CrossMark

Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity

**Hedy Attouch[1] · Zaki Chbani[2] ·
Juan Peypouquet[3] · Patrick Redont[1]**

Abstract In a Hilbert space setting $\mathcal{H}$, we study the fast convergence properties as $t \to +\infty$ of the trajectories of the second-order differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = g(t),$$

where $\nabla\Phi$ is the gradient of a convex continuously differentiable function $\Phi : \mathcal{H} \to \mathbb{R}$, α is a positive parameter, and $g : [t_0, +\infty[\to \mathcal{H}$ is a *small* perturbation term. In this inertial system, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes asymptotically, but not too rapidly. For $\alpha \geq 3$, and $\int_{t_0}^{+\infty} t\|g(t)\|dt < +\infty$, just assuming that argmin $\Phi \neq$

✉ Juan Peypouquet
  juan.peypouquet@usm.cl

  Hedy Attouch
  hedy.attouch@univ-montp2.fr

  Zaki Chbani
  chbaniz@uca.ma

  Patrick Redont
  patrick.redont@univ-montp2.fr

[1] Institut Montpelliérain Alexander Grothendieck, UMR 5149 CNRS, Université Montpellier 2, Place Eugéne Bataillon, 34095 Montpellier Cedex 5, France

[2] Laboratoire IBN AL-BANNA de Mathématiques et Applications (LIBMA), Faculty of Sciences Semlalia, Mathematics, Cadi Ayyad University, 40000 Marrakech, Morocco

[3] Universidad Técnica Federico Santa María, Av España 1680, Valparaiso, Chile

Ø, we show that any trajectory of the above system satisfies the fast convergence property

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{t^2}.$$

Moreover, for $\alpha > 3$, any trajectory converges weakly to a minimizer of Φ. The strong convergence is established in various practical situations. These results complement the $\mathcal{O}(t^{-2})$ rate of convergence for the values obtained by Su, Boyd and Candès in the unperturbed case $g = 0$. Time discretization of this system, and some of its variants, provides new fast converging algorithms, expanding the field of rapid methods for structured convex minimization introduced by Nesterov, and further developed by Beck and Teboulle with FISTA. This study also complements recent advances due to Chambolle and Dossal.

Keywords Convex optimization · Fast convergent methods · Dynamical systems · Gradient flows · Inertial dynamics · Vanishing viscosity · Nesterov method

Mathematics Subject Classification 34D05 · 49M25 · 65K05 · 65K10 · 90C25 · 90C30

1 Introduction

Let $\mathcal{H}$ be a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, and let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function. In this paper, we first study the solution trajectories of the second-order differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0, \tag{1}$$

with $\alpha > 0$, in terms of their asymptotic behavior as $t \to +\infty$. This will serve us as guideline to the study of corresponding algorithms.

We take for granted the existence and uniqueness of a global solution to the Cauchy problem associated with (1). Although this is not our main concern, we point out that, given $t_0 > 0$, for any $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{H}$, the existence of a unique global solution on $[t_0, +\infty[$ for the Cauchy problem with initial condition $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ can be guaranteed, for instance, if $\nabla\Phi$ is Lipschitz-continuous on bounded sets, and Φ is minorized.

Throughout the paper, unless otherwise indicated, we simply assume that

$$\Phi : \mathcal{H} \to \mathbb{R} \text{ is a continuously differentiable convex function.}$$

As we shall see, most of the convergence properties of the trajectories are valid under this general assumption. This approach paves the way for the extension of our results to non-smooth convex potential functions (replacing the gradient by the subdifferential).

In preparation for a stability study of this system and the associated algorithms, we will also consider the following perturbed version of (1):

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = g(t), \tag{2}$$

where the right-hand side $g : [t_0, +\infty[\to \mathcal{H}$ is an integrable source term, such that $g(t)$ is small for large t.

The importance of the evolution system (1) is threefold:

1. Mechanical interpretation: It describes the motion of a particle, with mass equal to one, subject to a potential energy function Φ, and an isotropic linear damping with a viscosity parameter that vanishes asymptotically. This provides a simple model for a progressive reduction of the friction, possibly due to material fatigue.

2. Fast minimization of function Φ: Equation (1) is a particular case of the inertial gradient-like system

$$\ddot{x}(t) + a(t)\dot{x}(t) + \nabla\Phi(x(t)) = 0, \tag{3}$$

with *asymptotic vanishing damping*, studied by Cabot et al. [17,18]. As shown in [17, Corollary 3.1] (under some additional conditions on Φ), every solution $x(\cdot)$ of (3) satisfies $\lim_{t \to +\infty} \Phi(x(t)) = \min \Phi$, provided $\int_0^\infty a(t)dt = +\infty$. The specific case (1) was studied by Su, Boyd and Candès in [35] in terms of the rate of convergence of the values. More precisely, [35, Theorem 4.1] establishes that $\Phi(x(t)) - \min \Phi = \mathcal{O}(t^{-2})$, whenever $\alpha \geq 3$. Unfortunately, their analysis does not entail the convergence of the trajectory itself.

3. Relationship with fast numerical optimization methods: As pointed out in [35, Section 2], for $\alpha = 3$, (1) can be seen as a continuous version of the fast convergent method of Nesterov (see [26–29]), and its widely used successors, such as the *Fast Iterative Shrinkage-Thresholding Algorithm* (FISTA), studied in [14]. These methods have a convergence rate of $\Phi(x_k) - \min \Phi = \mathcal{O}(k^{-2})$, where k is the number of iterations. As for the continuous-time system (1), convergence of the sequences generated by FISTA and related methods, has not been established so far. This is a central and long-standing question in the study of numerical optimization methods.

The purpose of this research is to establish the convergence of the trajectories satisfying (1), as well as the sequences generated by the corresponding numerical methods with Nesterov-type acceleration. We also complete the study with several stability properties concerning both the continuous-time system and the algorithms.

The main contributions of this work are the following:

In Sect. 2, we first establish the minimizing property in the general case where $\alpha > 0$, and $\inf \Phi$ is not necessarily attained. As a consequence, every weak limit point of the trajectory must be a minimizer of Φ, and so, the existence of a bounded trajectory characterizes the existence of minimizers. Assuming $\operatorname{argmin} \Phi \neq \emptyset$ and $\alpha \geq 3$, we recover the $\mathcal{O}(t^{-2})$ convergence rates, and give several examples and counterexamples concerning the optimality of these results. Next, we show that every solution of (1) converges weakly to a minimizer of Φ provided $\alpha > 3$ and $\operatorname{argmin} \Phi \neq \emptyset$. For the limiting case $\alpha = 3$, which corresponds exactly to Nesterov's method, the convergence of the trajectories is still a puzzling open question. We finish this section

by providing an ergodic convergence result for the acceleration of the system in case $\nabla\Phi$ is Lipschitz-continuous on sublevel sets of Φ.

In Sect. 3, strong convergence is established in various practical situations enjoying further geometric features, such as strong convexity, symmetry, or nonemptyness of the interior of the solution set. In the strongly convex case, we obtained a surprising result: convergence of the values occurs at a $\mathcal{O}(t^{-\frac{2\alpha}{3}})$ rate.

In Sect. 4, we analyze the asymptotic behavior, as $t \to +\infty$, of the solutions of the perturbed differential system (2), and obtain similar convergence results under integrability assumptions on the perturbation term g.

Section 5 contains the analogous results for the associated Nesterov-type algorithms (also for $\alpha > 3$). To avoid repeating similar arguments, we state the results and develop the proofs directly for the perturbed version. As a guideline, we follow the proof of the convergence of the continuous dynamic. We provide discrete versions of the differential inequalities that we used in the Lyapunov convergence analysis. The convergence results are parallel to those for the continuous case, under summability conditions on the errors.

As we were preparing the final version of this manuscript, we discovered the preprint [19] by Chambolle and Dossal, where the weak convergence result of the algorithm in the unperturbed case is obtained by a similar, but different argument (see [19, Theorem 3]). This approach has been further developed by Aujol–Dossal [10] in the perturbed case.

2 Minimizing property, convergence rates and weak convergence of the trajectories

We begin this section by providing some preliminary estimations concerning the global energy of the system (1) and the distance to the minimizers of Φ. These allow us to show the minimizing property of the trajectories under minimal assumptions. Next, we recover the convergence rates for the values originally given in [35], and obtain further decay estimates that ultimately imply the convergence of the solutions of (1). We finish the study by proving an ergodic convergence result for the acceleration. Several examples and counterexamples are given throughout the section.

2.1 Preliminary remarks and estimations

The existence of global solutions to (1) has been examined, for instance, in [17, Proposition 2.2] in the case of a general asymptotic vanishing damping coefficient. In our setting, for any $t_0 > 0, \alpha > 0$, and $(x_0, v_0) \in \mathcal{H} \times \mathcal{H}$, there exists a unique global solution $x : [t_0, +\infty[\to \mathcal{H}$ of (1), satisfying the initial condition $x(t_0) = x_0, \dot{x}(t_0) = v_0$, under the sole assumption that $\nabla\Phi$ is Lipschitz-continuous on bounded sets, and inf $\Phi > -\infty$. Taking $t_0 > 0$ comes from the singularity of the damping coefficient $a(t) = \frac{\alpha}{t}$ at zero. Indeed, since we are only concerned about the asymptotic behavior of the trajectories, we do not really care about the origin of time. If one insists in starting from $t_0 = 0$, then all the results remain valid with $a(t) = \frac{\alpha}{t+1}$.

At different points, we shall use the *global energy* of the system, given by $W : [t_0, +\infty[\to \mathbb{R}$

$$W(t) = \frac{1}{2}\|\dot{x}(t)\|^2 + \Phi(x(t)). \tag{4}$$

Using (1), we immediately obtain

Lemma 2.1 *Let W be defined by* (4). *For each $t > t_0$, we have*

$$\dot{W}(t) = -\frac{\alpha}{t}\|\dot{x}(t)\|^2.$$

Hence, W is nonincreasing,[1] *and $W_\infty = \lim_{t \to +\infty} W(t)$ exists in $\mathbb{R} \cup \{-\infty\}$. If Φ is bounded from below, W_∞ is finite.*

Now, given $z \in \mathcal{H}$, we define $h_z : [t_0, +\infty[\to \mathbb{R}$ by

$$h_z(t) = \frac{1}{2}\|x(t) - z\|^2. \tag{5}$$

By the Chain Rule, we have

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

Using (1), we obtain

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) = \|\dot{x}(t)\|^2 + \langle x(t) - z, \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) \rangle$$
$$= \|\dot{x}(t)\|^2 + \langle x(t) - z, -\nabla\Phi(x(t)) \rangle. \tag{6}$$

The convexity of Φ implies

$$\langle x(t) - z, \nabla\Phi(x(t)) \rangle \geq \Phi(x(t)) - \Phi(z),$$

and we deduce that

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \Phi(x(t)) - \Phi(z) \leq \|\dot{x}(t)\|^2. \tag{7}$$

We have the following relationship between h_z and W:

Lemma 2.2 *Take $z \in \mathcal{H}$, and let W and h_z be defined by* (4) *and* (5), *respectively. There is a constant C such that*

$$\int_{t_0}^t \frac{1}{s}(W(s) - \Phi(z))\,ds \leq C - \frac{1}{t}\dot{h}_z(t) - \frac{3}{2\alpha}W(t).$$

Proof Divide (7) by t, and use the definition of W given in (4), to obtain

---

[1] In fact, W decreases strictly, as long as the trajectory is not stationary.

$$\frac{1}{t}\ddot{h}_z(t) + \frac{\alpha}{t^2}\dot{h}_z(t) + \frac{1}{t}\left(W(t) - \Phi(z)\right) \leq \frac{3}{2t}\|\dot{x}(t)\|^2.$$

Integrate this expression from t_0 to $t > t_0$ (use integration by parts for the first term), to obtain

$$\int_{t_0}^{t} \frac{1}{s}\left(W(s) - \Phi(z)\right) ds \leq \frac{1}{t_0}\dot{h}_z(t_0) - \frac{1}{t}\dot{h}_z(t)$$

$$-(\alpha + 1)\int_{t_0}^{t} \frac{1}{s^2}\dot{h}_z(s)ds + \int_{t_0}^{t} \frac{3}{2s}\|\dot{x}(s)\|^2 ds. \tag{8}$$

On the one hand, Lemma 2.1 gives

$$\int_{t_0}^{t} \frac{3}{2s}\|\dot{x}(s)\|^2 ds = \frac{3}{2\alpha}\left(W(t_0) - W(t)\right).$$

On the other hand, another integration by parts yields

$$\int_{t_0}^{t} \frac{1}{s^2}\dot{h}_z(s)ds = \frac{1}{t^2}h_z(t) - \frac{1}{t_0^2}h_z(t_0) + \int_{t_0}^{t} \frac{2}{s^3}h_z(s)ds \geq -\frac{1}{t_0^2}h_z(t_0).$$

Combining these inequalities with (8), we get

$$\int_{t_0}^{t} \frac{1}{s}\left(W(s) - \Phi(z)\right) ds \leq \frac{1}{t_0}\dot{h}_z(t_0) - \frac{1}{t}\dot{h}_z(t) + (\alpha + 1)\frac{1}{t_0^2}h_z(t_0)$$

$$+ \frac{3}{2\alpha}\left(W(t_0) - W(t)\right) = C - \frac{1}{t}\dot{h}_z(t) - \frac{3}{2\alpha}W(t),$$

where C collects the constant terms. □

2.2 Minimizing property

It turns out that the trajectories of (1) minimize Φ in the completely general setting, where $\alpha > 0$, argmin Φ is possibly empty, and Φ is not necessarily bounded from below (recall that we assume the existence and uniqueness of a global solution to the Cauchy problem associated with (1), which is not guaranteed by these general assumptions). This property was obtained by Alvarez in [2, Theorem 2.1] for the heavy ball with friction (where the damping is constant). Similar results can be found in [17].

  We have the following:

Theorem 2.3 *Let $\alpha > 0$, and suppose $x : [t_0, +\infty[\to \mathcal{H}$ is a solution of (1). Then*

(i) $W_\infty = \lim_{t \to +\infty} W(t) = \lim_{t \to +\infty} \Phi(x(t)) = \inf \Phi \in \mathbb{R} \cup \{-\infty\}$.
(ii) As $t \to +\infty$, every weak limit point of $x(t)$ lies in argmin Φ.
(iii) If argmin $\Phi = \emptyset$, then $\lim_{t \to +\infty} \|x(t)\| = +\infty$.

(iv) *If x is bounded, then* argmin $\Phi \neq \emptyset$.

 (v) *If Φ is bounded from below, then* $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$.

(vi) *If Φ is bounded from below and x is bounded, then* $\lim_{t\to+\infty} \dot{h}_z(t) = 0$ *for each* $z \in \mathcal{H}$. *Moreover,*

$$\int_{t_0}^{\infty} \frac{1}{t}(\Phi(x(t)) - \min \Phi)dt < +\infty.$$

Proof To prove (i), first set $z \in \mathcal{H}$ and $\tau \geq t > t_0$. By Lemma 2.1, W is nonincreasing. Hence, Lemma 2.2 gives

$$(W(\tau) - \Phi(z)) \int_{t_0}^{t} \frac{ds}{s} + \frac{3}{2\alpha}W(\tau) \leq C - \frac{1}{t}\dot{h}_z(t),$$

which we rewrite as

$$(W(\tau) - \Phi(z)) \left(\int_{t_0}^{t} \frac{ds}{s} + \frac{3}{2\alpha} \right) \leq C - \frac{3}{2\alpha}\Phi(z) - \frac{1}{t}\dot{h}_z(t),$$

and then

$$(W(\tau) - \Phi(z)) \left(\ln(t) + \frac{3}{2\alpha} - \ln(t_0) \right) \leq C - \frac{3}{2\alpha}\Phi(z) - \frac{1}{t}\dot{h}_z(t).$$

Integrate from $t = t_0$ to $t = \tau$ to obtain

$$(W(\tau) - \Phi(z)) \left(\tau \ln(\tau) - t_0 \ln(t_0) + t_0 - \tau + \left(\frac{2}{3\alpha} - \ln(t_0) \right)(\tau - t_0) \right)$$

$$\leq \left(C - \frac{3}{2\alpha}\Phi(z) \right)(\tau - t_0) - \int_{t_0}^{\tau} \frac{1}{t}\dot{h}_z(t)dt.$$

But

$$\int_{t_0}^{\tau} \frac{\dot{h}_z(t)}{t}dt = \frac{h_z(\tau)}{\tau} - \frac{h_z(t_0)}{t_0} + \int_{t_0}^{\tau} \frac{h_z(t)}{t^2}dt \geq -\frac{h_z(t_0)}{t_0}.$$

Hence,

$$(W(\tau) - \Phi(z))(\tau \ln(\tau) + A\tau + B) \leq \tilde{C}\tau + D,$$

for suitable constants A, B, $\tilde{C}$ and D. This immediately yields $W_\infty \leq \Phi(z)$, and hence $W_\infty \leq \inf \Phi$. It suffices to observe that

$$\inf \Phi \leq \liminf_{t\to+\infty} \Phi(x(t)) \leq \limsup_{t\to+\infty} \Phi(x(t)) \leq \lim_{t\to+\infty} W(t) = W_\infty$$

to obtain (i).

Next, (ii) follows from (i) by the lower-semicontinuity of Φ for the weak topology. Clearly, (iii) and (iv) are immediate consequences of (ii). We obtain (v) by using (i) and the definition of W given in (4). For (vi), since $\dot{h}_z(t) = \langle x(t) - z, \dot{x}(t) \rangle$ and x is bounded, (v) implies $\lim_{t \to +\infty} \dot{h}_z(t) = 0$. Finally, using the definition of W together with Lemma 2.2 with $z \in \operatorname{argmin} \Phi$, we get

$$\int_{t_0}^{\infty} \frac{1}{t}(\Phi(x(t)) - \min \Phi)dt \leq C - \frac{3}{2\alpha} \min \Phi < +\infty,$$

which completes the proof.   □

Remark 2.4 We shall see in Theorem 2.14 that, for $\alpha \geq 3$, the existence of minimizers implies that every solution of (1) is bounded. This gives a converse to part iv) of Theorem 2.3.

If Φ is not bounded from below, it may be the case that $\|\dot{x}(t)\|$ does not tend to zero, as shown in the following example:

Example 2.5 Let $\mathcal{H} = \mathbb{R}$ and $\alpha > 0$. The function $x(t) = t^2$ satisfies (1) with $\Phi(x) = -2(\alpha + 1)x$. Then $\lim_{t \to +\infty} \Phi(x(t)) = -\infty = \inf \Phi$, and $\lim_{t \to +\infty} \|\dot{x}(t)\| = +\infty$.

2.3 Two "anchored" energy functions

We begin by introducing two important auxiliary functions, and showing their basic properties. From now on, we assume $\operatorname{argmin} \Phi \neq \emptyset$. Fix $\lambda \geq 0, \xi \geq 0, p \geq 0$ and $x^* \in \operatorname{argmin} \Phi$. Let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of (1). For $t \geq t_0$ define

$$\mathcal{E}_{\lambda,\xi}(t) = t^2(\Phi(x(t)) - \min \Phi) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t) - x^*\|^2,$$

$$\mathcal{E}_{\lambda}^p(t) = t^p \mathcal{E}_{\lambda,0}(t) = t^p \left(t^2(\Phi(x(t)) - \min \Phi) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 \right),$$

and notice that $\mathcal{E}_{\lambda,\xi}$ and $\mathcal{E}_{\lambda}^p$ are sums of nonnegative terms. These generalize the energy functions $\mathcal{E}$ and $\tilde{\mathcal{E}}$ introduced in [35]. More precisely, $\mathcal{E} = \mathcal{E}_{\alpha-1,0}$ and $\tilde{\mathcal{E}} = \mathcal{E}_{(2\alpha-3)/3}^1$.

We need some preparatory calculations prior to differentiating $\mathcal{E}_{\lambda,\xi}$ and $\mathcal{E}_{\lambda}^p$. For simplicity of notation, we do not make the dependence of x or $\dot{x}$ on t explicit. Notice that we use (1) in the second line to dispose of $\ddot{x}$.

$$\frac{d}{dt}t^2(\Phi(x) - \min \Phi) = 2t(\Phi(x) - \min \phi) + t^2\langle \dot{x}, \nabla\Phi(x) \rangle$$

$$\frac{d}{dt}\frac{1}{2}\|\lambda(x - x^*) + t\dot{x}\|^2 = -\lambda t\langle x - x^*, \nabla\Phi(x) \rangle - \lambda(\alpha - \lambda - 1)\langle x - x^*, \dot{x} \rangle$$

$$- (\alpha - \lambda - 1)t\|\dot{x}\|^2 - t^2\langle \dot{x}, \nabla\Phi(x) \rangle$$

$$\frac{d}{dt}\frac{1}{2}\|x - x^*\|^2 = \langle x - x^*, \dot{x} \rangle.$$

Whence, we deduce

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) = 2t(\Phi(x) - \min \Phi) - \lambda t \langle x - x^*, \nabla\Phi(x)\rangle$$

$$+ (\xi - \lambda(\alpha - \lambda - 1))\langle x - x^*, \dot{x}\rangle - (\alpha - \lambda - 1)t\|\dot{x}\|^2 \qquad (9)$$

$$\frac{d}{dt}\mathcal{E}_{\lambda}^{p}(t) = (p+2)t^{p+1}(\Phi(x) - \min \Phi) - \lambda t^{p+1} \langle x - x^*, \nabla\Phi(x)\rangle$$

$$- \lambda(\alpha - \lambda - 1 - p)t^{p}\langle x - x^*, \dot{x}\rangle$$

$$+ \frac{\lambda^2 p}{2}t^{p-1}\|x - x^*\|^2 - \left(\alpha - \lambda - 1 - \frac{p}{2}\right)t^{p+1}\|\dot{x}\|^2. \qquad (10)$$

Remark 2.6 If $y \in \mathcal{H}$ and $x^* \in \operatorname{argmin} \Phi$, the convexity of Φ gives $\min \Phi = \Phi(x^*) \geq \Phi(y) + \langle \nabla\Phi(y), x^* - y\rangle$. Using this in (9) with $y = x(t)$, we obtain

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \leq (2 - \lambda)\, t\, (\Phi(x) - \min \Phi) + (\xi - \lambda(\alpha - \lambda - 1))\langle x - x^*, \dot{x}\rangle$$

$$- (\alpha - \lambda - 1)\, t\, \|\dot{x}\|^2.$$

If $\alpha \geq 3$ and $2 \leq \lambda \leq \alpha - 1$ and if one chooses $\xi^* = \lambda(\alpha - \lambda - 1)$ (which is nonnegative), then

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi^*}(t) \leq (2 - \lambda)\, t\, (\Phi(x) - \min \Phi) - (\alpha - \lambda - 1)\, t\, \|\dot{x}\|^2.$$

Hence $\mathcal{E}_{\lambda,\xi^*}$ is nonincreasing. The extreme cases $\lambda = 2$ and $\lambda = \alpha - 1$ are of special importance, as we shall see shortly.

2.4 Rate of convergence for the values

We now recover convergence rate results for the value of Φ along a trajectory, already established in [35, Theorem 4.1]:

Theorem 2.7 *Let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution of* (1)*, and assume* $\operatorname{argmin} \Phi \neq \emptyset$. *If* $\alpha \geq 3$*, then*

$$\Phi(x(t)) - \min \Phi \leq \frac{\mathcal{E}_{\alpha-1,0}(t_0)}{t^2}.$$

If $\alpha > 3$*, then*

$$\int_{t_0}^{+\infty} t\big(\Phi(x(t)) - \min \Phi\big)\, dt \leq \frac{\mathcal{E}_{\alpha-1,0}(t_0)}{\alpha - 3} < +\infty.$$

Proof Suppose $\alpha \geq 3$. Choose $\lambda = \alpha - 1$ and $\xi = 0$, so that $\xi - \lambda(\alpha - \lambda - 1) = \alpha - \lambda - 1 = 0$ and $\lambda - 2 = \alpha - 3$. Remark 2.6 gives

$$\frac{d}{dt}\mathcal{E}_{\alpha-1,0}(t) \leq -(\alpha - 3)\, t\, (\Phi(x) - \min \Phi), \qquad (11)$$

and $\mathcal{E}_{\alpha-1,0}$ is nonincreasing. Since $t^2(\Phi(x)-\min \Phi) \leq \mathcal{E}_{\alpha-1,0}(t)$ (recall the definition of $\mathcal{E}_{\lambda,\xi}$), we obtain

$$\Phi(x(t)) - \min \Phi \leq \frac{\mathcal{E}_{\alpha-1,0}(t_0)}{t^2}.$$

If $\alpha > 3$, integrating (11) from t_0 to t we obtain

$$\int_{t_0}^{t} s(\Phi(x(s)) - \min \Phi)ds \leq \frac{1}{\alpha-3}(\mathcal{E}_{\alpha-1,0}(t_0) - \mathcal{E}_{\alpha-1,0}(t)) \leq \frac{1}{\alpha-3}\mathcal{E}_{\alpha-1,0}(t_0),$$

which allows us to conclude.   □

Remark 2.8 It would be interesting to know whether $\alpha = 3$ is critical for the convergence rate given above.

Remark 2.9 For the (first-order) steepest descent dynamical system, the typical rate of convergence is $\mathcal{O}(1/t)$ (see, for instance, [33, Section 3.1]). For the second-order system (1), we have obtained a rate of $\mathcal{O}(1/t^2)$. It would be interesting to know whether higher-order systems give the corresponding rates of convergence. Another challenging question is the convergence rate of the trajectories defined by differential equations involving fractional time derivatives, as well as integro-differential equations.

Remark 2.10 The constant in the order of convergence given by Theorem 2.7 is

$$K(x_0, v_0) = \mathcal{E}_{\alpha-1,0}(t_0) = t_0^2(\Phi(x_0) - \min \Phi) + \frac{1}{2}\|(\alpha-1)(x_0 - x^*) + t_0 v_0\|^2,$$

where $x_0 = x(t_0)$ and $v_0 = \dot{x}(t_0)$. This quantity is minimized when $x_0^* \in \operatorname{argmin} \Phi$ and $v_0^* = \frac{(\alpha-1)}{t_0}(x^* - x_0^*)$, with $\min K = 0$. If $x_0^* \neq x^*$, the trajectory will not be stationary, but the value $\Phi(x(t))$ will be constantly equal to $\min \Phi$. Of course, selecting $x_0^* \in \operatorname{argmin} \Phi$ is not realistic, and the point x^* is unknown. Keeping $\hat{x}_0$ fixed, the function $v_0 \mapsto K(\hat{x}_0, v_0)$ is minimized at $\hat{v}_0 = \frac{(\alpha-1)}{t_0}(x^* - \hat{x}_0)$. This suggests taking the initial velocity as a multiple of an approximation of $x^* - \hat{x}_0$, such as the gradient direction $\hat{v}_0 = \nabla\Phi(\hat{x}_0)$, Newton or Levenberg-Marquardt direction $\hat{v}_0 = [\varepsilon I + \nabla^2\Phi(\hat{x}_0)^{-1}]\nabla\Phi(\hat{x}_0)$ ($\varepsilon \geq 0$), or the proximal point direction $\hat{v}_0 = [(I + \gamma\nabla\Phi)^{-1}(\hat{x}_0) - \hat{x}_0]$ ($\gamma >> 0$).

2.5 Some examples and counterexamples

A convergence rate of $\mathcal{O}(1/t^2)$ may be attained, even if $\operatorname{argmin} \Phi = \emptyset$ and $\alpha < 3$. This is illustrated in the following example:

Example 2.11 Let $\mathcal{H} = \mathbb{R}$ and take $\Phi(x) = \frac{\alpha-1}{2}e^{-2x}$ with $\alpha \geq 1$. Let us verify that $x(t) = \ln t$ is a solution of (1). On the one hand,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = \frac{\alpha-1}{t^2}.$$

On the other hand, $\nabla\Phi(x) = -(\alpha - 1)e^{-2x}$ which gives

$$\nabla\Phi(x(t)) = -(\alpha - 1)e^{-2\ln t} = -\frac{\alpha - 1}{t^2}.$$

Thus, $x(t) = \ln t$ is a solution of (1). Let us examine the minimizing property. We have inf $\Phi = 0$, and

$$\Phi(x(t)) = \frac{\alpha - 1}{2}e^{-2\ln t} = \frac{\alpha - 1}{2t^2}.$$

Therefore, one may wonder whether the rapid convergence of the values is true in general. The following example shows that this is not the case:

Example 2.12 Let $\mathcal{H} = \mathbb{R}$ and take $\Phi(x) = \frac{c}{x^\theta}$, with $\theta > 0$, $\alpha \geq \frac{\theta}{(2+\theta)}$ and $c = \frac{2(2\alpha+\theta(\alpha-1))}{\theta(2+\theta)^2}$. Let us verify that $x(t) = t^{\frac{2}{2+\theta}}$ is a solution of (1). On the one hand,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = \frac{2}{(2+\theta)^2}(2\alpha + \theta(\alpha - 1))t^{-\frac{2(1+\theta)}{2+\theta}}.$$

On the other hand, $\nabla\Phi(x) = -c\theta x^{-\theta-1}$ which gives

$$\nabla\Phi(x(t)) = -c\theta t^{-\frac{2(1+\theta)}{2+\theta}} = -\frac{2}{(2+\theta)^2}(2\alpha + \theta(\alpha - 1))t^{-\frac{2(1+\theta)}{2+\theta}}.$$

Thus, $x(t) = t^{\frac{2}{2+\theta}}$ is solution of (1). Let us examine the minimizing property. We have inf $\Phi = 0$, and

$$\Phi(x(t)) = c\frac{1}{t^{\frac{2\theta}{2+\theta}}} \quad \text{with} \quad \frac{2\theta}{2+\theta} < 2.$$

We conclude that the order of convergence may be strictly slower than $\mathcal{O}(1/t^2)$ when argmin $\Phi = \emptyset$. In the Example 2.12, this occurs no matter how large α is. The speed of convergence of $\Phi(x(t))$ to inf Φ depends on the behavior of $\Phi(x)$ as $\|x\| \to +\infty$. The above examples suggest that, when $\Phi(x)$ decreases rapidly and attains its infimal value as $\|x\| \to \infty$, we can expect fast convergence of $\Phi(x(t))$.

Even when argmin $\Phi \neq \emptyset$, $\mathcal{O}(1/t^2)$ is the worst possible case for the rate of convergence, attained as a limit in the following example:

Example 2.13 Take $\mathcal{H} = \mathbb{R}$ and $\Phi(x) = c|x|^\gamma$, where c and γ are positive parameters. Let us look for nonnegative solutions of (1) of the form $x(t) = \frac{1}{t^\theta}$, with $\theta > 0$. This means that the trajectory is not oscillating, it is a completely damped trajectory. We begin by determining the values of c, γ and θ that provide such solutions. On the one hand,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = \theta(\theta + 1 - \alpha)\frac{1}{t^{\theta+2}}.$$

On the other hand, $\nabla \Phi(x) = c\gamma |x|^{\gamma-2}x$, which gives

$$\nabla \Phi(x(t)) = c\gamma \frac{1}{t^{\theta(\gamma-1)}}.$$

Thus, $x(t) = \frac{1}{t^\theta}$ is solution of (1) if, and only if,

(i) $\theta + 2 = \theta(\gamma - 1)$, which is equivalent to $\gamma > 2$ and $\theta = \frac{2}{\gamma-2}$; and

(ii) $c\gamma = \theta(\alpha - \theta - 1)$, which is equivalent to $\alpha > \frac{\gamma}{\gamma-2}$ and $c = \frac{2}{\gamma(\gamma-2)}(\alpha - \frac{\gamma}{\gamma-2})$.

We have $\min \Phi = 0$ and

$$\Phi(x(t)) = \frac{2}{\gamma(\gamma-2)}(\alpha - \frac{\gamma}{\gamma-2})\frac{1}{t^{\frac{2\gamma}{\gamma-2}}}.$$

The speed of convergence of $\Phi(x(t))$ to 0 depends on the parameter γ. As γ tends to infinity, the exponent $\frac{2\gamma}{\gamma-2}$ tends to 2. This limiting situation is obtained by taking a function Φ that becomes very flat around the set of its minimizers. Therefore, without other geometric assumptions on Φ, we cannot expect a convergence rate better than $\mathcal{O}(1/t^2)$. By contrast, in Sect. 3, we will show better rates of convergence under some geometrical assumptions, like strong convexity of Φ.

2.6 Weak convergence of the trajectories

In this subsection, we show the convergence of the solutions of (1), provided $\alpha > 3$. We begin by establishing some preliminary estimations that cannot be derived from the analysis carried out in [35]. The first statement improves part (v) of Theorem 2.3, while the second one is the key to proving the convergence of the trajectories of (1):

Theorem 2.14 *Let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of (1) with $\operatorname{argmin} \Phi \neq \emptyset$.*

(i) If $\alpha \geq 3$ and x is bounded, then $\|\dot{x}(t)\| = \mathcal{O}(1/t)$. More precisely,

$$\|\dot{x}(t)\| \leq \frac{1}{t}\left(\sqrt{2\mathcal{E}_{\alpha-1,0}(t_0)} + (\alpha - 1)\sup_{t \geq t_0} \|x(t) - x^*\|\right). \tag{12}$$

(ii) If $\alpha > 3$, then x is bounded and

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 \, dt \leq \frac{\mathcal{E}_{2,2(\alpha-3)}(t_0)}{\alpha - 3} < +\infty. \tag{13}$$

Proof To prove (i), assume $\alpha \geq 3$ and x is bounded. From the definition of $\mathcal{E}_{\lambda,\xi}$, we have $\frac{1}{2}\|\lambda(x - x^*) + t\dot{x}\|^2 \leq \mathcal{E}_{\lambda,\xi}(t)$, and so $\|t\dot{x}\| \leq \sqrt{2\mathcal{E}_{\lambda,\xi}(t)} + \lambda\|x - x^*\|$. By Remark 2.6, $\mathcal{E}_{\alpha-1,0}$ is nonincreasing, and we immediately obtain (12).

In order to show (ii), suppose now that $\alpha > 3$. Choose $\lambda = 2$ and $\xi^* = 2(\alpha - 3)$. By Remark 2.6, we have

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi^*}(t) \leq -(\alpha - 3)t\|\dot{x}\|^2, \tag{14}$$

and $\mathcal{E}_{\lambda,\xi^*}$ is nonincreasing. From the definition of $\mathcal{E}_{\lambda,\xi}$, we deduce that $\|x(t) - x^*\|^2 \leq \frac{2}{\xi}\mathcal{E}_{\lambda,\xi}(t)$, which gives

$$\|x(t) - x^*\|^2 \leq \frac{\mathcal{E}_{2,2(\alpha-3)}(t)}{\alpha - 3} \leq \frac{\mathcal{E}_{2,2(\alpha-3)}(t_0)}{\alpha - 3}, \tag{15}$$

and establishes the boundedness of x. Integrating (14) from t_0 to t, and recalling that $\mathcal{E}_{\lambda,\xi^*}$ is nonnegative, we obtain

$$\int_{t_0}^t s\|\dot{x}(s)\|^2 ds \leq \frac{\mathcal{E}_{2,2(\alpha-3)}(t_0)}{\alpha - 3},$$

as required. □

Remark 2.15 In view of (12) and (15), when $\alpha > 3$, we obtain the following explicit bound for $\|\dot{x}\|$, namely

$$\|\dot{x}(t)\| \leq \frac{1}{t}\left(\sqrt{2\mathcal{E}_{\alpha-1,0}(t_0)} + (\alpha - 1)\sqrt{\frac{\mathcal{E}_{2,2(\alpha-3)}(t_0)}{\alpha - 3}}\right).$$

Since $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$ by Theorem 2.3, we also have $\lim_{t\to+\infty} t\|\dot{x}(t)\|^2 = 0$.

We are now in a position to prove the weak convergence of the trajectories of (1), which is the main result of this section. The proof relies on a Lyapunov analysis, which was first used by Alvarez [2] in the context of the heavy ball with friction.

Theorem 2.16 *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a continuously differentiable convex function such that $\operatorname{argmin} \Phi \neq \emptyset$, and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of (1) with $\alpha > 3$. Then $x(t)$ converges weakly, as $t \to +\infty$, to a point in $\operatorname{argmin} \Phi$.*

Proof We shall use Opial's Lemma 5.7. To this end, let $x^* \in \operatorname{argmin} \Phi$ and recall from (7) that

$$\ddot{h}_{x^*}(t) + \frac{\alpha}{t}\dot{h}_{x^*}(t) + \Phi(x(t)) - \min \Phi \leq \|\dot{x}(t)\|^2,$$

where h_z is given by (5). This yields

$$t\ddot{h}_{x^*}(t) + \alpha \dot{h}_{x^*}(t) \leq t\|\dot{x}(t)\|^2.$$

In view of Theorem 2.14, part (ii), the right-hand side is integrable on $[t_0, +\infty[$. Lemma 5.9 then implies $\lim_{t\to+\infty} h_{x^*}(t)$ exists. This gives the first hypothesis in Opial's Lemma. The second one was established in part (ii) of Theorem 2.3. □

Remark 2.17 A puzzling question concerns the convergence of the trajectories for $\alpha = 3$, a question which is directly related to the convergence of the sequences generated by Nesterov's method [26] (see [35, Section 3]).

2.7 Further stabilization results

Let us complement the study of Eq. (1) by examining the asymptotic behavior of the acceleration $\ddot{x}$. To this end, we shall use an additional regularity assumption on the gradient of Φ.

Proposition 2.18 *Let $\alpha > 3$ and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of (1) with argmin $\Phi \neq \emptyset$. Assume $\nabla\Phi$ Lipschitz-continuous on bounded sets. Then $\ddot{x}$ is bounded, globally Lipschitz continuous on $[t_0, +\infty[$, and satisfies*

$$\lim_{t\to+\infty} \frac{1}{t^\alpha} \int_{t_0}^t s^\alpha \|\ddot{x}(s)\|^2 ds = 0.$$

Proof First recall that x and $\dot{x}$ are bounded, by virtue of Theorems 2.14 and 2.3, respectively. By (1), we have

$$\ddot{x}(t) = -\frac{\alpha}{t}\dot{x}(t) - \nabla\Phi(x(t)). \tag{16}$$

Since $\nabla\Phi$ is Lipschitz-continuous on bounded sets, it follows from (16), and the boundedness of x and $\dot{x}$, that $\ddot{x}$ is bounded on $[t_0, +\infty[$. As a consequence, $\dot{x}$ is Lipschitz-continuous on $[t_0, +\infty[$. Returning to (16), we deduce that $\ddot{x}$ is Lipschitz-continuous on $[t_0, +\infty[$.

Pick $x^* \in$ argmin Φ, set $h = h_{x^*}$ (to simplify the notation) and use (6) to obtain

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \langle x(t) - x^*, \nabla\Phi(x(t))\rangle = \|\dot{x}(t)\|^2. \tag{17}$$

Let L be a Lipschitz constant for $\nabla\Phi$ on some ball containing the minimizer x^* and the trajectory x. By virtue of the Baillon–Haddad Theorem (see, for instance, [12], [32, Theorem 3.13] or [27, Theorem 2.1.5]), $\nabla\Phi$ is $\frac{1}{L}$-cocoercive on that ball, which means that

$$\langle x(t) - x^*, \nabla\Phi(x(t)) - \nabla\Phi(x^*)\rangle \geq \frac{1}{L}\|\nabla\Phi(x(t)) - \nabla\Phi(x^*)\|^2.$$

Substituting this inequality in (17), and using the fact that $\nabla\Phi(x^*) = 0$, we obtain

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \frac{1}{L}\|\nabla\Phi(x(t))\|^2 \leq \|\dot{x}(t)\|^2.$$

In view of (16), this gives

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \frac{1}{L}\left\|\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t)\right\|^2 \leq \|\dot{x}(t)\|^2.$$

Developing the square on the left-hand side, and neglecting the nonnegative term $(\alpha\|\dot{x}(t)\|/t)^2/L$, we obtain

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \frac{1}{L}\|\ddot{x}(t)\|^2 + \frac{\alpha}{Lt}\frac{d}{dt}\|\dot{x}(t)\|^2 \le \|\dot{x}(t)\|^2.$$

We multiply this inequality by t^α to obtain

$$\frac{d}{dt}\left(t^\alpha \dot{h}(t)\right) + \frac{1}{L}t^\alpha\|\ddot{x}(t)\|^2 + \frac{\alpha}{L}t^{\alpha-1}\frac{d}{dt}\|\dot{x}(t)\|^2 \le t^\alpha\|\dot{x}(t)\|^2.$$

Integration from t_0 to t yields

$$t^\alpha \dot{h}(t) - t_0^\alpha \dot{h}(t_0) + \frac{1}{L}\int_{t_0}^t s^\alpha\|\ddot{x}(s)\|^2 ds$$

$$+ \frac{\alpha}{L}\left(t^{\alpha-1}\|\dot{x}(t)\|^2 - t_0^{\alpha-1}\|\dot{x}(t_0)\|^2 - (\alpha-1)\right.$$

$$\left.\int_{t_0}^t \|\dot{x}(s)\|^2 s^{\alpha-2} ds\right) \le \int_{t_0}^t s^\alpha\|\dot{x}(s)\|^2 ds.$$

Neglecting the nonnegative term $\alpha t^{\alpha-1}\|\dot{x}(t)\|^2/L$, we obtain

$$t^\alpha \dot{h}(t) + \frac{1}{L}\int_{t_0}^t s^\alpha\|\ddot{x}(s)\|^2 ds \le C + (\alpha-1)\int_{t_0}^t \|\dot{x}(s)\|^2 s^{\alpha-2} ds$$

$$+ \int_{t_0}^t s^\alpha\|\dot{x}(s)\|^2 ds, \tag{18}$$

where $C = t_0^\alpha \dot{h}(t_0) + \alpha t_0^{\alpha-1}\|\dot{x}(t_0)\|^2/L$.

If $t_0 < 1$, we have

$$\frac{1}{t^\alpha}\int_{t_0}^t s^\alpha\|\ddot{x}(s)\|^2 ds = \frac{1}{t^\alpha}\int_{t_0}^1 s^\alpha\|\ddot{x}(s)\|^2 ds + \frac{1}{t^\alpha}\int_1^t s^\alpha\|\ddot{x}(s)\|^2 ds$$

for all $t \ge 1$. Since the first term on the right-hand side tends to 0 as $t \to +\infty$, we may assume, without loss of generality, that $t_0 \ge 1$.

Observe now that $s^{\alpha-2} \le s^\alpha$, whenever $s \ge 1$. Whence, inequality (18) simplifies to

$$t^\alpha \dot{h}(t) + \frac{1}{L}\int_{t_0}^t s^\alpha\|\ddot{x}(s)\|^2 ds \le C + \alpha\int_{t_0}^t s^\alpha\|\dot{x}(s)\|^2 ds.$$

Dividing by t^α and integrating again, we obtain

$$h(t) - h(t_0) + \frac{1}{L}\int_{t_0}^t \tau^{-\alpha}\left(\int_{t_0}^\tau s^\alpha\|\ddot{x}(s)\|^2 ds\right) d\tau \le \frac{C}{\alpha-1}\left(t_0^{-\alpha+1} - t^{-\alpha+1}\right)$$

$$+ \alpha\int_{t_0}^t \tau^{-\alpha}\left(\int_{t_0}^\tau s^\alpha\|\dot{x}(s)\|^2 ds\right) d\tau.$$

Setting $C' = h(t_0) + Ct_0^{-\alpha+1}/(\alpha-1)$, and neglecting the nonnegative term $h(t)$ of the left-hand side and the nonpositive term $-Ct^{-\alpha+1}/(\alpha-1)$ of the right-hand side, we get

$$\frac{1}{L}\int_{t_0}^{t}\tau^{-\alpha}\left(\int_{t_0}^{\tau}s^{\alpha}\|\ddot{x}(s)\|^2ds\right)d\tau \leq C' + \alpha\int_{t_0}^{t}\tau^{-\alpha}\left(\int_{t_0}^{\tau}s^{\alpha}\|\dot{x}(s)\|^2ds\right)d\tau.$$

Set $g(\tau) = \tau^{-\alpha}\left(\int_{t_0}^{\tau}s^{\alpha}\|\ddot{x}(s)\|^2ds\right)$ and use Fubini's Theorem on the second integral to get

$$\frac{1}{L}\int_{t_0}^{t}g(\tau)d\tau \leq C' + \frac{\alpha}{\alpha-1}\int_{t_0}^{t}s^{\alpha}\|\dot{x}(s)\|^2\left(s^{-\alpha+1}-t^{-\alpha+1}\right)ds$$

$$\leq C' + \frac{\alpha}{\alpha-1}\int_{t_0}^{t}s\|\dot{x}(s)\|^2ds.$$

By part (ii) of Theorem 2.14, the integral on the right-hand side is finite. We have

$$\int_{t_0}^{+\infty}g(\tau)d\tau < +\infty. \tag{19}$$

The derivative of g is

$$\dot{g}(\tau) = -\alpha\tau^{-\alpha-1}\int_{t_0}^{\tau}s^{\alpha}\|\ddot{x}(s)\|^2ds + \|\ddot{x}(\tau)\|^2.$$

Let C'' be an upper bound for $\|\ddot{x}\|^2$. We have

$$|\dot{g}(\tau)| \leq C''\left(1 + \alpha\tau^{-\alpha-1}\int_{t_0}^{\tau}s^{\alpha}ds\right) = C''\left(1 + \frac{\alpha}{\alpha+1}\tau^{-\alpha-1}\left(\tau^{\alpha+1}-t^{\alpha+1}\right)\right)$$

$$\leq C''\left(1 + \frac{\alpha}{\alpha+1}\right). \tag{20}$$

From (19) and (20) we deduce that $\lim_{\tau\to+\infty}g(\tau) = 0$ by virtue of Lemma 5.6. $\square$

Remark 2.19 Since $\int_{t_0}^{t}s^{\alpha}ds = \frac{1}{\alpha+1}\left(t^{\alpha+1}-t_0^{\alpha+1}\right)$, Proposition 2.18 expresses a fast ergodic convergence of $\|\ddot{x}(s)\|^2$ to 0 with respect to the weight s^{α} as $t\to+\infty$, namely

$$\frac{\int_{t_0}^{t}s^{\alpha}\|\ddot{x}(s)\|^2ds}{\int_{t_0}^{t}s^{\alpha}ds} = o\left(\frac{1}{t}\right).$$

3 Strong convergence results

A counterexample due to Baillon [11] shows that the trajectories of the steepest descent dynamical system may converge weakly but not strongly. Nevertheless, under some additional geometrical or topological assumptions on Φ, the steepest descent trajectories do converge strongly. This has been proved in the case where the function Φ is either even or strongly convex (see [16]), or when $\mathrm{int}(\mathrm{argmin}\,\Phi) \neq \emptyset$ (see [15, theorem 3.13]). Some of these results have been extended to inertial dynamics, see [2] for the heavy ball with friction, and [4] for an inertial version of Newton's method. This suggests that convexity alone may not be sufficient for the trajectories of (1) to converge strongly, but one can reasonably expect it to be the case under some additional conditions. The purpose of this section is to establish this fact. The different types of hypotheses will be studied in independent subsections since different techniques are required.

3.1 Set of minimizers with nonempty interior

Let us begin by studying the case where $\mathrm{int}(\mathrm{argmin}\,\Phi) \neq \emptyset$.

Theorem 3.1 *Let* $\Phi : \mathcal{H} \to \mathbb{R}$ *be a continuously differentiable convex function. Let* $\mathrm{int}(\mathrm{argmin}\,\Phi) \neq \emptyset$, *and let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution of* (1) *with* $\alpha > 3$. *Then* $x(t)$ *converges strongly, as* $t \to +\infty$, *to a point in* $\mathrm{argmin}\,\Phi$. *Moreover,*

$$\int_{t_0}^{\infty} t\,\|\nabla\Phi(x(t))\|dt < +\infty.$$

Proof Since $\mathrm{int}(\mathrm{argmin}\,\Phi) \neq \emptyset$, there exist $x^* \in \mathrm{argmin}\,\Phi$ and some $\rho > 0$ such that $\nabla\Phi(z) = 0$ for all $z \in \mathcal{H}$ such that $\|z - x^*\| < \rho$. By the monotonicity of $\nabla\Phi$, for all $y \in \mathcal{H}$, we have

$$\langle\nabla\Phi(y), y - z\rangle \geq 0.$$

Hence,

$$\langle\nabla\Phi(y), y - x^*\rangle \geq \langle\nabla\Phi(y), z - x^*\rangle.$$

Taking the supremum with respect to $z \in \mathcal{H}$ such that $\|z - x^*\| < \rho$, we infer that

$$\langle\nabla\Phi(y), y - x^*\rangle \geq \rho\|\nabla\Phi(y)\|$$

for all $y \in \mathcal{H}$. In particular,

$$\langle\nabla\Phi(x(t)), x(t) - x^*\rangle \geq \rho\|\nabla\Phi(x(t))\|.$$

By using this inequality in (9) with $\lambda = \alpha - 1$ and $\xi = 0$, we obtain

$$\frac{d}{dt}\mathcal{E}_{\alpha-1,0}(t) + (\alpha - 1)\rho t \|\nabla\Phi(x(t))\| \le 2t\big(\Phi(x(t)) - \min\Phi\big),$$

whence we derive, by integrating from t_0 to t

$$\mathcal{E}_{\alpha-1,0}(t) - \mathcal{E}_{\alpha-1,0}(t_0) + (\alpha - 1)\rho \int_{t_0}^{t} s\|\nabla\Phi(x(s))\|ds$$

$$\le 2\int_{t_0}^{t} s\big(\Phi(x(s)) - \min\Phi\big)\,ds.$$

Since $\mathcal{E}_{\alpha-1,0}(t)$ is nonnegative, part (ii) of Theorem 2.7 gives

$$\int_{t_0}^{\infty} t\|\nabla\Phi(x(t))\|dt < +\infty. \tag{21}$$

Finally, rewrite (1) as
$$t\ddot{x}(t) + \alpha\dot{x}(t) = -t\nabla\Phi(x(t)).$$

Since the right-hand side is integrable, we conclude by applying Lemma 5.12 and Theorem 2.16. □

3.2 Even functions

Let us recall that $\Phi : \mathcal{H} \to \mathbb{R}$ is *even* if $\Phi(-x) = \Phi(x)$ for every $x \in \mathcal{H}$. In this case the set argmin Φ is nonempty, and contains the origin.

Theorem 3.2 *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a continuously differentiable convex even function, and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of (1) with $\alpha > 3$. Then $x(t)$ converges strongly, as $t \to +\infty$, to a point in* argmin Φ.

Proof For $t_0 \le \tau \le s$, set

$$q(\tau) = \|x(\tau)\|^2 - \|x(s)\|^2 - \frac{1}{2}\|x(\tau) - x(s)\|^2.$$

We have

$$\dot{q}(\tau) = \langle \dot{x}(\tau), x(\tau) + x(s)\rangle \quad \text{and} \quad \ddot{q}(\tau) = \|\dot{x}(\tau)\|^2 + \langle\ddot{x}(\tau), x(\tau) + x(s)\rangle.$$

Combining these two equalities and using (1), we obtain

$$\ddot{q}(\tau) + \frac{\alpha}{\tau}\dot{q}(\tau) = \|\dot{x}(\tau)\|^2 + \langle\ddot{x}(\tau) + \frac{\alpha}{\tau}\dot{x}(\tau), x(\tau) + x(s)\rangle$$

$$= \|\dot{x}(\tau)\|^2 - \langle\nabla\Phi(x(\tau)), x(\tau) + x(s)\rangle. \tag{22}$$

Recall that the energy $W(\tau) = \frac{1}{2}\|\dot{x}(\tau)\|^2 + \Phi(x(\tau))$ is nonincreasing. Therefore,

$$\begin{aligned}
\frac{1}{2}\|\dot{x}(\tau)\|^2 + \Phi(x(\tau)) &\geq \frac{1}{2}\|\dot{x}(s)\|^2 + \Phi(x(s)) \\
&= \frac{1}{2}\|\dot{x}(s)\|^2 + \Phi(-x(s)) \\
&\geq \frac{1}{2}\|\dot{x}(s)\|^2 + \Phi(x(\tau)) - \langle\nabla\Phi(x(\tau)), x(\tau) + x(s)\rangle,
\end{aligned}$$

by convexity. After simplification, we obtain

$$\frac{1}{2}\|\dot{x}(\tau)\|^2 \geq -\langle\nabla\Phi(x(\tau)), x(\tau) + x(s)\rangle. \tag{23}$$

Combining (22) and (23), we obtain

$$\tau\ddot{q}(\tau) + \alpha\dot{q}(\tau) \leq \frac{3}{2}\tau\|\dot{x}(\tau)\|^2.$$

As in the proof of Lemma 5.9, we have

$$\dot{q}(\tau) \leq k(\tau) := \frac{C}{\tau^\alpha} + \frac{3}{2\tau^\alpha}\int_{t_0}^\tau u^\alpha\|\dot{x}(u)\|^2 du,$$

where $C = 2\|\dot{x}(t_0)\|\,\|x\|_\infty$. The function k does not depend on s. Moreover, using Fubini's Theorem, we deduce that

$$\int_{t_0}^{+\infty} k(\tau)\,d\tau \leq \frac{C}{t_0^{\alpha-1}(\alpha-1)} + \frac{3}{2(\alpha-1)}\int_{t_0}^{+\infty} u\|\dot{x}(u)\|^2\,du < +\infty,$$

by part ii) of Theorem 2.14. Integrating $\dot{q}(\tau) \leq k(\tau)$ from t to s, we obtain

$$\frac{1}{2}\|x(t) - x(s)\|^2 \leq \|x(t)\|^2 - \|x(s)\|^2 + \int_t^s k(\tau)d\tau.$$

Since Φ is even, we have $0 \in \operatorname{argmin}\Phi$. Hence $\lim_{t\to+\infty}\|x(t)\|^2$ exists (see the proof of Theorem 2.16). As a consequence, $x(t)$ has the Cauchy property as $t \to +\infty$, and hence converges. □

3.3 Uniformly convex functions

Following [13], a function $\Phi : \mathcal{H} \to \mathbb{R}$ is *uniformly convex on bounded sets* if, for each $r > 0$, there is an increasing function $\omega_r : [0, +\infty[\to [0, +\infty[$ vanishing only at 0, and such that

$$\Phi(y) \geq \Phi(x) + \langle\nabla\Phi(x), y - x\rangle + \omega_r(\|x - y\|) \tag{24}$$

for all $x, y \in \mathcal{H}$ such that $\|x\| \le r$ and $\|y\| \le r$. Uniformly convex functions are strictly convex, so the set of their minimum points is at most a singleton.

Theorem 3.3 *Let Φ be uniformly convex on bounded sets with* argmin $\Phi \ne \emptyset$, *and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of* (1) *with $\alpha > 3$. Then $x(t)$ converges strongly, as $t \to +\infty$, to the unique $x^* \in$ argmin Φ.*

Proof Recall that the trajectory $x(\cdot)$ is bounded, by part (ii) in Theorem 2.14. Let $r > 0$ be such that x is contained in the ball of radius r centered at the origin. This ball also contains x^*, which is the weak limit of the trajectory in view of the weak lower-semicontinuity of the norm and Theorem 2.16. Writing $y = x(t)$ and $x = x^*$ in (24), we obtain

$$\omega_r (\|x(t) - x^*\|) \le \Phi(x(t)) - \min \Phi.$$

The right-hand side tends to 0 as $t \to +\infty$ by virtue of Theorem 2.3. It follows that $x(t)$ converges strongly to x^* as $t \to +\infty$. □

Let us recall that a function $\Phi : \mathcal{H} \to \mathbb{R}$ is *strongly convex* if there exists a positive constant μ such that

$$\Phi(y) \ge \Phi(x) + \langle \nabla \Phi(x), y - x \rangle + \frac{\mu}{2} \|x - y\|^2$$

for all $x, y \in \mathcal{H}$. Clearly, strongly convex functions are uniformly convex on bounded sets. However, interestingly, convergence rates increase indefinitely with larger values of α for these functions.

Theorem 3.4 *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be strongly convex, and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of* (1) *with $\alpha > 3$. Then $x(t)$ converges strongly, as $t \to +\infty$, to the unique element $x^* \in$ argmin Φ. Moreover*

$$\Phi(x(t)) - \min \Phi = \mathcal{O}\left(t^{-\frac{2}{3}\alpha}\right), \qquad \|x(t) - x^*\|^2 = \mathcal{O}\left(t^{-\frac{2}{3}\alpha}\right), \qquad and$$
$$\|\dot{x}(t)\|^2 = \mathcal{O}\left(t^{-\frac{2}{3}\alpha}\right). \tag{25}$$

Proof Strong convergence follows from Theorem 3.3 because strongly convex functions are uniformly convex on bounded sets. From (10) and the strong convexity of Φ, we deduce that

$$\frac{d}{dt}\mathcal{E}_\lambda^p(t) \le (p + 2 - \lambda)t^{p+1}(\Phi(x) - \min \Phi) - \lambda(\alpha - \lambda - 1 - p)t^p \langle x - x^*, \dot{x} \rangle$$
$$- \frac{\lambda}{2}(\mu t^2 - p\lambda)t^{p-1}\|x - x^*\|^2 - \left(\alpha - \lambda - 1 - \frac{p}{2}\right)t^{p+1}\|\dot{x}\|^2$$

for any $\lambda \ge 0$ and any $p \ge 0$. Now fix $p = \frac{2}{3}(\alpha - 3)$ and $\lambda = \frac{2}{3}\alpha$, so that $p + 2 - \lambda = \alpha - \lambda - 1 - p/2 = 0$ and $\alpha - \lambda - 1 - p = -p/2$. The above inequality becomes

$$\frac{d}{dt}\mathcal{E}_\lambda^p(t) \le \frac{\lambda p}{2}t^p \langle x - x^*, \dot{x} \rangle - \frac{\lambda}{2}(\mu t^2 - p\lambda)t^{p-1}\|x - x^*\|^2.$$

Define $t_1 = \max\left\{t_0, \sqrt{\frac{p\lambda}{\mu}}\right\}$, so that

$$\frac{d}{dt}\mathcal{E}_\lambda^p(t) \le \frac{\lambda p}{2}t^p\langle x - x^*, \dot{x}\rangle$$

for all $t \ge t_1$. Integrate this inequality from t_1 to t (use integration by parts on the right-hand side) to get

$$\mathcal{E}_\lambda^p(t) \le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{4}\left(t^p\|x(t) - x^*\|^2 - t_1^p\|x(t_1) - x^*\|^2 \right.$$
$$\left. - p\int_{t_1}^t s^{p-1}\|x(s) - x^*\|^2 ds\right).$$

Hence,

$$\mathcal{E}_\lambda^p(t) \le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{4}t^p\|x(t) - x^*\|^2 \le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{2\mu}t^p(\Phi(x(t)) - \min\Phi), \quad (26)$$

in view of the strong convexity of Φ. By the definition of $\mathcal{E}_\lambda^p$, we have

$$t^{p+2}((\Phi(x(t)) - \min\Phi) \le \mathcal{E}_\lambda^p(t) \le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{2\mu}t^p(\Phi(x(t)) - \min\Phi).$$

Dividing by t^{p+2} and using the definition of t_1, along with the fact that $t \ge t_1$, we obtain

$$\Phi(x(t)) - \min\Phi \le \mathcal{E}_\lambda^p(t_1)t^{-p-2} + \frac{\lambda p}{2\mu}t^{-2}(\Phi(x(t)) - \min\Phi)$$
$$\le \mathcal{E}_\lambda^p(t_1)t^{-p-2} + \frac{\lambda p}{2\mu}t_1^{-2}(\Phi(x(t)) - \min\Phi)$$
$$\le \mathcal{E}_\lambda^p(t_1)t^{-p-2} + \frac{1}{2}(\Phi(x(t)) - \min\Phi).$$

Recalling that $p = \frac{2}{3}(\alpha - 3)$ and $\lambda = \frac{2}{3}\alpha$, we deduce that

$$\Phi(x(t)) - \min\Phi \le 2\mathcal{E}_\lambda^p(t_1)t^{-p-2} = \left[2\mathcal{E}_{\frac{2}{3}\alpha}^{\frac{2}{3}(\alpha-3)}(t_1)\right]t^{-\frac{2}{3}\alpha}. \quad (27)$$

The strong convexity of Φ then gives

$$\|x(t) - x^*\|^2 \le \frac{2}{\mu}(\Phi(x(t)) - \min\Phi) \le \left[\frac{4}{\mu}\mathcal{E}_\lambda^p(t_1)\right]t^{-p-2}$$
$$= \left[\frac{4}{\mu}\mathcal{E}_{\frac{2}{3}\alpha}^{\frac{2}{3}(\alpha-3)}(t_1)\right]t^{-\frac{2}{3}\alpha}. \quad (28)$$

Inequalities (27) and (28) settle the first two points in (25).

Now, using (26) and (27), we derive

$$\mathcal{E}_\lambda^p(t) \le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{2\mu} t^p (\Phi(x(t)) - \min \Phi) \le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{\mu} \mathcal{E}_\lambda^p(t_1) t^{-2}$$

$$\le \mathcal{E}_\lambda^p(t_1) + \frac{\lambda p}{\mu} \mathcal{E}_\lambda^p(t_1) t_1^{-2} \le 2\mathcal{E}_\lambda^p(t_1).$$

The definition of $\mathcal{E}_\lambda^p$ then gives

$$\frac{t^p}{2} \|\lambda \left(x(t) - x^*\right) + t\dot{x}(t)\|^2 \le \mathcal{E}_\lambda^p(t) \le 2\mathcal{E}_\lambda^p(t_1).$$

Hence

$$\|\lambda \left(x(t) - x^*\right) + t\dot{x}(t)\|^2 \le 4t^{-p} \mathcal{E}_\lambda^p(t_1),$$

and

$$t\|\dot{x}(t)\| \le 2t^{-p/2} \sqrt{\mathcal{E}_\lambda^p(t_1)} + \lambda \|x(t) - x^*\|.$$

But using (28), we deduce that

$$\lambda \|x(t) - x^*\| \le \frac{2\lambda}{\sqrt{\mu}} t^{-p/2-1} \sqrt{\mathcal{E}_\lambda^p(t_1)}.$$

The last two inequalities together give

$$t\|\dot{x}(t)\| \le 2t^{-p/2} \sqrt{\mathcal{E}_\lambda^p(t_1)} \left(1 + \frac{\lambda t^{-1}}{\sqrt{\mu}}\right) \le 2t^{-p/2} \sqrt{\mathcal{E}_\lambda^p(t_1)} \left(1 + \sqrt{\frac{\lambda}{p}}\right).$$

Taking squares, and rearranging the terms, we obtain

$$\|\dot{x}(t)\|^2 \le \left[4\left(1 + \sqrt{\frac{\alpha}{\alpha-3}}\right)^2 \mathcal{E}_{\frac{2}{3}\alpha}^{\frac{2}{3}(\alpha-3)}(t_1)\right] t^{-\frac{2}{3}\alpha},$$

which shows the last point in (25) and completes the proof.   □

The preceding theorem extends [35, Theorem 4.2], which states that if $\alpha > 9/2$, then $\Phi(x(t)) - \min \Phi = \mathcal{O}(1/t^3)$.

4 Asymptotic behavior of the trajectory under perturbations

In this section, we analyze the asymptotic behavior, as $t \to +\infty$, of the solutions of the differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = g(t). \tag{29}$$

From the Cauchy–Lipschitz–Picard Theorem (see, for instance, [7, Theorem 17.1. 2b)]), for any initial condition $x(t_0) = x_0 \in \mathcal{H}$, $\dot{x}(t_0) = v_0 \in \mathcal{H}$, we deduce the existence and uniqueness of a maximal local solution x to (29), with $\dot{x}$ locally absolutely continuous, if $\nabla \Phi$ is Lipschitz-continuous on bounded sets and g is locally integrable. If Φ is minorized, the global existence follows from the energy estimate proved in Lemma 4.1, in the next subsection.

This being said, our main concern here is to obtain sufficient conditions on the perturbation g in order to guarantee that the convergence properties established in the preceding sections are preserved. The analysis follows very closely the arguments given in Sects. 2 and 3. It is developed in the same general setting, just assuming that $\Phi : \mathcal{H} \to \mathbb{R}$ is a continuously differentiable convex function. Therefore, we shall state the main results and sketch the proofs, underlining the parts where additional techniques are required.

4.1 Energy estimates and minimization property

The following result is obtained by considering the global energy of the system, and showing that it is a strict Lyapunov function:

Lemma 4.1 *Let Φ be bounded from below, and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of (29) with $\alpha > 0$ and $\int_{t_0}^{\infty} \|g(t)\| \, dt < +\infty$. Then, $\sup_{t>t_0} \|\dot{x}(t)\| < +\infty$ and $\int_{t_0}^{\infty} \frac{1}{\tau} \|\dot{x}(\tau)\|^2 d\tau < +\infty$. Moreover, $\lim_{t \to +\infty} \Phi(x(t)) = \inf_{\mathcal{H}} \Phi$.*

Proof Set $T > t_0$. For $t_0 \leq t \leq T$, define the energy function

$$W_T(t) := \frac{1}{2} \|\dot{x}(t)\|^2 + \left(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi \right) + \int_t^T \langle \dot{x}(\tau), g(\tau) \rangle d\tau. \qquad (30)$$

Since $\dot{x}$ is continuous and g is integrable, the function W_T is well defined. Derivating W_T with respect to time, and using (29), we obtain

$$\dot{W}_T(t) = \langle \dot{x}(t), \ddot{x}(t) + \nabla \Phi(x(t)) - g(t) \rangle = \langle \dot{x}(t), -\frac{\alpha}{t} \dot{x}(t) \rangle = -\frac{\alpha}{t} \|\dot{x}(t)\|^2 \leq 0.$$

Hence $W_T(\cdot)$ is a decreasing function. In particular, $W_T(t) \leq W_T(t_0)$, that is

$$\frac{1}{2} \|\dot{x}(t)\|^2 + \left(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi \right) + \int_t^T \langle \dot{x}(\tau), g(\tau) \rangle d\tau$$

$$\leq \frac{1}{2} \|\dot{x}(t_0)\|^2 + \left(\Phi(x_0) - \inf_{\mathcal{H}} \Phi \right) + \int_{t_0}^T \langle \dot{x}(\tau), g(\tau) \rangle d\tau.$$

As a consequence,

$$\frac{1}{2} \|\dot{x}(t)\|^2 \leq \frac{1}{2} \|\dot{x}(t_0)\|^2 + \left(\Phi(x_0) - \inf_{\mathcal{H}} \Phi \right) + \int_{t_0}^t \|\dot{x}(\tau)\| \|g(\tau)\| d\tau.$$

Applying Lemma 5.13, we obtain

$$\|\dot{x}(t)\| \leq \left(\|\dot{x}(t_0)\|^2 + 2\left(\Phi(x_0) - \inf_{\mathcal{H}} \Phi\right)\right)^{\frac{1}{2}} + \int_{t_0}^{t} \|g(\tau)\| d\tau.$$

It follows that

$$\sup_{t > t_0} \|\dot{x}(t)\| \leq \left(\|\dot{x}(t_0)\|^2 + 2\left(\Phi(x_0) - \inf_{\mathcal{H}} \Phi\right)\right)^{\frac{1}{2}} + \int_{t_0}^{\infty} \|g(\tau)\| d\tau < +\infty. \tag{31}$$

As a consequence, we may define a function $W :]t_0, +\infty[\to \mathbb{R}$ by

$$W(t) := \frac{1}{2}\|\dot{x}(t)\|^2 + \left(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi\right) + \int_{t}^{\infty} \langle \dot{x}(\tau), g(\tau) \rangle d\tau$$

$$\geq -\left[\sup_{t > t_0} \|\dot{x}(t)\|\right] \int_{t_0}^{\infty} \|g(\tau)\| d\tau,$$

by (31). From the definition of W_T and W, we have

$$\dot{W}(t) = \dot{W}_T(t) = -\frac{\alpha}{t} \|\dot{x}(t)\|^2. \tag{32}$$

Integrating from t_0 to t, and using (31), we obtain

$$\int_{t_0}^{\infty} \frac{\alpha}{\tau} \|\dot{x}(\tau)\|^2 d\tau = W(t_0) - W(t) \leq \frac{1}{2}\|\dot{x}(t_0)\|^2 + \left(\Phi(x(t_0)) - \inf_{\mathcal{H}} \Phi\right)$$

$$+ \left[\sup_{t > t_0} \|\dot{x}(t)\|\right] \int_{t_0}^{\infty} \|g(\tau)\| d\tau < +\infty,$$

which gives a bound for the second improper integral.

For the minimization property, consider the function $h :]t_0, +\infty[\to \mathbb{R}$, defined by $h(t) = \frac{1}{2}\|x(t) - z\|^2$, where z is an arbitrary element of $\mathcal{H}$. We can easily verify that

$$\ddot{h}(t) + \frac{\alpha}{t} \dot{h}(t) = \|\dot{x}(t)\|^2 - \langle \nabla \Phi(x(t)), x(t) - z \rangle$$

$$+ \langle g(t), x(t) - z \rangle.$$

By convexity of Φ, we obtain

$$\ddot{h}(t) + \frac{\alpha}{t} \dot{h}(t) + \Phi(x(t)) - \Phi(z) \leq \|\dot{x}(t)\|^2 + \langle g(t), x(t) - z \rangle. \tag{33}$$

Recall that the function W defined above is nonincreasing and bounded from below. Hence, $W(t)$ converges, as $t \to +\infty$, to some $W_\infty \in \mathbb{R}$. Moreover, using the definition of W in (33), we deduce that

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + W(t) + \inf \Phi - \Phi(z)$$

$$\leq \frac{3}{2}\|\dot{x}(t)\|^2 + \langle g(t), x(t) - z\rangle + \int_t^\infty \langle \dot{x}(s), g(s)\rangle ds. \tag{34}$$

Setting $B_\infty = W_\infty + \inf \Phi - \Phi(z)$, we may write

$$B_\infty \leq \frac{3}{2}\|\dot{x}(t)\|^2 + \|g(t)\|\|x(t) - z\|$$

$$+ \left(\sup_{t \geq t_0} \|\dot{x}(t)\|\right) \int_t^\infty \|g(s)\|ds - \frac{1}{t^\alpha}\frac{d}{dt}(t^\alpha \dot{h}(t)).$$

Multiplying this last equation by $\frac{1}{t}$, and integrating between t_0 and $\theta > t_0$, we get

$$B_\infty \ln\left(\frac{\theta}{t_0}\right) \leq \frac{3}{2}\int_{t_0}^\theta \frac{1}{t}\|\dot{x}(t)\|^2 dt + \int_{t_0}^\theta \frac{\|g(t)\|\|x(t) - z\|}{t}dt$$

$$+ \left(\sup_{t \geq t_0} \|\dot{x}(t)\|\right) \int_{t_0}^\theta \left(\frac{1}{t}\int_t^\infty \|g(s)\|ds\right) dt$$

$$- \int_{t_0}^\theta \frac{1}{t^{\alpha+1}}\frac{d}{dt}(t^\alpha \dot{h}(t))dt.$$

Let us estimate the integrals in the second member of the last inequality:

(1) The first term is finite, in view of Lemma 4.1.
(2) The second term is also finite, since the relation $\|x(t) - z\| \leq \|x(t_0) - z\| + \int_{t_0}^t \|\dot{x}(s)\|ds$ implies

$$\int_{t_0}^\theta \frac{\|g(t)\|\|x(t) - z\|}{t}dt \leq \left(\frac{\|x_0 - z\|}{t_0} + \sup_{t \geq t_0}\|\dot{x}(t)\|\right) \int_{t_0}^{+\infty} \|g(t)\|dt < +\infty.$$

(3) For the third term, integration by parts gives

$$\int_{t_0}^\theta \left(\frac{1}{t}\int_t^\infty \|g(s)\|ds\right) dt = \ln \theta \int_\theta^\infty \|g(s)\|ds - \ln t_0 \int_{t_0}^\infty \|g(s)\|ds$$

$$+ \int_{t_0}^\theta \|g(t)\| \ln t\, dt.$$

(4) For the fourth term, set $I = \int_{t_0}^\theta \frac{1}{t^{\alpha+1}}\frac{d}{dt}(t^\alpha \dot{h}(t))dt$, and integrate by parts twice to obtain

$$I = \left[\frac{1}{t}\dot{h}(t)\right]_{t_0}^{\theta} + (\alpha + 1)\int_{t_0}^{\theta}\frac{1}{t^2}\dot{h}(t)dt$$

$$= C_0 + \frac{1}{\theta}\dot{h}(\theta) + \frac{(1+\alpha)}{\theta^2}h(\theta)$$

$$+ 2(1+\alpha)\int_{t_0}^{\theta}\frac{1}{t^3}h(t)dt \geq C_0 + \frac{1}{\theta}\dot{h}(\theta),$$

for some constant C_0, because $h \geq 0$. Finally, notice that

$$|\dot{h}(\theta)| = |\langle\dot{x}(\theta), x(\theta) - z\rangle| \leq \sup_{t\geq t_0}\|\dot{x}(t)\| \left(\|x(0) - z\| + \theta\sup_{t\geq t_0}\|\dot{x}(t)\|\right).$$

Collecting the above results, we deduce that

$$B_\infty \ln\left(\frac{\theta}{t_0}\right) \leq C_1 + \ln\theta\int_{\theta}^{\infty}\|g(s)\|ds + \left(\sup_{t\geq t_0}\|\dot{x}(t)\|\right)\int_{t_0}^{\theta}\|g(t)\|\ln t\,dt,$$

for some other constant C_1. Dividing by $\ln(\frac{\theta}{t_0})$, and letting $\theta \to +\infty$, we conclude that $B_\infty \leq 0$, by using Lemma 5.11 with $\psi(t) = \ln t$. This implies that $W_\infty \leq \Phi(z) - \inf\Phi$ for every $z \in \mathcal{H}$, which leads to $W_\infty \leq 0$.

On the other hand, it is easy to see that

$$W(t) \geq \Phi(x(t)) - \inf\Phi - \left(\sup_{t\geq t_0}\|\dot{x}(t)\|\right)\int_{t}^{+\infty}g(s)ds.$$

Passing to the limit, as $t \to +\infty$, we deduce that

$$0 \geq W_\infty \geq \limsup\Phi(x(t)) - \inf\Phi.$$

Since we always have $\inf\Phi \leq \liminf\Phi(x(t))$, we conclude that $\lim_{t\to+\infty}\Phi(x(t)) = \inf\Phi$. □

4.2 Fast convergence of the values

We are now in position to prove the following:

Theorem 4.2 *Let* $\operatorname{argmin}\Phi \neq \emptyset$, *and let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution of* (29) *with* $\alpha \geq 3$ *and* $\int_{t_0}^{\infty}t\|g(t)\|dt < +\infty$. *Then* $\Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}\left(\frac{1}{t^2}\right)$.

Proof The proof follows the arguments used for Theorem 2.7. Take $x^* \in S = \operatorname{argmin}\Phi$. For $t_0 \leq t \leq T$, define the energy function

$$\mathcal{E}_{\alpha,g,T}(t) := \frac{2}{\alpha - 1}t^2(\Phi(x(t)) - \inf_{\mathcal{H}}\Phi) + (\alpha - 1)\|x(t) - x^*$$

$$+ \frac{t}{\alpha - 1}\dot{x}(t)\|^2 + 2\int_{t}^{T}\tau\langle x(\tau) - x^* + \frac{\tau}{\alpha - 1}\dot{x}(\tau), g(\tau)\rangle d\tau.$$

Let us show that

$$\dot{\mathcal{E}}_{\alpha,g,T}(t) + 2\frac{\alpha - 3}{\alpha - 1}t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \leq 0.$$

Derivation of $\mathcal{E}_{\alpha,g,T}$ gives

$$\begin{aligned}
\dot{\mathcal{E}}_{\alpha,g,T}(t) := {} & \frac{4}{\alpha - 1}t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + \frac{2}{\alpha - 1}t^2\langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\
& - 2t\left\langle x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t), g(t)\right\rangle \\
& + 2(\alpha - 1)\left\langle x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t), \dot{x}(t) + \frac{1}{\alpha - 1}\dot{x}(t) + \frac{t}{\alpha - 1}\ddot{x}(t)\right\rangle \\
= {} & \frac{4}{\alpha - 1}t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + \frac{2}{\alpha - 1}t^2\langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\
& + 2(\alpha - 1)\left\langle x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t), \frac{t}{\alpha - 1}\left(\frac{\alpha}{t}\dot{x}(t) + \ddot{x}(t) - g(t)\right)\right\rangle \\
= {} & \frac{4}{\alpha - 1}t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + \frac{2}{\alpha - 1}t^2\langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\
& - 2t\left\langle x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t), \nabla\Phi(x(t))\right\rangle \\
= {} & \frac{4}{\alpha - 1}t\left(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi\right) - 2t\langle x(t) - x^*, \nabla\Phi(x(t))\rangle.
\end{aligned}$$

Using the subdifferential inequality for Φ, and rearranging the terms, we obtain

$$\dot{\mathcal{E}}_{\alpha,g,T}(t) + 2\frac{\alpha - 3}{\alpha - 1}t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) \leq 0. \tag{35}$$

As a consequence, for $\alpha \geq 3$, the function $\mathcal{E}_{\alpha,g,T}$ is nonincreasing. In particular, $\mathcal{E}_{\alpha,g,T}(t) \leq \mathcal{E}_{\alpha,g,T}(t_0)$, which gives

$$\begin{aligned}
& \frac{2}{\alpha - 1}t^2(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + (\alpha - 1)\|x(t) - x^* \\
& + \frac{t}{\alpha - 1}\dot{x}(t)\|^2 \leq C + 2\int_{t_0}^{t} \tau\langle x(\tau) - x^* + \frac{\tau}{\alpha - 1}\dot{x}(\tau), g(\tau)\rangle d\tau, \tag{36}
\end{aligned}$$

with

$$C = \frac{2}{\alpha - 1}t_0^2(\Phi(x_0) - \inf_{\mathcal{H}} \Phi) + (\alpha - 1)\|x_0 - x^* + \frac{t_0}{\alpha - 1}\dot{x}(t_0)\|^2.$$

From (36), we infer that

$$\frac{1}{2}\|x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t)\|^2 \leq \frac{C}{2(\alpha - 1)} + \frac{1}{\alpha - 1}\int_{t_0}^{t} \|x(\tau) - x^*$$
$$+ \frac{\tau}{\alpha - 1}\dot{x}(\tau)\|\|\tau g(\tau)\|d\tau.$$

Applying Lemma 5.13, we obtain

$$\|x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t)\| \leq \left(\frac{C}{\alpha - 1}\right)^{\frac{1}{2}} + \frac{1}{\alpha - 1}\int_{t_0}^{t} \tau\|g(\tau)\|d\tau, \qquad (37)$$

and so

$$\sup_{t \geq t_0} \|x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t)\| < +\infty. \qquad (38)$$

Using (37) in (36), we conclude that

$$\frac{2}{\alpha - 1}t^2(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) \leq C + 2\left(\left(\frac{C}{\alpha - 1}\right)^{\frac{1}{2}} + \frac{1}{\alpha - 1}\int_{t_0}^{\infty} \tau\|g(\tau)\|d\tau\right)$$
$$\int_{t_0}^{\infty} \tau\|g(\tau)\|d\tau,$$

and the result follows.   □

Remark 4.3 As a consequence, the energy function

$$\mathcal{E}_{\alpha,g}(t) := \frac{2}{\alpha - 1}t^2(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + (\alpha - 1)\|x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t)\|^2$$
$$+ 2\int_{t}^{+\infty} \tau\langle x(\tau) - x^* + \frac{\tau}{\alpha - 1}\dot{x}(\tau), g(\tau)\rangle d\tau$$

is well defined on $[t_0, +\infty[$, and is a Lyapunov function for the dynamical system (29).

4.3 Convergence of the trajectories

In the case $\alpha > 3$, provided that the second member $g(t)$ is sufficiently small for large t, we are going to show the convergence of the trajectories of (29), as it occurs for the unperturbed system studied in the previous sections

Theorem 4.4 *Let* $\mathrm{argmin}\,\Phi \neq \emptyset$, *and let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution of* (29) *with* $\alpha > 3$ *and* $\int_{t_0}^{\infty} t\,\|g(t)\|\,dt < +\infty$. *Then,* $x(t)$ *converges weakly, as* $t \to +\infty$, *to a point in* $\mathrm{argmin}\,\Phi$.

Proof **Step 1** Recall, from the proof of Theorem 4.2, that the energy function $\mathcal{E}_{\alpha,g}$ defined in Remark 4.3 satisfies

$$\dot{\mathcal{E}}_{\alpha,g}(t) + 2\frac{\alpha - 3}{\alpha - 1}t\left(\Phi(x(t)) - \inf_{\mathcal{H}}\Phi\right) \le 0.$$

Integrating this inequality, we obtain

$$\mathcal{E}_{\alpha,g}(t) + 2\frac{\alpha - 3}{\alpha - 1}\int_{t_0}^{t}\tau\left(\Phi(x(\tau)) - \inf_{\mathcal{H}}\Phi\right)d\tau \le \mathcal{E}_{\alpha,g}(t_0).$$

By the definition of $\mathcal{E}_{\alpha,g}$, and neglecting its nonnegative terms, we infer that

$$2\int_{t}^{+\infty}\tau\langle x(\tau) - x^* + \frac{\tau}{\alpha - 1}\dot{x}(\tau), g(\tau)\rangle d\tau$$

$$+2\frac{\alpha - 3}{\alpha - 1}\int_{t_0}^{t}\tau\left(\Phi(x(\tau)) - \inf_{\mathcal{H}}\Phi\right)d\tau \le \mathcal{E}_{\alpha,g}(t_0),$$

and so

$$2\frac{\alpha - 3}{\alpha - 1}\int_{t_0}^{t}\tau\left(\Phi(x(\tau)) - \inf_{\mathcal{H}}\Phi\right)d\tau$$

$$\le \mathcal{E}_{\alpha,g}(t_0) + 2\int_{t_0}^{+\infty}\left\|x(\tau) - x^* + \frac{\tau}{\alpha - 1}\dot{x}(\tau)\right\|\|\tau g(\tau)\|\rangle d\tau$$

$$\le \mathcal{E}_{\alpha,g}(t_0) + 2\sup_{t\ge t_0}\left\|x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t)\right\|\int_{t_0}^{+\infty}\|\tau g(\tau)\|\rangle d\tau.$$

by (38). Since $\alpha > 3$, we deduce that

$$\int_{t_0}^{+\infty}\tau(\Phi(x(\tau)) - \inf_{\mathcal{H}}\Phi)d\tau < +\infty. \tag{39}$$

Step 2 Let us show that

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

By taking the scalar product of (29) by $t^2\dot{x}(t)$, we get

$$t^2\langle\ddot{x}(t), \dot{x}(t)\rangle + \alpha t\|\dot{x}(t)\|^2 + t^2\langle\nabla\Phi(x(t)), \dot{x}(t)\rangle = t^2\langle g(t), \dot{x}(t)\rangle.$$

Using the Chain Rule and the Cauchy–Schwarz inequality, we obtain

$$\frac{1}{2}t^2\frac{d}{dt}\|\dot{x}(t)\|^2 + \alpha t\|\dot{x}(t)\|^2 + t^2\frac{d}{dt}\Phi(x(t)) \le \|tg(t)\|\|t\dot{x}(t)\|.$$

Integration by parts yields

$$\frac{t^2}{2}\|\dot{x}(t)\|^2 - \frac{t_0{}^2}{2}\|\dot{x}(t_0)\|^2 - \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds + \alpha \int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds$$

$$+ t^2(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) - t_0{}^2 \left(\Phi(x(t_0)) - \inf_{\mathcal{H}} \Phi \right)$$

$$- 2\int_{t_0}^{t} s \left(\Phi(x(s)) - \inf_{\mathcal{H}} \Phi \right) ds \le \int_{t_0}^{t} \|sg(s)\| \|s\dot{x}(s)\| ds.$$

As a consequence,

$$\frac{t^2}{2}\|\dot{x}(t)\|^2 + (\alpha - 1)\int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds \le C_0$$

$$+ 2\int_{t_0}^{t} s(\Phi(x(s)) - \inf_{\mathcal{H}} \Phi)ds + \int_{t_0}^{t} \|sg(s)\| \|s\dot{x}(s)\| ds \qquad (40)$$

for some constant C_0 depending only on the Cauchy data. Since $\int_{t_0}^{\infty} s(\Phi(x(s)) - \inf_{\mathcal{H}} \Phi)ds < +\infty$ by (39), and $\alpha > 1$, we deduce that

$$\frac{1}{2}\|t\dot{x}(t)\|^2 \le C_1 + \int_{t_0}^{t} \|sg(s)\| \|s\dot{x}(s)\| ds \qquad (41)$$

for some other constant C_1, which we may assume to be nonnegative. Applying Lemma 5.13, we obtain

$$\|t\dot{x}(t)\| \le \sqrt{2C_1} + \int_{t_0}^{t} \|sg(s)\| ds,$$

and so

$$\sup_{t \ge t_0} \|t\dot{x}(t)\| < +\infty. \qquad (42)$$

Returning to (40), we deduce that

$$(\alpha - 1)\int_{t_0}^{t} s\|\dot{x}(s)\|^2 ds \le C + 2\int_{t_0}^{\infty} s(\Phi(x(s)) - \inf_{\mathcal{H}} \Phi)ds$$

$$+ \sup_{t \ge t_0} \|t\dot{x}(t)\| \int_{t_0}^{\infty} \|sg(s)\| ds, \qquad (43)$$

which gives

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

Moreover, combining (38) and (42), we deduce that

$$\sup_{t \geq t_0} \|x(t)\| < +\infty, \tag{44}$$

and the trajectory x is bounded.

Step 3 As before, we prove the weak convergence by means of Opial's Lemma 5.7. Take $x^* \in \operatorname{argmin} \Phi$, and define $h : [0, +\infty[\to \mathbb{R}^+$ by

$$h(t) = \frac{1}{2}\|x(t) - x^*\|^2.$$

By the Chain Rule,

$$\dot{h}(t) = \langle x(t) - x^*, \dot{x}(t)\rangle,$$
$$\ddot{h}(t) = \langle x(t) - x^*, \ddot{x}(t)\rangle + \|\dot{x}(t)\|^2.$$

Combining these two equations, and using (29), we obtain

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) = \|\dot{x}(t)\|^2 + \langle x(t) - x^*, \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t)\rangle,$$
$$= \|\dot{x}(t)\|^2 + \langle x(t) - x^*, -\nabla\Phi(x(t)) + g(t)\rangle.$$

By the monotonicity of $\nabla\Phi$ and the fact that $\nabla\Phi(x^*) = 0$, we have

$$\langle x(t) - x^*, -\nabla\Phi(x(t))\rangle \leq 0, \tag{45}$$

and we infer that

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) \leq \|\dot{x}(t)\|^2 + \|x(t) - x^*\|\|g(t)\|.$$

Equivalently

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) \leq k(t), \tag{46}$$

with

$$k(t) := \|\dot{x}(t)\|^2 + \|x(t) - x^*\|\|g(t)\| \leq \|\dot{x}(t)\|^2 + C_2\|g(t)\|$$

because the trajectory x is bounded, by (44). Recall that $\int_{t_0}^{+\infty} t\|g(t)\|dt < +\infty$ by assumption, and $\int_{t_0}^{\infty} t\|\dot{x}(t)\|^2 dt < +\infty$ by (13). Hence, the function $t \mapsto tk(t)$ belongs to $L^1(t_0, +\infty)$. Applying Lemma 5.9, with $w(t) = h(t)$, we deduce that $\dot{h}^+(t) \in L^1(t_0, +\infty)$, which implies that the limit of $h(t)$ exists, as $t \to +\infty$. This proves item (i) of Opial's Lemma 5.7. For item (ii), observe that every weak limit point of $x(t)$ as $t \to +\infty$ must minimize Φ, since $\lim_{t\to+\infty} \Phi(x(t)) = \inf \Phi$. $\qquad\square$

Remark 4.5 Throughout the proof of Theorem 4.4, we proved that

$$\int_{t_0}^{\infty} t \left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi \right) dt < +\infty, \quad \text{and} \quad \int_{t_0}^{\infty} t \|\dot{x}(t)\|^2 dt < +\infty.$$

We also proved that $\sup_{t \geq t_0} t \|\dot{x}(t)\| < +\infty$, and hence $\lim_{t \to \infty} \|\dot{x}(t)\| = 0$.

4.4 Strong convergence results

For strong convergence, we have the following:

Theorem 4.6 *Let* $\Phi : \mathcal{H} \to \mathbb{R}$ *be a continuously differentiable convex function, and let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution of* (29) *with* $\alpha > 3$ *and* $\int_{t_0}^{\infty} t \|g(t)\| dt < +\infty$. *Then* $x(t)$ *converges strongly, as* $t \to +\infty$, *in any of the following cases:*

 (i) *The set* argmin Φ *has nonempty interior;*
 (ii) *The function* Φ *is even; or*
 (iii) *The function* Φ *is uniformly convex.*

In order to prove this result, it suffices to adapt the arguments given in Sect. 3 for the unperturbed case. Since it is relatively straightforward, we leave it as an exercise to the reader.

5 Convergence of the associated algorithms

In this section, we analyze the fast convergence properties of the associated Nesterov-type algorithms. To avoid repeating similar arguments, we state the results and develop the proofs directly for the perturbed version.

5.1 A dynamical introduction of the algorithm

Time discretization of dissipative gradient-based dynamical systems leads naturally to algorithms, which, under appropriate assumptions, have similar convergence properties. This approach has been followed successfully in a variety of situations. For a general abstract discussion see [5,6]. For dynamics with inertial features, see [2–4,9]. To cover practical situations involving constraints or non-smooth data, we need to broaden our scope. This leads us to consider the non-smooth structured convex minimization problem

$$\min \{\Psi(x) + \Phi(x) : x \in \mathcal{H}\} \tag{47}$$

where $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous convex function; and $\Phi : \mathcal{H} \to \mathbb{R}$ is a continuously differentiable convex function, whose gradient is Lipschitz continuous.

The optimal solutions of (47) satisfy

$$\partial \Psi(x) + \nabla \Phi(x) \ni 0,$$

where $\partial \Psi$ is the subdifferential of Ψ, in the sense of convex analysis. In order to adapt our dynamic to this non-smooth situation, we will consider the corresponding differential inclusion

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \partial \Psi(x(t)) + \nabla \Phi(x(t)) \ni g(t). \tag{48}$$

This dynamical system is within the following framework

$$\ddot{x}(t) + a(t)\dot{x}(t) + \partial \Theta(x(t)) \ni g(t), \tag{49}$$

where $\Theta : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous convex function, and $a(\cdot)$ is a positive damping parameter.

It is interesting to establish the asymptotic properties, as $t \to +\infty$, of the solutions of the differential inclusion (48). Beyond global existence issues, one must check that the Lyapunov analysis is still valid. In view of the validity of the subdifferential inequality for convex functions, the (generalized) chain rule for derivatives over curves (see [15]), most results presented in the previous sections can be transposed to this more general context, except for the stabilization of the acceleration, which relies on the Lipschitz character of the gradient. However, a detailed study of this differential inclusion goes far beyond the scope of the present article. See [8] for some results in the case of a fixed positive damping parameter, i.e., $a(t) = \gamma > 0$ fixed, and $g = 0$. Thus, setting $\Theta(x) = \Psi(x) + \Phi(x)$, we can reasonably assume that, for $\alpha > 3$, and $\int_{t_0}^{+\infty} t\|g(t)\|dt < +\infty$, for each trajectory of (48), there is rapid convergence of the values

$$\Theta(x(t)) - \min \Theta \leq \frac{C}{t^2},$$

and weak convergence of the trajectory to an optimal solution.

We shall use these ideas as a guideline, in order to introduce corresponding fast converging algorithms, making the link with Nesterov [26–29] and Beck–Teboulle [14]; and so, extending the recent works of Chambolle–Dossal [19] and Su–Boyd–Candès [35] to the perturbed case.

In order to preserve the fast convergence properties of the dynamical system (48), we are going to discretize it *implicitly* with respect to the nonsmooth function Ψ, and *explicitly* with respect to the smooth function Φ.

Taking a fixed time step size $h > 0$, and setting $t_k = kh$, $x_k = x(t_k)$ the implicit/explicit finite difference scheme for (48) gives

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2}(x_k - x_{k-1}) + \partial \Psi(x_{k+1}) + \nabla \Phi(y_k) \ni g_k, \tag{50}$$

where y_k is a linear combination of x_k and x_{k-1}, that will be made precise later on. After developing (50), we obtain

$$x_{k+1} + h^2 \partial \Psi(x_{k+1}) \ni x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) - h^2 \nabla \Phi(y_k) + h^2 g_k. \quad (51)$$

A natural choice for y_k leading to a simple formulation of the algorithm (other choices are possible, offering new directions of research for the future) is

$$y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}). \quad (52)$$

Using the classical proximal operator (equivalently, the resolvent of the maximal monotone operator $\partial \Psi$)

$$\text{prox}_{\gamma \Psi}(x) = \text{argmin}_{\xi \in \mathcal{H}} \left\{ \Psi(\xi) + \frac{1}{2\gamma} \|\xi - x\|^2 \right\} = (I + \gamma \partial \Psi)^{-1}(x), \quad (53)$$

and setting $s = h^2$, the algorithm can be written as

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s(\nabla \Phi(y_k) - g_k)). \end{cases} \quad (54)$$

For practical purposes, and in order to fit with the existing literature on the subject, it is convenient to work with the following equivalent formulation

$$\begin{cases} y_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s(\nabla \Phi(y_k) - g_k)). \end{cases} \quad (55)$$

Indeed, we have $\frac{k-1}{k+\alpha-1} = 1 - \frac{\alpha}{k+\alpha-1}$. When α is an integer, we obtain the same sequences (x_k) and (y_k), up to the reindexation $k \mapsto k+\alpha-1$. For general $\alpha > 0$, we can easily verify that the algorithm (55) is still associated with the dynamical system (48).

This algorithm is within the scope of the proximal-based inertial algorithms [3,24, 25], and forward-backward methods. In the unperturbed case, $g_k = 0$, it has been recently considered by Chambolle–Dossal [19] and Su–Boyd–Candès [35]. It enjoys fast convergence properties which are very similar to that of the continuous dynamic.

For $\alpha = 3$ and $g_k = 0$, we recover the classical FISTA algorithm developed by Beck–Teboulle [14], based on the acceleration method introduced by Nesterov [26] in the smooth case and by Güler [21] in the proximal setting:

$$\begin{cases} y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s\nabla \Phi(y_k)). \end{cases} \quad (56)$$

An important question regarding the (FISTA) method, as described in (56), is the convergence of sequences (x_k) and (y_k), which is still an open question. A major

interest to consider the broader context of algorithms (55) is that, for $\alpha > 3$, these sequences converge, even when inexactly computed, provided the errors or perturbations are sufficiently small.

5.2 Fast convergence of the values

We will see that the fast convergence properties of algorithm (54) can be obtained in a parallel way with the convergence analysis in the continuous case in Theorem 4.2.

Theorem 5.1 *Let $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and convex, and let $\Phi : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient. Suppose that $S = \operatorname{argmin}(\Psi + \Phi) \neq \emptyset$, and let (x_k) be a sequence generated by algorithm (55) with $\alpha \geq 3, 0 < s < \frac{1}{L}$, and $\sum_{k \in \mathbb{N}} k \|g_k\| < +\infty$. Then,*

$$(\Psi + \Phi)(x_k) - \min(\Psi + \Phi) \leq \frac{C(\alpha - 1)}{2s\,(k + \alpha - 2)^2}, \tag{57}$$

where

$$
\begin{aligned}
C = &\frac{2s}{\alpha - 1}(\alpha - 2)^2\left(\Theta(x_0) - \Theta^*\right) + (\alpha - 1)\|y_0 - x^*\|^2 \\
&+ 2s\left[\sum_{j=0}^{\infty}(j + \alpha - 1)\|g_j\|\right]\left[\sqrt{\frac{\mathcal{E}(0)}{\alpha - 1}} + \frac{2s}{\alpha - 1}\sum_{j=0}^{\infty}(j + \alpha - 1)\|g_j\|\right]
\end{aligned}
$$

Proof To simplify notations, we set $\Theta = \Psi + \Phi$, and take $x^* \in \operatorname{argmin}\Theta$. As in the continuous case, we shall prove that the energy sequence $(\mathcal{E}(k))$ given by

$$
\begin{aligned}
\mathcal{E}(k) := &\frac{2s}{\alpha - 1}(k + \alpha - 2)^2\left(\Theta(x_k) - \Theta(x^*)\right) + (\alpha - 1)\|z_k - x^*\|^2 \\
&+ \sum_{j=k}^{\infty} 2s\,(j + \alpha - 1)\langle g_j, z_{j+1} - x^*\rangle,
\end{aligned}
\tag{58}
$$

with

$$z_k := \frac{k + \alpha - 1}{\alpha - 1}y_k - \frac{k}{\alpha - 1}x_k, \tag{59}$$

is non-increasing (we shall justify further that it is well defined). Note that $\mathcal{E}(k)$ equals the Lyapunov function considered by Su–Boyd–Candès in [35, Theorem 4.3], plus a perturbation term. For each $y \in \mathcal{H}$, we set $\Phi_k(y) := \Phi(y) - \langle g_k, y\rangle$, and $\Theta_k(y) = \Psi(y) + \Phi_k(y)$. Since $\nabla\Phi_k(y) = \nabla\Phi(y) - g_k$, we deduce that $\nabla\Phi_k$ is still L-Lipschitz continuous. By introducing the operator $G_{s,k} : \mathcal{H} \to \mathcal{H}$, defined by

$$G_{s,k}(y) = \frac{1}{s}\left(y - \operatorname{prox}_{s\Psi}\left(y - s\nabla\Phi_k(y)\right)\right)$$

for each $y \in \mathcal{H}$, we can write

$$\text{prox}_{s\Psi}(y - s\nabla\Phi_k(y)) = y - sG_{s,k}(y),$$

and rewrite algorithm (55) as

$$\begin{cases} y_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}); \\ x_{k+1} = y_k - sG_{s,k}(y_k). \end{cases} \tag{60}$$

The variable z_k, defined in (59), will play an important role. It comes naturally into play as a discrete version of the term $\frac{t}{\alpha-1}\dot{x}(t) + x(t) - x^*$ which enters $\mathcal{E}_{\alpha,g}(t)$. Simple algebraic manipulations give

$$\begin{aligned} z_{k+1} &= \frac{k+\alpha-1}{\alpha-1}x_{k+1} - \frac{k}{\alpha-1}x_k \\ &= \frac{k+\alpha-1}{\alpha-1}(x_{k+1} - x_k) + x_k, \end{aligned} \tag{61}$$

and also

$$\begin{aligned} z_{k+1} &= \frac{k+\alpha-1}{\alpha-1}\left(y_k - sG_{s,k}(y_k)\right) - \frac{k}{\alpha-1}x_k \\ &= z_k - \frac{s}{\alpha-1}(k+\alpha-1)G_{s,k}(y_k). \end{aligned} \tag{62}$$

The operator $G_{s,k}$ satisfies

$$\Theta_k(y - sG_{s,k}(y)) \le \Theta_k(x) + \langle G_{s,k}(y), y - x\rangle - \frac{s}{2}\|G_{s,k}(y)\|^2 \tag{63}$$

for all $x, y \in \mathcal{H}$ (see [14,19,31,35]), since $s \le \frac{1}{L}$, and $\nabla\Phi_k$ is L-lipschitz continuous. Let us write successively this formula at $y = y_k$ and $x = x_k$, then at $y = y_k$ and $x = x^*$. We obtain

$$\Theta_k(y_k - sG_{s,k}(y_k)) \le \Theta_k(x_k) + \langle G_{s,k}(y_k), y_k - x_k\rangle - \frac{s}{2}\|G_{s,k}(y_k)\|^2, \quad \text{and}$$

$$\Theta_k(y_k - sG_{s,k}(y_k)) \le \Theta_k(x^*) + \langle G_{s,k}(y_k), y_k - x^*\rangle - \frac{s}{2}\|G_{s,k}(y_k)\|^2,$$

respectively. Multiplying the first inequality by $\frac{k}{k+\alpha-1}$, and the second one by $\frac{\alpha-1}{k+\alpha-1}$, then adding the two resulting inequalities, and using the fact that $x_{k+1} = y_k - sG_{s,k}(y_k)$, we obtain

$$\begin{aligned} \Theta_k(x_{k+1}) \le{}& \frac{k}{k+\alpha-1}\Theta_k(x_k) + \frac{\alpha-1}{k+\alpha-1}\Theta_k(x^*) - \frac{s}{2}\|G_{s,k}(y_k)\|^2 \\ &+ \left\langle G_{s,k}(y_k), \frac{k}{k+\alpha-1}(y_k - x_k) + \frac{\alpha-1}{k+\alpha-1}(y_k - x^*)\right\rangle. \end{aligned}$$

We rewrite the scalar product above as

$$
\begin{aligned}
&\left\langle G_{s,k}(y_k), \frac{k}{k+\alpha-1}(y_k - x_k) + \frac{\alpha-1}{k+\alpha-1}(y_k - x^*) \right\rangle \\
&= \frac{\alpha-1}{k+\alpha-1}\left\langle G_{s,k}(y_k), \frac{k}{\alpha-1}(y_k - x_k) + y_k - x^* \right\rangle \\
&= \frac{\alpha-1}{k+\alpha-1}\left\langle G_{s,k}(y_k), \frac{k+\alpha-1}{\alpha-1}y_k - \frac{k}{\alpha-1}x_k - x^* \right\rangle \\
&= \frac{\alpha-1}{k+\alpha-1}\left\langle G_{s,k}(y_k), z_k - x^* \right\rangle.
\end{aligned}
$$

We obtain

$$
\begin{aligned}
\Theta_k(x_{k+1}) \le{} &\frac{k}{k+\alpha-1}\Theta_k(x_k) + \frac{\alpha-1}{k+\alpha-1}\Theta_k(x^*) \\
&+ \frac{\alpha-1}{k+\alpha-1}\left\langle G_{s,k}(y_k), z_k - x^* \right\rangle - \frac{s}{2}\|G_{s,k}(y_k)\|^2.
\end{aligned} \tag{64}
$$

We shall obtain a recursion from (64). To this end, observe that (62) gives

$$
z_{k+1} - x^* = z_k - x^* - \frac{s}{\alpha-1}(k+\alpha-1)\, G_{s,k}(y_k).
$$

After developing

$$
\begin{aligned}
\|z_{k+1} - x^*\|^2 ={} &\|z_k - x^*\|^2 - 2\frac{s}{\alpha-1}(k+\alpha-1)\left\langle z_k - x^*, G_{s,k}(y_k) \right\rangle \\
&+ \frac{s^2}{(\alpha-1)^2}(k+\alpha-1)^2 \|G_{s,k}(y_k)\|^2,
\end{aligned}
$$

and multiplying the above expression by $\frac{(\alpha-1)^2}{2s(k+\alpha-1)^2}$, we obtain

$$
\begin{aligned}
&\frac{(\alpha-1)^2}{2s\,(k+\alpha-1)^2}\left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right) \\
&= \frac{\alpha-1}{k+\alpha-1}\left\langle G_{s,k}(y_k), z_k - x^* \right\rangle - \frac{s}{2}\|G_{s,k}(y_k)\|^2.
\end{aligned}
$$

Replacing this expression in (64), we obtain

$$
\begin{aligned}
\Theta_k(x_{k+1}) \le{} &\frac{k}{k+\alpha-1}\Theta_k(x_k) + \frac{\alpha-1}{k+\alpha-1}\Theta_k(x^*) \\
&+ \frac{(\alpha-1)^2}{2s\,(k+\alpha-1)^2}\left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right).
\end{aligned}
$$

Equivalently,

$$\Theta_k(x_{k+1}) - \Theta_k(x^*) \le \frac{k}{k + \alpha - 1} \left(\Theta_k(x_k) - \Theta_k(x^*)\right)$$
$$+ \frac{(\alpha - 1)^2}{2s\,(k + \alpha - 1)^2} \left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right).$$

Recalling that $\Theta(y) = \Theta_k(y) + \langle g_k, y \rangle$, we obtain

$$\Theta(x_{k+1}) - \Theta(x^*) \le \frac{k}{k + \alpha - 1} \left(\Theta(x_k) - \Theta(x^*)\right)$$
$$+ \frac{(\alpha - 1)^2}{2s\,(k + \alpha - 1)^2} \left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right)$$
$$+ \langle g_k, x_{k+1} - x^* \rangle - \frac{k}{k + \alpha - 1} \langle g_k, x_k - x^* \rangle$$
$$= \frac{k}{k + \alpha - 1} \left(\Theta(x_k) - \Theta(x^*)\right)$$
$$+ \frac{(\alpha - 1)^2}{2s\,(k + \alpha - 1)^2} \left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right)$$
$$+ \left\langle g_k, x_{k+1} - x_k + \frac{\alpha - 1}{k + \alpha - 1}(x_k - x^*) \right\rangle.$$

Multiplying by $\frac{2s}{\alpha - 1}(k + \alpha - 1)^2$, we obtain

$$\frac{2s}{\alpha - 1}(k + \alpha - 1)^2 \left(\Theta(x_{k+1}) - \Theta(x^*)\right)$$
$$\le \frac{2s}{\alpha - 1} k\,(k + \alpha - 1) \left(\Theta(x_k) - \Theta(x^*)\right)$$
$$+ (\alpha - 1) \left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right)$$
$$+ \frac{2s}{\alpha - 1}(k + \alpha - 1)^2 \left\langle g_k, x_{k+1} - x_k + \frac{\alpha - 1}{k + \alpha - 1}(x_k - x^*) \right\rangle,$$

which implies

$$\frac{2s}{\alpha - 1}(k + \alpha - 1)^2 \left(\Theta(x_{k+1}) - \Theta(x^*)\right)$$
$$+ 2s\frac{\alpha - 3}{\alpha - 1} k \left(\Theta(x_k) - \Theta(x^*)\right) \le \frac{2s}{\alpha - 1}(k + \alpha - 2)^2 \left(\Theta(x_k) - \Theta(x^*)\right)$$
$$+ (\alpha - 1) \left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right)$$
$$+ \frac{2s}{\alpha - 1}(k + \alpha - 1)^2 \left\langle g_k, x_{k+1} - x_k + \frac{\alpha - 1}{k + \alpha - 1}(x_k - x^*) \right\rangle, \qquad (65)$$

in view of

$$k\,(k + \alpha - 1) = (k + \alpha - 2)^2 - k(\alpha - 3) - (\alpha - 2)^2$$
$$\leq (k + \alpha - 2)^2 - k(\alpha - 3).$$

Setting

$$\mathcal{G}(k) = \frac{2s}{\alpha - 1}\,(k + \alpha - 2)^2\,(\Theta(x_k) - \Theta^*) + (\alpha - 1)\|z_k - x^*\|^2,$$

we can reformulate (65) as

$$\mathcal{G}(k+1) + 2s\frac{\alpha - 3}{\alpha - 1}k\,\big(\Theta(x_k) - \Theta(x^*)\big)$$
$$\leq \mathcal{G}(k) + \frac{2s}{\alpha - 1}\,(k + \alpha - 1)^2\left\langle g_k, x_{k+1} - x_k + \frac{\alpha - 1}{k + \alpha - 1}(x_k - x^*)\right\rangle.$$

Equivalently,

$$\mathcal{G}(k+1) + 2s\frac{\alpha - 3}{\alpha - 1}k\,\big(\Theta(x_k) - \Theta(x^*)\big)$$
$$\leq \mathcal{G}(k) + 2s\,(k + \alpha - 1)\left\langle g_k, \frac{k + \alpha - 1}{\alpha - 1}\,(x_{k+1} - x_k) + x_k - x^*\right\rangle.$$

Using (61), we deduce that

$$\mathcal{G}(k+1) + 2s\frac{\alpha - 3}{\alpha - 1}k\,\big(\Theta(x_k) - \Theta(x^*)\big)$$
$$\leq \mathcal{G}(k) + 2s\,(k + \alpha - 1)\,\big\langle g_k, z_{k+1} - x^*\big\rangle. \tag{66}$$

Fix an integer K, and set

$$\mathcal{E}_K(k) = \mathcal{G}(k) + \sum_{j=k}^{K} 2s\,(j + \alpha - 1)\big\langle g_j, z_{j+1} - x^*\big\rangle,$$

so that (66) is equivalent to

$$\mathcal{E}_K(k+1) + 2s\frac{\alpha - 3}{\alpha - 1}k\,\big(\Theta(x_k) - \Theta(x^*)\big) \leq \mathcal{E}_K(k),$$

and we deduce that the sequence $(\mathcal{E}_K(k))_k$ is nonincreasing. In particular, $\mathcal{E}_K(k) \leq \mathcal{E}_K(0)$, which gives

$$\mathcal{G}(k) + \sum_{j=k}^{K} 2s\,(j + \alpha - 1)\,\langle g_j, z_{j+1} - x^* \rangle$$

$$\leq \mathcal{G}(0) + \sum_{j=0}^{K} 2s\,(j + \alpha - 1)\,\langle g_j, z_{j+1} - x^* \rangle.$$

As a consequence,

$$\mathcal{G}(k) \leq \mathcal{G}(0) + \sum_{j=0}^{k-1} 2s\,(j + \alpha - 1)\,\langle g_j, z_{j+1} - x^* \rangle. \tag{67}$$

By the definition of $\mathcal{G}(k)$, neglecting some positive terms, and using the Cauchy–Schwarz inequality, we infer that

$$\|z_k - x^*\|^2 \leq \frac{1}{\alpha - 1}\mathcal{G}(0) + \frac{2s}{\alpha - 1}\sum_{j=1}^{k}(j + \alpha - 2)\,\|g_{j-1}\|\,\|z_j - x^*\|.$$

Applying Lemma 5.14 with $a_k = \|z_k - x^*\|$, we deduce that

$$\|z_k - x^*\| \leq M := \sqrt{\frac{\mathcal{G}(0)}{\alpha - 1}} + \frac{2s}{\alpha - 1}\sum_{j=0}^{\infty}(j + \alpha - 1)\,\|g_j\|. \tag{68}$$

Note that M is finite, because $\sum_{k \in \mathbb{N}} k\|g_k\| < +\infty$. Returning to (67) we obtain

$$\mathcal{G}(k) \leq C := \mathcal{G}(0) + 2s \left(\sum_{j=0}^{\infty}(j + \alpha - 1)\,\|g_j\| \right)$$

$$\times \left(\sqrt{\frac{\mathcal{G}(0)}{\alpha - 1}} + \frac{2s}{\alpha - 1}\sum_{j=0}^{\infty}(j + \alpha - 1)\,\|g_j\| \right).$$

By the definition of $\mathcal{G}(k)$, we finally obtain

$$\frac{2s}{\alpha - 1}\,(k + \alpha - 2)^2\,(\Theta(x_k) - \Theta^*) \leq C.$$

which gives (57) and completes the proof.   □

Remark 5.2 In [22], Kim–Fessler introduce an extra inertial term in the FISTA method that allows them to reduce the constant by a factor of 2 in the complexity estimation. It would be interesting to know whether this variant can be obtained by another discretization in time of our inertial dynamic, or a different one.

5.3 Convergence of the sequence (x_k)

Let us now study the convergence of the sequence (x_k).

Theorem 5.3 *Let $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, lower-semicontinuous and convex, and let $\Phi : \mathcal{H} \to \mathbb{R}$ be convex and continuously differentiable with L-Lipschitz continuous gradient. Suppose that $S = \operatorname{argmin}(\Psi + \Phi) \neq \emptyset$, and let (x_k) be a sequence generated by algorithm (55) with $\alpha > 3$, $0 < s < \frac{1}{L}$, and $\sum_{k \in \mathbb{N}} k \|g_k\| < +\infty$. Then,*

(i) $\sum_k k \Big((\Psi + \Phi)(x_k) - \inf(\Psi + \Phi) \Big) < +\infty$;
(ii) $\sum k \|x_{k+1} - x_k\|^2 < +\infty$; and
(iii) x_k converges weakly, as $k \to +\infty$, to some $x^* \in \operatorname{argmin}(\Psi + \Phi)$.

Proof We follow the same steps as those of Theorem 4.4:
    Step 1 Recall from (66) that

$$\mathcal{G}(k+1) + 2s\frac{\alpha - 3}{\alpha - 1}k \left(\Theta(x_k) - \Theta(x^*) \right) \leq \mathcal{G}(k) + 2s \, (k + \alpha - 1) \langle g_k, z_{k+1} - x^* \rangle.$$

By (68), the sequence (z_k) is bounded. Summing the above inequalities, and using $\alpha > 3$, we obtain item i).
    Step 2 Rewrite inequality (63) as

$$\Theta_k(y - sG_{s,k}(y)) + \frac{1}{2s}\|y - sG_{s,k}(y) - x\|^2 \leq \Theta_k(x) + \frac{1}{2s}\|x - y\|^2.$$

Take $y = y_k$, and $x = x_k$. Since $x_{k+1} = y_k - sG_{s,k}(y_k)$, and $y_k - x_k = \frac{k-1}{k+\alpha-1}(x_k - x_{k-1})$, we obtain

$$\Theta_k(x_{k+1}) + \frac{1}{2s}\|x_{k+1} - x_k\|^2 \leq \Theta_k(x_k) + \frac{1}{2s}\frac{(k-1)^2}{(k+\alpha-1)^2}\|x_k - x_{k-1}\|^2.$$

By the definition of Θ_k, this is

$$\Theta(x_{k+1}) + \frac{1}{2s}\|x_{k+1} - x_k\|^2 \leq \Theta(x_k)$$

$$+ \frac{1}{2s}\frac{(k-1)^2}{(k+\alpha-1)^2}\|x_k - x_{k-1}\|^2 + \langle g_k, x_{k+1} - x_k \rangle. \tag{69}$$

Set $\theta_k = \Theta(x_k) - \Theta(x^*)$, $d_k = \frac{1}{2}\|x_k - x_{k-1}\|^2$, $a = \alpha - 1$. By the Cauchy–Schwarz inequality, (69) gives

$$\frac{1}{s}\left(d_{k+1} - \frac{(k-1)^2}{(k+a)^2}d_k \right) \leq (\theta_k - \theta_{k+1}) + \|g_k\|\|x_{k+1} - x_k\|.$$

Multiply by $(k + a)^2$ to get

$$\frac{1}{s} \left((k + a)^2 d_{k+1} - (k - 1)^2 d_k \right) \le (k + a)^2 \left(\theta_k - \theta_{k+1} \right)$$
$$+ (k + a)^2 \|g_k\| \|x_{k+1} - x_k\|.$$

Summing for $k = 1, \ldots, K$, we obtain

$$\sum_{k=1}^{K} \left((k + a)^2 d_{k+1} - (k - 1)^2 d_k \right) \le s \sum_{k=1}^{K} (k + a)^2 \left(\theta_k - \theta_{k+1} \right)$$
$$+ s \sum_{k=1}^{K} (k + a)^2 \|g_k\| \|x_{k+1} - x_k\|.$$

Performing a similar computation as in Chambolle–Dossal [19, Corollary 2], we can write

$$(K + a)^2 d_{K+1} + \sum_{k=2}^{K} a \left(2k + a - 2 \right) d_k$$
$$\le s \left((a + 1)^2 \theta_1 - (K + a)^2 \theta_{K+1} \right.$$
$$\left. + \sum_{k=2}^{K} (2k + 2a - 1) \theta_k + \sum_{k=1}^{K} (k + a)^2 \|g_k\| \|x_{k+1} - x_k\| \right). \tag{70}$$

By item (i), we have $\sum_k (2k + 2a - 1) \theta_k < +\infty$. Hence there exists some constant C such that

$$(K + a)^2 \|x_{K+1} - x_K\|^2 \le C + 2s \sum_{k=1}^{K} (k + a)^2 \|g_k\| \|x_{k+1} - x_k\| \tag{71}$$

for all $K \in \mathbb{N}$. We now proceed as in the proof of Theorem 4.4. To this end, write (71) as

$$a_k^2 \le C + 2s \sum_{j=1}^{k} (j + a) \|g_j\| a_j,$$

where $a_j := (j + a) \|x_{j+1} - x_j\|$. Recalling that $\sum_k k \|g_k\| < +\infty$, apply Lemma 5.14 with $\beta_j = (j + a) \|g_j\|$ to deduce that

$$\sup_k k \|x_{k+1} - x_k\| < +\infty. \tag{72}$$

Injecting this information in (70), we obtain

$$\sum_k a\,(2k + a - 2)\,d_k \le C + \sum_k (2k + 2a - 1)\,\theta_k$$

$$+ \sup_k((k + a)\|x_{k+1} - x_k\|) \sum_k (k + a)\|g_k\|.$$

Since $a = \alpha - 1 \ge 2$, item (i) and the definition of d_k, together give

$$\sum_k k\|x_{k+1} - x_k\|^2 < +\infty,$$

which is $ii)$.

Step 3 We finish by applying Opial's Lemma 5.8 with $S = \mathrm{argmin}(\Psi + \Phi)$. By Theorem 5.1, we have $(\Psi + \Phi)(x_k) \to \min(\Psi + \Phi)$. The weak lower-semicontinuity of $\Psi + \Phi$ gives item (ii) of Opial's Lemma. Thus, the only point to verify is that $\lim \|x_k - x^*\|$ exists for each $x^* \in \mathrm{argmin}(\Psi + \Phi)$. Take any such x^*. We shall show that $\lim_{k \to \infty} h_k$ exists, where $h_k := \frac{1}{2}\|x_k - x^*\|^2$.

The beginning of the proof is similar to [3] or [19], and consists in establishing a discrete version of the second-order differential inequality (46). We use the identity

$$\frac{1}{2}\|a - b\|^2 + \frac{1}{2}\|a - c\|^2 = \frac{1}{2}\|b - c\|^2 + \langle a - b, a - c \rangle,$$

which holds for any $a, b, c \in \mathcal{H}$. Taking $b = x^*, a = x_{k+1}, c = x_k$, we obtain

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

which is equivalent to

$$h_k - h_{k+1} = \frac{1}{2}\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x^*, x_k - x_{k+1} \rangle. \tag{73}$$

By the definition of y_k, we have

$$x_k - x_{k+1} = y_k - x_{k+1} - \frac{k - 1}{k + \alpha - 1}(x_k - x_{k-1}).$$

Therefore,

$$h_k - h_{k+1} = \frac{1}{2}\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x^*, y_k - x_{k+1} \rangle$$

$$- \frac{k - 1}{k + \alpha - 1}\langle x_{k+1} - x^*, x_k - x_{k-1} \rangle. \tag{74}$$

We now use the monotonicity of $\partial \Psi$. Since $-s \nabla \Phi(x^*) \in s \partial \Psi(x^*)$, and $y_k - x_{k+1} - s \nabla \Phi(y_k) + s g_k \in s \partial \Psi(x_{k+1})$, we have

$$\left\langle y_k - x_{k+1} - s \nabla \Phi(y_k) + s g_k + s \nabla \Phi(x^*), x_{k+1} - x^* \right\rangle \geq 0.$$

Equivalently,

$$\left\langle y_k - x_{k+1}, x_{k+1} - x^* \right\rangle + s \left\langle \nabla \Phi(x^*) - \nabla \Phi(y_k) + g_k, x_{k+1} - x^* \right\rangle \geq 0.$$

Replacing in (74), we obtain

$$h_{k+1} - h_k + \frac{1}{2} \|x_{k+1} - x_k\|^2 + s \left\langle \nabla \Phi(y_k) - \nabla \Phi(x^*) - g_k, x_{k+1} - x^* \right\rangle$$
$$- \frac{k-1}{k+\alpha-1} \left\langle x_{k+1} - x^*, x_k - x_{k-1} \right\rangle \leq 0. \tag{75}$$

On the other hand, the co-coercivity of $\nabla \Phi$ gives

$$\begin{aligned}
\left\langle \nabla \Phi(y_k) - \nabla \Phi(x^*), x_{k+1} - x^* \right\rangle &= \left\langle \nabla \Phi(y_k) - \nabla \Phi(x^*), x_{k+1} - y_k \right\rangle \\
&\quad + \left\langle \nabla \Phi(y_k) - \nabla \Phi(x^*), y_k - x^* \right\rangle \\
&\geq \frac{1}{L} \|\Phi(y_k) - \nabla \Phi(x^*)\|^2 \\
&\quad + \left\langle \nabla \Phi(y_k) - \nabla \Phi(x^*), x_{k+1} - y_k \right\rangle \\
&\geq \frac{1}{L} \|\Phi(y_k) - \nabla \Phi(x^*)\|^2 - \|\nabla \Phi(y_k) \\
&\quad - \nabla \Phi(x^*)\| \|x_{k+1} - y_k\| \\
&\geq -\frac{L}{2} \|x_{k+1} - y_k\|^2 \tag{76}
\end{aligned}$$

(vertex of the parabola). Combining (75) and (76), we deduce that

$$h_{k+1} - h_k + \frac{1}{2} \|x_{k+1} - x_k\|^2 - \frac{sL}{2} \|x_{k+1}$$
$$- y_k\|^2 - s \|g_k\| \|x_{k+1} - x^*\|$$
$$- \frac{k-1}{k+\alpha-1} \left\langle x_{k+1} - x^*, x_k - x_{k-1} \right\rangle \leq 0. \tag{77}$$

Replace k by $k-1$ in (73) to obtain

$$h_{k-1} - h_k = \frac{1}{2} \|x_k - x_{k-1}\|^2 - \left\langle x_k - x^*, x_k - x_{k-1} \right\rangle. \tag{78}$$

Combine (77) with (78) to deduce that

$$
\begin{aligned}
h_{k+1} - h_k - \frac{k-1}{k+\alpha-1} (h_k - h_{k-1}) \\
\leq -\frac{1}{2} \|x_{k+1} - x_k\|^2 + \frac{sL}{2} \|x_{k+1} - y_k\|^2 + s\|g_k\|\|x_{k+1} - x^*\| \\
+ \frac{k-1}{k+\alpha-1} \left(\frac{1}{2}\|x_k - x_{k-1}\|^2 + \langle x_k - x_{k-1}, x_{k+1} - x_k \rangle \right).
\end{aligned} \tag{79}
$$

By the definition of $y_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1})$, we have $x_{k+1} - y_k = x_{k+1} - x_k - \frac{k-1}{k+\alpha-1}(x_k - x_{k-1})$. Hence,

$$
\begin{aligned}
\|x_{k+1} - y_k\|^2 = \|x_{k+1} - x_k\|^2 + \left(\frac{k-1}{k+\alpha-1} \right)^2 \|x_k - x_{k-1}\|^2 \\
- 2\frac{k-1}{k+\alpha-1} \langle x_{k+1} - x_k, x_k - x_{k-1} \rangle
\end{aligned}
$$

Substituting this in (79), we obtain

$$
\begin{aligned}
h_{k+1} - h_k - \gamma_k (h_k - h_{k-1}) \leq -\left(1 - \frac{sL}{2} \right) \|x_{k+1} - y_k\|^2 + s\|g_k\|\|x_{k+1} \\
- x^*\| + \left(\gamma_k + \gamma_k{}^2 \right) \|x_k - x_{k-1}\|^2,
\end{aligned}
$$

where $\gamma_k = \frac{k-1}{k+\alpha-1}$. Since $0 < s < \frac{1}{L}$, we have $(1 - \frac{sL}{2}) > 0$. On the other hand, since $\gamma_k < 1$, we have $\gamma_k + \gamma_k{}^2 < 2\gamma_k$. Therefore,

$$
h_{k+1} - h_k - \gamma_k (h_k - h_{k-1}) \leq s\|g_k\|\|x_{k+1} - x^*\| + 2\gamma_k\|x_k - x_{k-1}\|^2. \tag{80}
$$

By (68), we know that the sequence (z_k) is bounded. By (72), we know that $\sup_k k\|x_{k+1} - x_k\| < +\infty$. Since $x_k = z_k - \frac{k+\alpha-1}{\alpha-1}(x_{k+1} - x_k)$, we deduce that the sequence (x_k) is bounded. Returning to (80), we have

$$
h_{k+1} - h_k - \gamma_k (h_k - h_{k-1}) \leq C\|g_k\| + 2\gamma_k\|x_k - x_{k-1}\|^2
$$

for some constant C. Now, item (ii), combined with the assumption $\sum_k k\|g_k\| < +\infty$, together give

$$
h_{k+1} - h_k - \gamma_k (h_k - h_{k-1}) \leq \omega_k,
$$

for some nonnegative sequence (ω_k) such that $\sum_{k\in\mathbb{N}} k\omega_k < +\infty$. Taking the positive parts and applying Lemma 5.10 with $a_k = (h_k - h_{k-1})^+$, to obtain

$$
\sum_k (h_k - h_{k-1})^+ < +\infty.
$$

Since (h_k) is nonnegative, one easily sees that it must converge. This completes the proof.   □

Remark 5.4 One reasonable conjecture is that strong convergence results can be obtained for algorithm (54), by transposing the results of Sect. 3 from the continuous to the discrete case. This direction will not be explored here, though.

Remark 5.5 The analysis carried out in this section for inertial forward-backward algorithm (55) is a reinterpretation of the proof of the corresponding results in the continuous case. In other words, we built a complete proof having the continuous setting as a guideline. It would be interesting to know whether the results in [5,6] can be applied in order to deduce the asymptotic properties without repeating the proofs.

5.4 Final comments

In the particular case $\alpha = 3$, for a perturbed version of the classical FISTA algorithm, Schmidt–Le Roux–Bach proved in [34] a result which is similar to Theorem 5.1, concerning the fast convergence of the values. In [36], Villa–Salzo–Baldassarres–Verri use the notion of ϵ-subdifferential in order to compute inexact proximal points in the algorithm. In a recent article [10], Aujol–Dossal extend this study by introducing additional perturbation terms to analyze the stability of the FISTA method, and prove convergence of the sequences generated by the algorithm (in the spirit of Chambolle–Dossal [19]). Although the results obtained in the previous papers have some similarities to those of Theorems 5.1 and 5.3, our dynamic approach is original and opens the door to new developments. One example is the following: In the dynamical system studied here, second-order information with respect to time ultimately induces fast convergence properties. On the other hand, in Newton-type methods, second-order information with respect to space has a similar consequence. In a forthcoming paper, and based on previous works [4] and [9], we study the solutions of the second-order (both in time and space) evolution equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \nabla^2 \Phi(x(t))\,\dot{x}(t) + \nabla\Phi(x(t)) = 0,$$

where Φ is a smooth convex function, and α, β are positive parameters. This inertial system combines an isotropic viscous damping which vanishes asymptotically, and a geometrical Hessian-driven damping, which makes it naturally related to Newton and Levenberg-Marquardt methods.

The rich literature on the optimal damping of oscillating systems offers interesting insight (see, for example, the recent paper by Ghisi–Gobbino–Haraux [20], which deals with periodic damping).

Appendix: Some auxiliary results

In this section, we present some auxiliary lemmas to be used later on. The following result can be found in [1]:

Lemma 5.6 *Let $\delta > 0$, $1 \leq p < \infty$ and $1 \leq r \leq \infty$. Suppose $F \in L^p([\delta, \infty[)$ is a locally absolutely continuous nonnegative function, $G \in L^r([\delta, \infty[)$ and*

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

for almost every $t > \delta$. Then $\lim_{t \to \infty} F(t) = 0$.

To establish the weak convergence of the solutions of (1), we will use Opial's Lemma [30], that we recall in its continuous form. This argument was first used in [16] to establish the convergence of nonlinear contraction semigroups.

Lemma 5.7 *Let S be a nonempty subset of $\mathcal{H}$ and let $x : [0, +\infty) \to \mathcal{H}$. Assume that*

(i) for every $z \in S$, $\lim_{t \to \infty} \|x(t) - z\|$ exists;
(ii) every weak sequential limit point of $x(t)$, as $t \to \infty$, belongs to S.

Then $x(t)$ converges weakly as $t \to \infty$ to a point in S.

Its discrete version is

Lemma 5.8 *Let S be a non empty subset of $\mathcal{H}$, and (x_k) a sequence of elements of $\mathcal{H}$. Assume that*

(i) for every $z \in S$, $\lim_{k \to +\infty} \|x_k - z\|$ exists;
(ii) every weak sequential limit point of (x_k), as $k \to \infty$, belongs to S.

Then x_k converges weakly as $k \to \infty$ to a point in S.

The following allows us to establish the existence of a limit for a real-valued function, as $t \to +\infty$:

Lemma 5.9 *Let $\delta > 0$, and let $w : [\delta, +\infty[\to \mathbb{R}$ be a continuously differentiable function which is bounded from below. Assume*

$$t\ddot{w}(t) + \alpha\dot{w}(t) \leq g(t), \tag{81}$$

for some $\alpha > 1$, almost every $t > \delta$, and some nonnegative function $g \in L^1(\delta, +\infty)$. Then, the positive part $[\dot{w}]_+$ of $\dot{w}$ belongs to $L^1(t_0, +\infty)$ and $\lim_{t \to +\infty} w(t)$ exists.

Proof Multiply (81) by $t^{\alpha-1}$ to obtain

$$\frac{d}{dt}\left(t^\alpha \dot{w}(t)\right) \leq t^{\alpha-1}g(t).$$

By integration, we obtain

$$\dot{w}(t) \leq \frac{\delta^\alpha |\dot{w}(\delta)|}{t^\alpha} + \frac{1}{t^\alpha}\int_\delta^t s^{\alpha-1}g(s)ds.$$

Hence,

$$[\dot{w}]_+(t) \le \frac{\delta^\alpha |\dot{w}(\delta)|}{t^\alpha} + \frac{1}{t^\alpha} \int_\delta^t s^{\alpha-1} g(s) ds,$$

and so,

$$\int_\delta^\infty [\dot{w}]_+(t) dt \le \frac{\delta^\alpha |\dot{w}(\delta)|}{(\alpha-1)\delta^{\alpha-1}} + \int_\delta^\infty \frac{1}{t^\alpha} \left(\int_\delta^t s^{\alpha-1} g(s) ds \right) dt.$$

Applying Fubini's Theorem, we deduce that

$$\int_\delta^\infty \frac{1}{t^\alpha} \left(\int_\delta^t s^{\alpha-1} g(s) ds \right) dt = \int_\delta^\infty \left(\int_s^\infty \frac{1}{t^\alpha} dt \right) s^{\alpha-1} g(s) ds$$

$$= \frac{1}{\alpha-1} \int_\delta^\infty g(s) ds.$$

As a consequence,

$$\int_\delta^\infty [\dot{w}]_+(t) dt \le \frac{\delta^\alpha |\dot{w}(\delta)|}{(\alpha-1)\delta^{\alpha-1}} + \frac{1}{\alpha-1} \int_\delta^\infty g(s) ds < +\infty.$$

Finally, the function $\theta : [\delta, +\infty) \to \mathbb{R}$, defined by

$$\theta(t) = w(t) - \int_\delta^t [\dot{w}]_+(\tau) d\tau,$$

is nonincreasing and bounded from below. It follows that

$$\lim_{t \to +\infty} w(t) = \lim_{t \to +\infty} \theta(t) + \int_\delta^{+\infty} [\dot{w}]_+(\tau) d\tau$$

exists.   □

In the study of the corresponding algorithms, we use the following result, which is a discrete version of Lemma 5.9:

Lemma 5.10 *Let $\alpha \ge 3$, and let (a_k) and (ω_k) be two sequences of nonnegative numbers such that*

$$a_{k+1} \le \frac{k-1}{k+\alpha-1} a_k + \omega_k$$

for all $k \ge 1$. If $\sum_k k\omega_{k \in \mathbb{N}} < +\infty$, then $\sum_{k \in \mathbb{N}} a_k < +\infty$.

Proof Since $\alpha \geq 3$ we have $\alpha - 1 \geq 2$, and hence

$$a_{k+1} \leq \frac{k-1}{k+2}a_k + \omega_k.$$

Multiplying this expression by $(k+1)^2$, we obtain

$$(k+1)^2 a_{k+1} \leq \frac{(k-1)(k+1)^2}{k+2}a_k + (k+1)^2\omega_k \leq k^2 a_k + (k+1)^2\omega_k.$$

Summing this inequality for $j = 1, 2, \ldots, k$, we obtain

$$k^2 a_k \leq a_1 + \sum_{j=1}^{k-1}(j+1)^2\omega_j.$$

Dividing by k^2, and summing for $k = 2, \ldots, K$, we obtain

$$\sum_{k=2}^{K} a_k \leq a_1 \sum_{k=2}^{K}\frac{1}{k^2} + \sum_{k=2}^{K}\frac{1}{k^2}\sum_{j=1}^{k-1}(j+1)^2\omega_j.$$

Applying Fubini's Theorem to this last sum, and observing that $\sum_{k=j+1}^{\infty}\frac{1}{k^2} \leq \int_j^{\infty}\frac{1}{t^2}dt = \frac{1}{j}$, we obtain

$$\sum_{k=2}^{K} a_k \leq a_1 \sum_{k=2}^{K}\frac{1}{k^2} + \sum_{j=1}^{K-1}\left(\sum_{k=j+1}^{\infty}\frac{1}{k^2}\right)(j+1)^2\omega_j$$

$$\leq a_1 \sum_{k=2}^{K}\frac{1}{k^2} + \sum_{j=1}^{K-1}\frac{(j+1)^2}{j}\omega_j$$

$$\leq a_1 \sum_{k=2}^{\infty}\frac{1}{k^2} + 4\sum_{j=1}^{\infty} j\omega_j < +\infty,$$

and the result follows. □

The following is a continuous version of Kronecker's Theorem for series (see, for example, [23, page 129]):

Lemma 5.11 *Take $\delta > 0$, and let $f \in L^1(\delta, +\infty)$ be nonnegative and continuous. Consider a nondecreasing function $\psi : (\delta, +\infty) \to (0, +\infty)$ such that $\lim_{t \to +\infty} \psi(t) = +\infty$. Then,*

$$\lim_{t \to +\infty} \frac{1}{\psi(t)}\int_{\delta}^{t} \psi(s)f(s)ds = 0.$$

Proof Given $\epsilon > 0$, fix t_ϵ sufficiently large so that

$$\int_{t_\epsilon}^\infty f(s)ds \le \epsilon.$$

Then, for $t \ge t_\epsilon$, split the integral $\int_\delta^t \psi(s)f(s)ds$ into two parts to obtain

$$\frac{1}{\psi(t)} \int_\delta^t \psi(s)f(s)ds = \frac{1}{\psi(t)} \int_\delta^{t_\epsilon} \psi(s)f(s)ds + \frac{1}{\psi(t)} \int_{t_\epsilon}^t \psi(s)f(s)ds$$

$$\le \frac{1}{\psi(t)} \int_\delta^{t_\epsilon} \psi(s)f(s)ds + \int_{t_\epsilon}^t f(s)ds.$$

Now let $t \to +\infty$ to deduce that

$$0 \le \limsup_{t \to +\infty} \frac{1}{\psi(t)} \int_\delta^t \psi(s)f(s)ds \le \epsilon.$$

Since this is true for any $\epsilon > 0$, the result follows. $\qquad\qquad\qquad\square$

Using the previous result, we also obtain the following vector-valued version of Lemma 5.9:

Lemma 5.12 *Take $\delta > 0$, and let $F \in L^1(\delta, +\infty; \mathcal{H})$ be continuous. Let $x : [\delta, +\infty[\to \mathcal{H}$ be a solution of*

$$t\ddot{x}(t) + \alpha\dot{x}(t) = F(t) \tag{82}$$

with $\alpha > 1$. Then, $x(t)$ converges strongly in $\mathcal{H}$ as $t \to +\infty$.

Proof As in the proof of Lemma 5.9, multiply (82) by $t^{\alpha-1}$ and integrate to obtain

$$\dot{x}(t) = \frac{\delta^\alpha \dot{x}(\delta)}{t^\alpha} + \frac{1}{t^\alpha} \int_\delta^t s^{\alpha-1} F(s)ds.$$

Integrate again to deduce that

$$x(t) = x(\delta) + \delta^\alpha \dot{x}(\delta) \int_\delta^t \frac{1}{s^\alpha} ds + \int_\delta^t \frac{1}{s^\alpha} \left(\int_\delta^s \tau^{\alpha-1} F(\tau)d\tau \right) ds.$$

Fubini's Theorem applied to the last integral gives

$$x(t) = x(\delta) + \frac{\delta^\alpha \dot{x}(\delta)}{\alpha - 1} \left(\frac{1}{\delta^{\alpha-1}} - \frac{1}{t^{\alpha-1}} \right)$$

$$+ \frac{1}{\alpha - 1} \left(\int_\delta^t F(\tau)d\tau - \frac{1}{t^{\alpha-1}} \int_\delta^t \tau^{\alpha-1} F(\tau)d\tau \right). \tag{83}$$

Finally, apply Lemma 5.11 to the last integral with $\psi(s) = s^{\alpha-1}$ and $f(s) = \|F(s)\|$ to conclude that all the terms in the right-hand side of (83) have a limit as $t \to +\infty$.
□

Finally, in the analysis of the solutions for the perturbed system, we shall use the following Gronwall-Bellman Lemma (see [15, Lemme A.5]):

Lemma 5.13 *Let $m : [\delta, T] \to [0, +\infty[$ be integrable, and let $c \geq 0$. Suppose $w : [\delta, T] \to \mathbb{R}$ is continuous and*

$$\frac{1}{2}w^2(t) \leq \frac{1}{2}c^2 + \int_\delta^t m(\tau)w(\tau)d\tau$$

for all $t \in [\delta, T]$. Then, $|w(t)| \leq c + \int_\delta^t m(\tau)d\tau$ for all $t \in [\delta, T]$.

We shall also make use of the following discrete version of the preceding result:

Lemma 5.14 *Let (a_k) be a sequence of nonnegative numbers such that*

$$a_k^2 \leq c^2 + \sum_{j=1}^k \beta_j a_j$$

for all $k \in \mathbb{N}$, where (β_j) is a summable sequence of nonnegative numbers, and $c \geq 0$. Then, $a_k \leq c + \sum_{j=1}^\infty \beta_j$ for all $k \in \mathbb{N}$.

Proof For $k \in \mathbb{N}$, set $A_k := \max_{1 \leq m \leq k} a_m$. Then, for $1 \leq m \leq k$, we have

$$a_m^2 \leq c^2 + \sum_{j=1}^m \beta_j a_j \leq c^2 + A_k \sum_{j=1}^\infty \beta_j.$$

Taking the maximum over $1 \leq m \leq k$, we obtain

$$A_k^2 \leq c^2 + A_k \sum_{j=1}^\infty \beta_j.$$

Bounding by the roots of the corresponding quadratic equation, we obtain the result.
□

References

1. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward–backward methods for structured monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **161**(2), 331–360 (2014)
2. Alvarez, F.: On the minimizing property of a second-order dissipative system in Hilbert spaces. SIAM J. Control Optim. **38**(4), 1102–1119 (2000)

3. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set Valued Anal. **9**(1–2), 3–11 (2001)

4. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. J. Math. Pures Appl. **81**(8), 747–779 (2002)

5. Alvarez, F., Peypouquet, J.: Asymptotic almost-equivalence of Lipschitz evolution systems in Banach spaces. Nonlinear Anal. **73**(9), 3018–3033 (2010)

6. Alvarez, F., Peypouquet, J.: A unified approach to the asymptotic almost-equivalence of evolution systems without Lipschitz conditions. Nonlinear Anal. **74**(11), 3440–3444 (2011)

7. Attouch, H., Buttazzo, G., Michaille, G.: Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization, 2nd edn. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, xii+793 pp (2014)

8. Attouch, H., Cabot, A., Redont, P.: The dynamics of elastic shocks via epigraphical regularization of a differential inclusion. Adv. Math. Sci. Appl. **12**(1), 273–306 (2002)

9. Attouch, H., Peypouquet, J., Redont, P.: A dynamical approach to an inertial forward-backward algorithm for convex minimization. SIAM J. Optim. **24**(1), 232–256 (2014)

10. Aujol, J.-F., Dossal, C.: Stability of over-relaxations for the forward–backward algorithm, application to FISTA. SIAM J. Optim. Society for Industrial and Applied Mathematics (2015). hal-01163432

11. Baillon, J.-B.: Un exemple concernant le comportement asymptotique de la solution du problème $\frac{du}{dt} + \partial\phi(u) \ni 0$. J. Funct. Anal. **28**, 369–376 (1978)

12. Baillon, J.-B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. Isr. J. Math. **26**, 137–150 (1977)

13. Bauschke, H., Combettes, P.: Convex Analysis and Monotone Operator Theory in Hilbert spaces. CMS Books in Mathematics, Springer (2011)

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

15. Brézis, H.: Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution, Lecture Notes 5, North Holland (1972)

16. Bruck, R.E.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces. J. Funct. Anal. **18**, 15–26 (1975)

17. Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation. Trans. Am. Math. Soc. **361**, 5983–6017 (2009)

18. Cabot, A., Engler, H., Gadat, S.: Second order differential equations with asymptotically small dissipation and piecewise flat potentials. Electron. J. Differ. Equ. **17**, 33–38 (2009)

19. Chambolle, A., Dossal, Ch.: On the convergence of the iterates of Fista, HAL Id: hal-01060130 https://hal.inria.fr/hal-01060130v3. Submitted on 20 Oct 2014

20. Ghisi, M., Gobbino, M., Haraux, A.: The remarkable effectiveness of time-dependent damping terms for second order evolution equations (2015). arXiv:1506.06915v1 [math.AP]

21. Güler, O.: New proximal point algorithms for convex minimization. SIAM J. Optim. **2**(4), 649–664 (1992)

22. Kim, D., Fessler, J.A.: Optimized first-order methods for smooth convex minimization (2015). arXiv:1406.5468v2 [math.OC]

23. Knopp, K.: Theory and Application of Infinite Series. Blackie & Son, Glasgow (1951)

24. Lorenz, D.A., Pock, T.: An inertial forward–backward algorithm for monotone inclusions. J. Math. Imaging Vis. **51**(2), 1–15 (2015)

25. Moudafi, A., Oliny, M.: Convergence of a splitting inertial proximal method for monotone operators. J. Comput. Appl. Math. **155**(2), 447–454 (2003)

26. Nesterov, Y.: A method of solving a convex programming problem with convergence rate O(1/k2). Sov. Math. Dokl. **27**, 372–376 (1983)

27. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course, Volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston (2004)

28. Nesterov, Y.: Smooth minimization of non-smooth functions. Math. Program. **103**(1), 127–152 (2005)

29. Nesterov, Y.: Gradient methods for minimizing composite objective function. CORE Discussion Papers (2007)

30. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. **73**, 591–597 (1967)

31. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**, 123–231 (2013)
32. Peypouquet, J.: Convex Optimization in Normed Spaces: Theory, Methods and Examples. Springer, Berlin (2015)
33. Peypouquet, J., Sorin, S.: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. J. Convex Anal. **17**(3–4), 1113–1163 (2010)
34. Schmidt, M., Le Roux, N., Bach, F.: Convergence rates of inexact proximal-gradient methods for convex optimization. In: NIPS'11—25th Annual Conference on Neural Information Processing Systems, Grenada, Spain (2011) HAL inria-00618152v3
35. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. Neural Inf. Process. Syst. **27**, 2510–2518 (2014)
36. Villa, S., Salzo, S., Baldassarres, L., Verri, A.: Accelerated and inexact forward–backward. SIAM J. Optim. **23**(3), 1607–1633 (2013)