - m), \end{aligned}$$

which the last inequality is due to  $\alpha > 1$ . Recall the assumption that  $t \dot{\beta}(t) \leq (\alpha - 3 - b) \beta(t)$ , we immediately have

$$\frac{d}{dt} \left( \frac{t^2}{(\alpha - 1)^2} W(t) \right) \leq \frac{\alpha - 1 - b}{(\alpha - 1)^2} t \beta(t) (\Phi_t(x(t)) - m).$$

According to (ii), we see that the positive part of  $\frac{d}{dt} \left( \frac{t^2}{(\alpha - 1)^2} W(t) \right)$  belongs to  $L^1(t_0, +\infty)$ . As a result,

$$\lim_{t \rightarrow +\infty} \frac{t^2}{(\alpha - 1)^2} W(t) \text{ exists,}$$

which implies that

$$\lim_{t \rightarrow +\infty} \frac{t^2}{(\alpha - 1)^2} W(t) = 0.$$

Otherwise, there would exist a constant  $\bar{C} > 0$  such that  $\frac{t^2}{(\alpha-1)^2} W(t) \geq \bar{C}$  for  $t$  sufficiently large, i.e.  $\frac{t}{(\alpha-1)^2} W(t) \geq \frac{\bar{C}}{t}$ . According to (4.10), we have  $\int_{t_0}^{+\infty} \frac{\bar{C}}{t} dt < +\infty$ , which leads to a contradiction. According to the definition of  $W(t)$  in (2.1) again, we have

$$\lim_{t \rightarrow +\infty} \frac{t^2}{(\alpha-1)^2} \beta(t) (\Phi_t(x(t)) - m) = 0.$$

Recall that  $\Phi_t(x) = \Phi(x(t)) + \frac{\varepsilon(t)}{2} \|x(t)\|^2$  and  $\lim_{t \rightarrow +\infty} \varepsilon(t) = 0$ . Combining these facts with the above relation, we obtain further that

$$\Phi(x(t)) - m = o\left(\frac{1}{\beta(t)t^2}\right),$$

which proves (iii).

Finally, we prove (iv). According to (4.8), we have

$$\dot{E}(t) + g(t)\beta(t)\frac{\varepsilon(t)}{2}\|x(t) - z\|^2 \leq g(t)\beta(t)\frac{\varepsilon(t)}{2}\|z\|^2. \tag{4.11}$$

By integrating (4.11) from  $t_0$  to  $\infty$ , we obtain that

$$\int_{t_0}^{+\infty} g(t)\beta(t)\frac{\varepsilon(t)}{2}\|x(t) - z\|^2 dt < +\infty, \tag{4.12}$$

which follows from the fact  $g(t) = \frac{t}{\alpha-1}$ , the assumption  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t) dt < +\infty$ , and the positive part of  $\dot{E}(t)$  belongs to  $L^1(t_0, +\infty)$ . Using again the definition of  $g(t) = \frac{t}{\alpha-1}$ , the assumption  $\alpha > 1$ , and (4.12), we immediately have

$$\int_{t_0}^{+\infty} t\beta(t)\frac{\varepsilon(t)}{2}\|x(t) - z\|^2 dt < +\infty.$$

Combining this relation with the assumption that  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t) dt < +\infty$  and the fact  $\|x(t)\|^2 \leq 2\|x(t) - z\|^2 + 2\|z\|^2$ ,  $t > 0$ ,  $\beta(t) \geq 0$ ,  $\varepsilon(t) \geq 0$ , we obtain further that

$$\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t)\|x(t)\|^2 dt < +\infty. \tag{4.13}$$

Next, we begin to establish the boundedness of the trajectory of (1.4). Recall the definition of  $E(t)$  in (4.3) and the result  $E(t)$  is bounded from the discussion in the proof of (ii), we see there exists  $C_2 > 0$  such that

$$\frac{1}{2}\|x(t) - z + g(t)\dot{x}(t)\|^2 \leq C_2. \tag{4.14}$$



Moreover, we have

$$\|x(t) - z\|^2 + 2g(t) \langle x(t) - z, \dot{x}(t) \rangle \leq 2C_2. \tag{4.15}$$

After dividing (4.15) by  $p(t) = \left(\frac{t}{t_0}\right)^\alpha$ , we obtain that

$$\frac{\|x(t) - z\|^2}{p(t)} + \frac{2g(t)}{p(t)} \langle x(t) - z, \dot{x}(t) \rangle \leq \frac{2C_2}{p(t)}.$$

Combing this relation with the definition of  $h_z(t) = \frac{1}{2}\|x(t) - z\|^2$  in (2.5) and  $\dot{h}_z(t) = \langle \dot{x}(t), x(t) - z \rangle$  from (2.6), we obtain further that

$$\frac{h_z(t)}{p(t)} + q(t) \dot{h}_z(t) \leq \frac{C_2}{p(t)}, \tag{4.16}$$

where  $q(t) = \frac{g(t)}{p(t)}$ . Using the definition of  $g(t) = \frac{t}{\alpha-1}$  and  $p(t) = \left(\frac{t}{t_0}\right)^\alpha$ , we can easily compute  $q(t)$  as  $q(t) = \frac{g(t)}{p(t)} = \frac{t}{\alpha-1} \cdot \frac{t_0^\alpha}{t^\alpha} = \frac{t_0^\alpha t^{1-\alpha}}{\alpha-1}$ . Hence, we have  $\dot{q}(t) = -\frac{1}{p(t)}$  and  $q(t)$  is bounded due to the fact  $\alpha > 1$  from the assumption. From these discussion, we can rewrite (4.16) as

$$q(t) \dot{h}_z(t) - \dot{q}(t) (h_z(t) - C_2) \leq 0,$$

dividing this equation by  $q(t)^2$ , we have

$$\frac{q(t) \dot{h}_z(t) - \dot{q}(t) (h_z(t) - C_2)}{q(t)^2} \leq 0,$$

which is equivalent to

$$\frac{d}{dt} \left( \frac{h_z(t) - C_2}{q(t)} \right) \leq 0.$$

Hence, by integrating the above inequality from  $t_0$  to  $t$ , we see that there exists  $C_3 > 0$  such that

$$h_z(t) \leq C_3 (1 + q(t)).$$

Note that  $q(t)$  is bounded due to the fact  $\alpha > 1$  from the assumption, combining this with the definition of  $h_z(t) = \frac{1}{2}\|x(t) - z\|^2$ , we immediately obtain that  $x(t)$  is bounded. This completes the proof.

**Remark 4.1** From Theorem 4.1, we see that if  $\beta(t) = 1$ , then we have

$$\Phi(x(t)) - m = o\left(\frac{1}{t^2}\right),$$

which is just the result obtained by Attouch, Chbani, Riahi [4]. Furthermore, the assumption  $t\dot{\beta}(t) \leq (\alpha - 3 - b)\beta(t)$  and the assumption  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t)dt < +\infty$  in Theorem 4.1 reduced to  $\alpha > 3$  and  $\int_{t_0}^{+\infty} t\varepsilon(t)dt < +\infty$  in [4]. Hence our results are more general.

**Remark 4.2** From Theorem 4.1, we see that if  $\beta(t) = t$ , then we have

$$\Phi(x(t)) - m = o\left(\frac{1}{t^3}\right).$$

Furthermore, the assumption  $t\dot{\beta}(t) \leq (\alpha - 3 - b)\beta(t)$  and the assumption  $\int_{t_0}^{+\infty} t\beta(t)\varepsilon(t)dt < +\infty$  in Theorem 4.1 reduced to  $\alpha \geq 4 + b$  and  $\int_{t_0}^{+\infty} t^2\varepsilon(t)dt < +\infty$ , respectively. This also shows that the convergence rate of the function value is faster than  $O\left(\frac{1}{t^2}\right)$  if  $\alpha > 3$ .

**4.2 Case  $\int_{t_0}^{+\infty} \frac{\varepsilon(t)\beta(t)}{t}dt = +\infty$**

For each  $\varepsilon > 0$ , we use  $x_\varepsilon$  to denote the unique solution of the strongly convex minimization problem

$$x_\varepsilon = \arg \min_{x \in \mathcal{H}} \left\{ \Phi(x) + \frac{\varepsilon}{2} \|x\|^2 \right\}.$$

From the first order optimality condition, we immediately have

$$\nabla\Phi(x_\varepsilon) + \varepsilon x_\varepsilon = 0.$$

Let us recall the Tikhonov approximation curve,  $\varepsilon \mapsto x_\varepsilon$ , which satisfies the well-known strong convergence property:

$$\lim_{\varepsilon \rightarrow 0} x_\varepsilon = p, \tag{4.17}$$

where  $p$  is the element of minimal norm of the closed convex nonempty set  $\arg \min \Phi$ . Moreover, by the monotonicity property of  $\nabla\Phi$ , and  $\nabla\Phi(p) = 0, \nabla\Phi(x_\varepsilon) = -\varepsilon x_\varepsilon$ , we have

$$\langle x_\varepsilon - p, -\varepsilon x_\varepsilon \rangle \geq 0,$$

which, after dividing by  $\varepsilon > 0$ , and by Cauchy-Schwarz inequality gives

$$\|x_\varepsilon\| \leq \|p\| \quad \text{for all } \varepsilon > 0. \tag{4.18}$$

**Theorem 4.2** Let  $\Phi : \mathcal{H} \rightarrow \mathbb{R}$  be a convex continuously differentiable function such that  $\arg \min \Phi$  is nonempty. Suppose that  $\varepsilon(t), \beta(t)$  satisfies condition  $(H_1)$ ,  $\beta(t)$  is

a nonincreasing function such that  $\int_{t_0}^{+\infty} \frac{\varepsilon(t)\beta(t)}{t} dt = +\infty$  and  $\alpha > 1$  hold. Let  $x(\cdot)$  be a classical global solution of (1.4). Then  $\liminf_{t \rightarrow \infty} \|x(t) - p\| = 0$ , where  $p$  is the element of minimal norm of  $\arg \min \Phi$ . Moreover, the ergodic convergence property holds, which is

$$\lim_{t \rightarrow \infty} \frac{1}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \|x(\tau) - p\|^2 d\tau = 0.$$

**Proof** From Lemma 2.2, we have

$$\frac{dW(t)}{dt} \leq -\frac{\alpha}{t} \|\dot{x}(t)\|^2 + \dot{\beta}(t) (\Phi_t(x(t)) - m).$$

According to the assumption that  $\beta(t)$  is a nonincreasing function, we have

$$\frac{dW(t)}{dt} \leq -\frac{\alpha}{t} \|\dot{x}(t)\|^2.$$

Hence,  $W(t)$  is nonincreasing, and  $\lim_{t \rightarrow +\infty} W(t)$  exists in  $\mathbb{R}$ . Then, by the definition of  $W(t)$  in (2.1), we obtain further that  $\sup_{t \geq t_0} \|\dot{x}(t)\| < +\infty$  and that

$$\int_{t_0}^{+\infty} \frac{\|\dot{x}(t)\|^2}{t} dt \leq \frac{1}{\alpha} \left( W(t_0) - \lim_{t \rightarrow +\infty} W(t) \right) < +\infty. \tag{4.19}$$

Now, we introduce an auxiliary function  $h_p(t)$ , which is defined by

$$h_p(t) = \frac{1}{2} \|x(t) - p\|^2, \tag{4.20}$$

where  $p$  is the element of minimal norm of  $\arg \min \Phi$ . By taking the derivative and second derivative of the  $h_p(t)$ , we have

$$\dot{h}_p(t) = \langle \dot{x}(t), x(t) - p \rangle, \quad \ddot{h}_p(t) = \|\dot{x}(t)\|^2 + \langle \ddot{x}(t), x(t) - p \rangle. \tag{4.21}$$

Hence, we deduce that

$$\ddot{h}_p(t) + \frac{\alpha}{t} \dot{h}_p(t) = \|\dot{x}(t)\|^2 + \left\langle \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t), x(t) - p \right\rangle. \tag{4.22}$$

Moreover, recall the definition  $\Phi_t(x) = \Phi(x) + \frac{\varepsilon(t)}{2} \|x\|^2$  in (2.1), from this and the assumption  $\varepsilon(t) \geq 0$ , we see that  $\Phi_t$  is strongly convex with modulus  $\varepsilon(t)$ . Then we have

$$\Phi_t(p) \geq \Phi_t(x(t)) + \langle \nabla \Phi_t(x(t)), p - x(t) \rangle + \frac{\varepsilon(t)}{2} \|p - x(t)\|^2.$$

Combining this relation with system (1.4), we obtain further that

$$\Phi_t(p) \geq \Phi_t(x(t)) + \frac{1}{\beta(t)} \left\langle -\ddot{x}(t) - \frac{\alpha}{t} \dot{x}(t), p - x(t) \right\rangle + \frac{\varepsilon(t)}{2} \|p - x(t)\|^2.$$

From this relation and the definition of  $h(p)$  in (4.20), we have

$$\frac{1}{\beta(t)} \left\langle x(t) - p, \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\rangle + \varepsilon(t) h_p(t) \leq \Phi_t(p) - \Phi_t(x(t)). \quad (4.23)$$

By the definition of  $x_\varepsilon$  and  $\Phi_t$ , we immediately get

$$\Phi_t(x_\varepsilon(t)) = \Phi(x_\varepsilon(t)) + \frac{\varepsilon(t)}{2} \|x_\varepsilon(t)\|^2 \leq \Phi(x(t)) + \frac{\varepsilon(t)}{2} \|x(t)\|^2 = \Phi_t(x(t)).$$

Combining (4.23) with the above relation, we obtain further that

$$\frac{1}{\beta(t)} \left\langle x(t) - p, \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\rangle + \varepsilon(t) h_p(t) \leq \Phi_t(p) - \Phi_t(x_\varepsilon(t)). \quad (4.24)$$

Since  $p$  is the element of minimal norm of  $\arg \min \Phi$ , we have  $\Phi(p) \leq \Phi(x_\varepsilon(t))$ . Using this and the definition of  $\Phi_t$ , we obtain that

$$\begin{aligned} \Phi_t(p) - \Phi_t(x_\varepsilon(t)) &= \Phi(p) + \frac{\varepsilon(t)}{2} \|p\|^2 - \Phi(x_\varepsilon(t)) - \frac{\varepsilon(t)}{2} \|x_\varepsilon(t)\|^2 \\ &\leq \frac{\varepsilon(t)}{2} \left( \|p\|^2 - \|x_\varepsilon(t)\|^2 \right). \end{aligned} \quad (4.25)$$

Combining (4.24) with (4.25) together, we obtain further that

$$\frac{1}{\beta(t)} \left\langle x(t) - p, \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\rangle + \varepsilon(t) h_p(t) \leq \frac{\varepsilon(t)}{2} \left( \|p\|^2 - \|x_\varepsilon(t)\|^2 \right).$$

Multiply both sides of the above formula by  $\beta(t)$ , we have

$$\left\langle x(t) - p, \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) \right\rangle + \beta(t) \varepsilon(t) h_p(t) \leq \frac{\beta(t) \varepsilon(t)}{2} \left( \|p\|^2 - \|x_\varepsilon(t)\|^2 \right). \quad (4.26)$$

Combining (4.22) with (4.26) together, we obtain that

$$\ddot{h}_p(t) + \frac{\alpha}{t} \dot{h}_p(t) + \beta(t) \varepsilon(t) h_p(t) \leq \|\dot{x}(t)\|^2 + \frac{\beta(t) \varepsilon(t)}{2} \left( \|p\|^2 - \|x_\varepsilon(t)\|^2 \right). \quad (4.27)$$

On the other hand, by simply computing, we have

$$\ddot{h}_p(t) + \frac{\alpha}{t} \dot{h}_p(t) = \frac{1}{t^\alpha} \frac{d}{dt} (t^\alpha \dot{h}_p(t)).$$

Hence, from the above relation and (4.27), we obtain that

$$\beta(t) \varepsilon(t) h_p(t) \leq \|\dot{x}(t)\|^2 + \frac{\beta(t) \varepsilon(t)}{2} \left( \|p\|^2 - \|x_\varepsilon(t)\|^2 \right) - \frac{1}{t^\alpha} \frac{d}{dt} \left( t^\alpha \dot{h}_p(t) \right).$$

Dividing both sides of the above formula by  $t$ , we deduce that

$$\frac{\beta(t) \varepsilon(t) h_p(t)}{t} \leq \frac{\|\dot{x}(t)\|^2}{t} + \frac{\beta(t) \varepsilon(t)}{2t} \left( \|p\|^2 - \|x_\varepsilon(t)\|^2 \right) - \frac{1}{t^{\alpha+1}} \frac{d}{dt} \left( t^\alpha \dot{h}_p(t) \right).$$

Define  $\delta(t) = \frac{1}{2} (\|p\|^2 - \|x_\varepsilon(t)\|^2)$ , from the assumption  $\lim_{t \rightarrow \infty} \varepsilon(t) = 0$  and (4.17),(4.18) we see that  $\lim_{t \rightarrow \infty} \delta(t) = 0$ . Moreover,

$$\frac{\beta(t) \varepsilon(t) h_p(t)}{t} \leq \frac{\|\dot{x}(t)\|^2}{t} + \frac{\beta(t) \varepsilon(t) \delta(t)}{t} - \frac{1}{t^{\alpha+1}} \frac{d}{dt} \left( t^\alpha \dot{h}_p(t) \right).$$

By rearranging terms, we have

$$\frac{\beta(t) \varepsilon(t)}{t} (h_p(t) - \delta(t)) \leq \frac{\|\dot{x}(t)\|^2}{t} - \frac{1}{t^{\alpha+1}} \frac{d}{dt} \left( t^\alpha \dot{h}_p(t) \right). \tag{4.28}$$

By integrating (4.28) on  $[t_0, t]$ , there exists  $C_4 > 0$  such that

$$\int_{t_0}^t \frac{\varepsilon(s) \beta(s)}{s} (h_p(s) - \delta(s)) ds \leq C_4 - \int_{t_0}^t \frac{1}{s^{\alpha+1}} \frac{d}{ds} \left( s^\alpha \dot{h}_p(s) \right) ds, \tag{4.29}$$

which follows from (4.19).

Next, we begin to analyze the right terms in the above formula, i.e.,  $\int_{t_0}^t \frac{1}{s^{\alpha+1}} \frac{d}{ds} \left( s^\alpha \dot{h}_p(s) \right) ds$ . According to the integration rule, we have

$$\begin{aligned} & \int_{t_0}^t \frac{1}{s^{\alpha+1}} \frac{d}{ds} \left( s^\alpha \dot{h}_p(s) \right) ds \\ &= \left( \frac{1}{s} \dot{h}_p(s) \right) \Big|_{t_0}^t + (\alpha + 1) \int_{t_0}^t \frac{1}{s^2} \dot{h}_p(s) ds \\ &= \frac{1}{t} \dot{h}_p(t) - \frac{1}{t_0} \dot{h}_p(t_0) + \frac{\alpha + 1}{t^2} h_p(t) - \frac{\alpha + 1}{t_0^2} h_p(t_0) + 2(\alpha + 1) \int_{t_0}^t \frac{1}{s^3} h_p(s) ds \\ &= C_5 + \frac{1}{t} \dot{h}_p(t) + \frac{\alpha + 1}{t^2} h_p(t) + 2(\alpha + 1) \int_{t_0}^t \frac{1}{s^3} h_p(s) ds, \end{aligned}$$

where  $C_5$  is other constant. Recall the definition of  $h_p(t)$  in (4.20), we see that  $h_p(t)$  is nonnegative. From this with the fact  $\alpha > 1$ , we have

$$\int_{t_0}^t \frac{1}{s^{\alpha+1}} \frac{d}{ds} (s^\alpha \dot{h}_p(s)) ds \geq C_5 + \frac{1}{t} \dot{h}_p(t).$$

Combining the above formula with (4.29), we have

$$\int_{t_0}^t \frac{\varepsilon(s)\beta(s)}{s} (h_p(s) - \delta(s)) ds \leq C_4 - C_5 - \frac{1}{t} \dot{h}_p(t) \leq C_6 + \frac{1}{t} |\dot{h}_p(t)|, \tag{4.30}$$

where  $C_6$  is other constant.

From the fact  $\sup_{t \geq t_0} \|\dot{x}(t)\| < +\infty$ , similar to Lemma 2.3 (ii), we have  $\sup_{t \geq t_0} \frac{|\dot{h}_p(t)|}{t} < +\infty$ . Using this result and (4.30), we obtain that there exists another constant  $\bar{C} > 0$  such that

$$\int_{t_0}^t \frac{\varepsilon(s)\beta(s)}{s} (h_p(s) - \delta(s)) ds \leq \bar{C}. \tag{4.31}$$

Since  $\int_{t_0}^{+\infty} \frac{\varepsilon(t)\beta(t)}{t} dt = +\infty$  from the assumption, by (4.31), we obtain further that

$$\liminf_{t \rightarrow \infty} (h_p(t) - \delta(t)) \leq 0.$$

Note that  $\lim_{t \rightarrow \infty} \delta(t) = 0$ , hence,  $\liminf_{t \rightarrow \infty} h_p(t) = 0$ , which implies that  $\liminf_{t \rightarrow \infty} \|x(t) - p\| = 0$ . This proves the strong convergence of the trajectory  $x(t)$ .

In the following, we will prove the trajectory  $x(t)$  is ergodically convergent to the solution with minimal norm of the solution of (1.4). Note that

$$\begin{aligned} & \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} h_p(\tau) d\tau \\ &= \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} (h_p(\tau) - \delta(\tau)) d\tau + \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \delta(\tau) d\tau \\ &\leq \bar{C} + \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \delta(\tau) d\tau, \end{aligned}$$

where the last inequality follows from (4.31). Dividing both sides of the above formula by  $\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau$ , then we have

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \frac{1}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} h_p(\tau) d\tau \\ & \leq \limsup_{t \rightarrow \infty} \left( \frac{\bar{C}}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} + \frac{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \delta(\tau) d\tau}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \right) \\ & \leq \limsup_{t \rightarrow \infty} \frac{\bar{C}}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} + \limsup_{t \rightarrow \infty} \frac{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \delta(\tau) d\tau}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \\ & = 0, \end{aligned}$$

where the first inequality follows from the fact  $\varepsilon(t) \geq 0, \beta(t) \geq 0$ , and the last equality follows from the assumption that  $\lim_{t \rightarrow \infty} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau = +\infty$  and the fact that  $\lim_{t \rightarrow \infty} \delta(t) = 0$ . Then, by the definition of  $h_p$ , we have

$$\limsup_{t \rightarrow \infty} \frac{1}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \|x(\tau) - p\|^2 d\tau \leq 0. \tag{4.32}$$

Since, all the terms in the left side of (4.32) are nonnegative, we obtain further that

$$\lim_{t \rightarrow \infty} \frac{1}{\int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\varepsilon(\tau)\beta(\tau)}{\tau} \|x(\tau) - p\|^2 d\tau = 0. \tag{4.33}$$

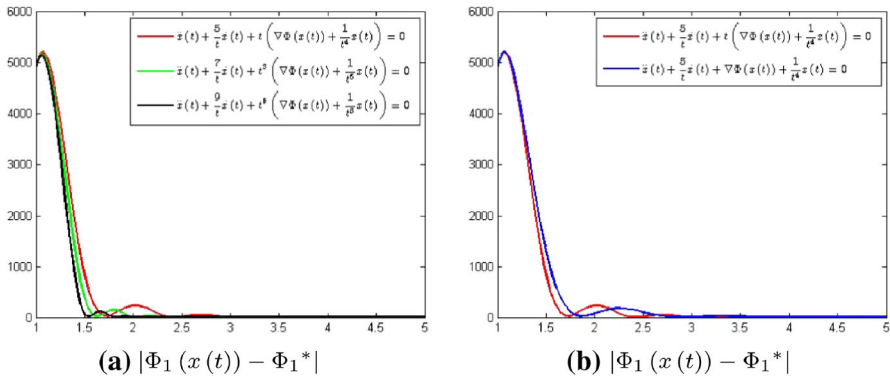
This completes the proof.

### 5 Numerical experiments

In this section, we perform numerical experiments to illustrate our theoretical results of dynamical system (1.4). All the experiments are performed by Matlab 2014b on a 64-bit Thinkpad laptop with an Intel(R) Core(TM) i7-6600U CPU (2.60GHz) and 12GB of RAM.

In our numerical tests, we consider three optimization problems: the first two examples are two dimensional strongly convex problem and convex problem respectively, the third is a convex and twice continuously differentiable one-dimensional problem and the minimizer is not unique, this example comes from reference [17]. We use Runge Kutta 4-5 adaptive method to solve them.

The first two examples are mainly to emphasize the fast convergence rate of the function value (Theorem 4.1), and the third example is to show the strong convergence of the trajectory (Theorem 4.2). A detailed description is given below.



**Fig. 1** Error analysis with different parameters in dynamical system (1.4) for a strong convex objective function  $\Phi_1$ . The red, green, black and blue line correspond to the three choices,  $\alpha = 5, \beta(t) = t, \varepsilon(t) = 1/t^4$ ;  $\alpha = 7, \beta(t) = t^3, \varepsilon(t) = 1/t^6$ ;  $\alpha = 9, \beta(t) = t^5, \varepsilon(t) = 1/t^8$ ;  $\alpha = 5, \beta(t) = 1, \varepsilon(t) = 1/t^4$ , respectively

In the next two subsections, we choose  $b \in (0, 1)$  and  $(\alpha, \beta(t), \varepsilon(t)) = (5, t, \frac{1}{t^4})$ ,  $(\alpha, \beta(t), \varepsilon(t)) = (7, t^3, \frac{1}{t^6})$ ,  $(\alpha, \beta(t), \varepsilon(t)) = (9, t^5, \frac{1}{t^8})$  respectively. All the choices of  $\alpha, \beta(t), \varepsilon(t)$  satisfy the assumptions in Theorem 4.1. Hence, by Theorem 4.1, the function value along the trajectory is convergent fast.

### 5.1 Strongly convex function

In this subsection, we consider the strongly convex optimization problem:

$$\min \Phi_1(x_1, x_2) = 2x_1^2 + 5x_2^2 - 4x_1 + 10x_2 + 7.$$

By simply computing, we obtain that  $\nabla\Phi_1(x_1, x_2) = (4x_1 - 4, 10x_2 + 10)^T$  and  $x^* = (1, -1)^T$  is the unique minimizer of  $\Phi_1$ , hence the optimal value is  $\Phi_1^* = \Phi_1(1, -1) = 0$ .

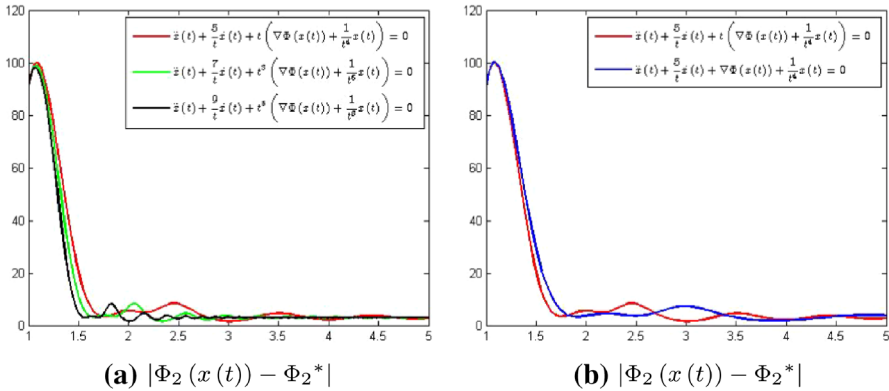
To illustrate the fast convergence rate of  $\Phi(x(t))$ , we plot in Fig. 1 the trajectory of  $|\Phi_1(x(t)) - \Phi_1^*|$  versus the time  $t$ , the horizontal axis represents  $t$ , the initial point is chosen as  $u_0 = v_0 = (-5, 30)^T$ . According to Fig. 1a, we see that  $\Phi_1(x(t))$  converges to  $\Phi_1^*$  fast for all the choices of  $\alpha, \beta(t)$  and  $\varepsilon(t)$ . Fig. 1b shows the performance of  $|\Phi_1(x(t)) - \Phi_1^*|$  under the choice of  $\alpha = 5, \beta(t) = t, \varepsilon(t) = 1/t^4$  and the case  $\alpha = 5, \beta(t) = 1, \varepsilon(t) = 1/t^4$ , where the latter choice is from [4]. We see from Fig. 1b that the choice  $\beta(t) = t$  in (1.4) are comparable with  $\beta(t) = 1$ .

### 5.2 Convex function

In this subsection, we consider convex optimization problem:

$$\min \Phi_2(x_1, x_2) = x_1^4 + 5x_2^2 - 4x_1 - 10x_2 + 8.$$





**Fig. 2** Error analysis with different parameters in dynamical system (1.4) for a convex objective function  $\Phi_2$ . The red, green, black and blue line correspond to the three choices,  $\alpha = 5, \beta(t) = t, \varepsilon(t) = 1/t^4$ ;  $\alpha = 7, \beta(t) = t^3, \varepsilon(t) = 1/t^6$ ;  $\alpha = 9, \beta(t) = t^5, \varepsilon(t) = 1/t^8$ ;  $\alpha = 5, \beta(t) = 1, \varepsilon(t) = 1/t^4$ , respectively

We can easily deduce that  $\nabla \Phi_2(x_1, x_2) = (4x_1^3 - 4, 10x_2 - 10)^T$  and  $x^* = (1, 1)^T$  is the minimizer of  $\Phi_2$ , thus the optimal value is  $\Phi_2^* = \Phi_2(1, 1) = 0$ .

The computational results are presented in Fig. 2. We plot  $|\Phi_1(x(t)) - \Phi_2^*|$  versus the time  $t$  in the following figures, the horizontal axis represents  $t$ , and the initial point is chosen as  $u_0 = v_0 = (-1, 5)^T$ . From Fig. 2a, we see that  $\Phi_2(x(t))$  converges to  $\Phi_2^*$  fast for all the choices of  $\alpha, \beta(t)$  and  $\varepsilon(t)$ . Figure 2b shows the comparison between the case  $\alpha = 5, \beta(t) = t, \varepsilon(t) = 1/t^4$  and the case  $\alpha = 5, \beta(t) = 1, \varepsilon(t) = 1/t^4$ , where the latter case is from [4]. From the numerical results, we see that the convergence rate of  $\Phi_2(x(t))$  are comparable under both choices of  $\alpha, \beta(t), \varepsilon(t)$ .

### 5.3 One-dimensional function

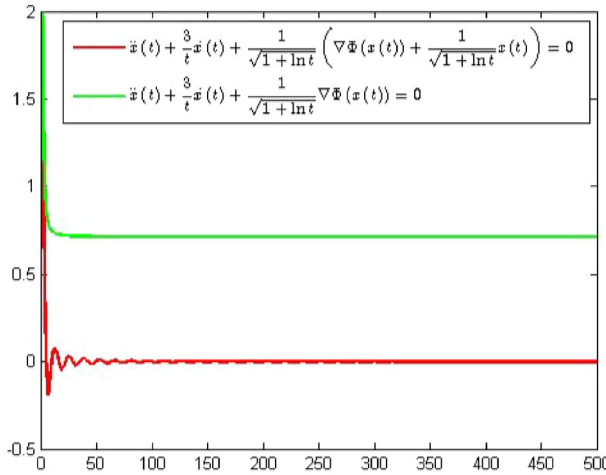
In this subsection, we conduct numerical experiments to illustrate the influence of Tikhonov regularization on the strong convergence of the trajectory  $x(t)$ . We consider  $(\alpha, \beta(t), \varepsilon(t)) = (3, \frac{1}{\sqrt{1+\ln t}}, \frac{1}{\sqrt{1+\ln t}})$  and  $(\alpha, \beta(t), \varepsilon(t)) = (3, \frac{1}{\sqrt{1+\ln t}}, 0)$  respectively, and the choice of  $(\alpha, \beta(t), \varepsilon(t)) = (3, \frac{1}{\sqrt{1+\ln t}}, \frac{1}{\sqrt{1+\ln t}})$  satisfies the assumptions in Theorem 4.2.

The optimization problem we consider in this part is as follows:

$$\min \Phi : \mathbb{R} \rightarrow \mathbb{R}, \quad \Phi(x) = \begin{cases} -(x + 1)^3, & \text{if } x < -1 \\ 0, & \text{if } -1 \leq x \leq 1 \\ (x - 1)^3, & \text{if } x > 1 \end{cases}$$

By easily computing, we can deduce that  $\arg \min \Phi = [-1, 1]$  and  $x^* = 0$  is its minimum norm solution.

Our computational results are presented in Fig. 3. We plot the trajectory  $x(t)$  generated by (1.4) versus the time  $t$  in the following figure, the horizontal axis represents  $t$ . We see from the figure that  $x(t)$  generated by (1.4) with the choice



**Fig. 3** The red line shows the trajectories of the dynamical system with Tikhonov regularization  $\varepsilon(t) = \frac{1}{\sqrt{1+\ln t}}$  are approaching the minimum norm solution  $x^* = 0$ ; the green line shows the trajectories of the dynamical system without Tikhonov regularization are approaching the optimal solution, but not the minimum norm solution

$(\alpha, \beta(t), \varepsilon(t)) = \left(3, \frac{1}{\sqrt{1+\ln t}}, \frac{1}{\sqrt{1+\ln t}}\right)$  converges to the minimum norm solution  $x^* = 0$ , which conforms with our theory. However, the trajectory  $x(t)$  under the case  $(\alpha, \beta(t), \varepsilon(t)) = \left(3, \frac{1}{\sqrt{1+\ln t}}, 0\right)$  (without the Tikhonov regularization) converges to the optimal solution, but not the minimum norm solution.

### 6 Conclusion, perspective

In this paper, we mainly study the convergence behavior of a second order gradient system with Tikhonov regularization (1.4). We first prove the existence and uniqueness of the  $C^2$ -global solution of (1.4). Next, under the assumption  $\int_{t_0}^{+\infty} t \beta(t) \varepsilon(t) dt < +\infty$ , we establish the global convergence of  $\Phi(x(t))$  to the optimal value of  $\Phi$ . Moreover, we show that the convergence rate of  $\Phi(x(t))$  to  $\min \Phi$  is  $o(1/t^2\beta(t))$ , which can be faster than  $o(1/t^2)$ . In the case  $\int_{t_0}^{+\infty} \frac{\varepsilon(t)\beta(t)}{t} dt = +\infty$ , by constructing proper energy function, we show that the trajectory  $x(t)$  strongly converges to  $p$ , where  $p$  is the element of minimal norm of  $\arg \min \Phi$ . In addition, we also prove the ergodic convergence of  $x(t)$ . Finally, we conduct some numerical experiments to illustrate the theoretical results.

At the end of this paper, we would like to list some possible directions of future research related to the dynamical system (1.4):

- (i) A natural direction is to propose some proper numerical algorithms via time discretization of (1.4). Furthermore, one can investigate their theoretical convergence properties, and confirm them with numerical experiments;

- (ii) One can also consider (1.4) endowed with an additional Hessian driven damping, see for example [8,17];
- (iii) Another direction is to consider the non-smooth optimization problems, which mean the objective functions are not differentiable, then we can not apply (1.4) directly. One can use the monotone inclusion to solve it, see for example [8,12,21].

**Acknowledgements** The authors would like to thank the editor and anonymous reviewers for their insight and helpful comments and suggestions which improve the quality of the paper. This work is also supported in part by NSFC 11801131, Natural Science Foundation of Hebei Province (Grant No. A2019202229), Science and Technology Project of Hebei Education Department (Grant No. QN2018101).

## References

1. Attouch, H., Cominetti, R.: A dynamical approach to convex minimization coupling approximation with the steepest descent method. *J. Differ. Equ.* **128**(2), 519–540 (1996)
2. Attouch, H., Czarnecki, M.O.: Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria. *J. Differ. Equ.* **179**(1), 278–310 (2002)
3. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.* **168**(1–2), 123–175 (2018)
4. Attouch, H., Chbani, Z., Riahi, H.: Combining fast inertial dynamics for convex optimization with Tikhonov regularization. *J. Math. Anal. Appl.* **457**(2), 1065–1094 (2018)
5. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterovs accelerated gradient method in the subcritical case  $\alpha \leq 3$ , ESAIM: control. *Optim. Calculus Var.* **25**, 2 (2019)
6. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via time scaling of damped inertial gradient dynamics, <https://hal.archives-ouvertes.fr/hal-02138954>
7. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. *SIAM J. Optim.* **29**(3), 2227–2256 (2019)
8. Attouch, H., Peypouquet, J., Redont, P.: Fast convex minimization via inertial dynamics with Hessian driven damping. *J. Differ. Equ.* **261**(10), 5734–5783 (2016)
9. Bolte, J.: Continuous gradient projection method in Hilbert spaces. *J. Optim. Theory Appl.* **119**(2), 235–259 (2003)
10. Bach, F.: Learning with submodular functions: a convex optimization perspective. *Found. Trends Mach. Learn.* **6**(2–3), 145–373 (2013)
11. Becker, S., Bobin, J., Cands, E.J.: NESTA: a fast and accurate first-order method for sparse recovery. *SIAM J. Imag. Sci.* **4**(1), 1–39 (2011)
12. Bot, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. *SIAM J. Control Optim.* **54**(3), 1423–1443 (2016)
13. Bot, R.I., Csetnek, E.R.: Approaching nonsmooth nonconvex optimization problems through first order dynamical systems with hidden acceleration and Hessian driven damping terms. *Set-Valued Variation. Anal.* **26**(2), 227–245 (2018)
14. Bot, R.I., Csetnek, E.R.: A second-order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities. *Optimization* **68**(7), 1265–1277 (2019)
15. Bot, R.I., Csetnek, E.R., László, S.C.: Second-order dynamical systems with penalty terms associated to monotone inclusions. *Anal. Appl.* **16**(05), 601–622 (2018)
16. Bot, R.I., Csetnek, E.R., László, S.C.: A second-order dynamical approach with variable damping to nonconvex smooth minimization. *Appl. Anal.* **99**(3), 361–378 (2020)
17. Bot, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian driven damping. *Math. Program.* (2020). <https://doi.org/10.1007/s10107-020-01528-8>
18. Bach, F., Jenatton, R., Mairal, J., Obozinski, G.: Structured sparsity through convex optimization. *Stat. Sci.* **27**(4), 450–468 (2012)
19. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2**(1), 183–202 (2009)
20. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* **158**(2), 460–479 (2013)

21. Csetnek, E.R.: Continuous dynamics related to monotone inclusions and non-smooth optimization problems. *Set Valued Variation. Anal.* (2020). <https://doi.org/10.1007/s11228-020-00548-y>
22. Haraux, A.: *Systèmes dynamiques dissipatifs et applications*, Rech. Math. Appl. **17**, Masson, Paris. (1991)
23. May, R.: On the strong convergence of the gradient projection algorithm with Tikhonov regularizing term, arXiv preprint [arXiv:1910.07873](https://arxiv.org/abs/1910.07873), (2019)
24. Muehlebach, M., Jordan, M. I.: A dynamical systems perspective on Nesterov acceleration, arXiv preprint [arXiv:1905.07436](https://arxiv.org/abs/1905.07436), (2019)
25. Mitchell, D., Ye, N., Sterck, H.D.: Nesterov acceleration of alternating least squares for canonical tensor decomposition. *Numer. Linear Algebra Appl.* (2020). <https://doi.org/10.1002/nla.2297>
26. Nesterov, Y.: A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ . *Soviet Math. Doklady* **27**(2), 372–376 (1983)
27. Sontag, E.D.: *Mathematical Control Theory Deterministic Finite-Dimensional Systems*. Texts in Applied Mathematics, vol. 6, 2nd edn. Springer, New York (1998)
28. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterovs accelerated gradient method: theory and insights. *J. Mach. Learn. Res.* **17**, 1–43 (2016)
29. Sutskever, I., Martens, J., Dahl, G., et al.: On the importance of initialization and momentum in deep learning. In: *International Conference on Machine Learning*, pp. 1139–1147 (2013)
30. Sra, S., Nowozin, S., Wright, S.J.: *Optimization for Machine Learning*. Mit Press, Cambridge (2012)
31. Tseng, P.: On accelerated proximal gradient methods for convex-concave optimization, <http://pages.cs.wisc.edu/~brecht/cs726docs/Tseng.APG.pdf> (2008). Accessed 2019
32. Wright, J., Ganesh, A., Rao, S.: Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. *Advances in Neural Information Processing Systems* **22**, 2080–2088 (2009)
33. Wilson, A. C., Recht, B., Jordan, M. I.: A Lyapunov analysis of momentum methods in optimization, arXiv preprint [arXiv:1611.02635](https://arxiv.org/abs/1611.02635), (2016)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.