



**HAL**  
open science

# Fast convex optimization via a third-order in time evolution equation

Hedy Attouch, Zaki Chbani, Hassan Riahi

► **To cite this version:**

Hedy Attouch, Zaki Chbani, Hassan Riahi. Fast convex optimization via a third-order in time evolution equation. 2020. hal-02432351

**HAL Id: hal-02432351**

**<https://hal.science/hal-02432351>**

Preprint submitted on 8 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# FAST CONVEX OPTIMIZATION VIA A THIRD-ORDER IN TIME EVOLUTION EQUATION

HEDY ATTOUCH, ZAKI CHBANI, AND HASSAN RIAHI

ABSTRACT. In a Hilbert space  $\mathcal{H}$ , we develop fast convex optimization methods, which are based on a third order in time evolution system. The function to minimize  $f : \mathcal{H} \rightarrow \mathbb{R}$  is convex, continuously differentiable, with  $\operatorname{argmin} f \neq \emptyset$ , and enters the dynamic via its gradient. On the basis of Lyapunov's analysis and temporal scaling techniques, we show a convergence rate of the values of the order  $1/t^3$ , and obtain the convergence of the trajectories towards optimal solutions. When  $f$  is strongly convex, an exponential rate of convergence is obtained. We complete the study of the continuous dynamic by introducing a damping term induced by the Hessian of  $f$ . This allows the oscillations to be controlled and attenuated. Then, we analyze the convergence of the proximal-based algorithms obtained by temporal discretization of this system, and obtain similar convergence rates. The algorithmic results are valid for a general convex, lower semicontinuous, and proper function  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ .

## 1. INTRODUCTION: INERTIAL DYNAMICS FOR OPTIMIZATION

Throughout the paper,  $\mathcal{H}$  is a real Hilbert space, endowed with the scalar product  $\langle \cdot, \cdot \rangle$  and the associated norm  $\| \cdot \|$ . Unless specified,  $f : \mathcal{H} \rightarrow \mathbb{R}$  is a  $C^1$  convex function with  $\operatorname{argmin} f \neq \emptyset$ . We take  $t_0 > 0$  as the origin of time <sup>1</sup>. We are going to study the convergence properties as  $t \rightarrow +\infty$  of the trajectories of the following evolution system

(TOGES): Third Order Gradient Evolution System

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \frac{2\alpha-6}{t^2}x(t) + \nabla f(x(t) + t\dot{x}(t)) = 0.$$

(TOGES) is a third order in time evolution system, where the function  $f$  to minimize enters via its gradient  $\nabla f$ , and  $\alpha$  is a positive damping parameter. Then, we will study the convergence properties of the proximal-based algorithms obtained by temporal discretization of this system. The system (TOGES) has its roots in recent developments concerning inertial optimization methods and their links with dynamic systems, which we briefly review.

**1.1. From the Heavy Ball with Friction to the Asymptotic Vanishing Damping.** B. Polyak initiated the use of inertial dynamics to accelerate the gradient method in optimization. In [38], based on the inertial system with a fixed viscous damping coefficient  $\gamma > 0$

$$\text{(HBF)} \quad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) = 0,$$

---

*Date:* December 24, 2019.

*1991 Mathematics Subject Classification.* 49M37, 65K05, 90C25.

*Key words and phrases.* Accelerated gradient system; convex optimization; Lyapunov analysis; proximal algorithms; time rescaling; third order in time evolution system.

<sup>1</sup>this is justified by the singularity at the origin of the damping coefficient  $\gamma(t) = \frac{\alpha}{t}$  which is used in the paper.

he introduced the Heavy Ball with Friction method. For a strongly convex function  $f$ , (HBF) provides convergence at exponential rate of  $f(x(t))$  to  $\inf_{\mathcal{H}} f$ . For general convex functions, the asymptotic convergence rate of (HBF) is  $\mathcal{O}(\frac{1}{t})$  (in the worst case). This is however not better than the steepest descent. It turns out that in this model, the friction is important and neutralizes the inertial effect, which prevents obtaining fast optimization methods for general convex functions. Indeed, considering the non autonomous system

$$(IGS)_{\gamma(\cdot)} \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla f(x(t)) = 0,$$

where the damping parameter  $\gamma(t)$  tends to zero as  $t \rightarrow +\infty$  (vanishing damping) is a key property for obtaining fast optimization methods. A decisive step in this direction was obtained by Su-Boyd-Candès [44], who showed that the Nesterov accelerated gradient method can be obtained as a temporal discretization of the dynamic system  $(IGS)_{\gamma(\cdot)}$  with  $\gamma(t) = \frac{3}{t}$ . Recently, Attouch-Chbani-Peypouquet-Redont [9] and May [31] studied the convergence of the trajectories of the  $(IGS)_{\gamma(\cdot)}$  system with  $\gamma(t) = \frac{\alpha}{t}$  and  $\alpha \geq 3$

$$(AVD)_{\alpha} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0.$$

As a specific feature, the viscous damping coefficient  $\frac{\alpha}{t}$  vanishes (tends to zero) as time  $t$  goes to infinity, hence the terminology Asymptotic Vanishing Damping with coefficient  $\alpha$ ,  $(AVD)_{\alpha}$  for short. Let us briefly recall the convergence rates obtained for this system:

- For  $\alpha \geq 3$ , each trajectory  $x(\cdot)$  of  $(AVD)_{\alpha}$  satisfies the asymptotic convergence rate of the values  $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/t^2)$ , see [5], [9], [31], [44].
- The case  $\alpha = 3$ , which corresponds to Nesterov's historical algorithm, is critical. In the case  $\alpha = 3$ , the question of the convergence of the trajectories remains an open problem (except in one dimension where convergence holds [10]).
- For  $\alpha > 3$ , it has been shown in [9] that each trajectory converges weakly to a minimizer. The corresponding algorithmic result has been obtained by Chambolle-Dossal [25]. In addition, it is shown in [15] and [31] that the asymptotic convergence rate of the values is  $o(1/t^2)$ .
- The subcritical case  $\alpha \leq 3$  has been examined by Apidopoulos-Aujol-Dossal [3] and Attouch-Chbani-Riahi [10], with the convergence rate of the objective values  $\mathcal{O}(1/t^{\frac{2\alpha}{3}})$ .

These rates are optimal, that is, they can be reached, or approached arbitrarily close. For further results concerning the system  $(AVD)_{\alpha}$  one can consult [3, 5, 6, 7, 9, 10, 15, 17, 18, 31, 44].

**1.2. Time rescaling.** These results are based on Attouch-Chbani-Riahi [11]. Let us start from  $(AVD)_{\alpha}$ . Suppose that  $\alpha \geq 3$ . Given a trajectory  $x(\cdot)$  of  $(AVD)_{\alpha}$ , we know that (see [5], [9], [44])

$$(1) \quad f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right) \text{ as } t \rightarrow +\infty.$$

Let's make the change of time variable  $t = \tau(s)$  in  $(AVD)_{\alpha}$ , where  $\tau(\cdot)$  is a  $\mathcal{C}^2$  strictly increasing function from  $\mathbb{R}$  to  $\mathbb{R}$ , which satisfies  $\lim_{s \rightarrow +\infty} \tau(s) = +\infty$ . We have

$$(2) \quad \ddot{x}(\tau(s)) + \frac{\alpha}{\tau(s)}\dot{x}(\tau(s)) + \nabla f(x(\tau(s))) = 0.$$

Set  $y(s) := x(\tau(s))$ . By the derivation chain rule, we have

$$\dot{y}(s) = \dot{\tau}(s)\dot{x}(\tau(s)), \quad \ddot{y}(s) = \ddot{\tau}(s)\dot{x}(\tau(s)) + \dot{\tau}(s)^2\ddot{x}(\tau(s)).$$

Reformulating (2) in terms of  $y(\cdot)$  and its derivatives, we obtain

$$\frac{1}{\dot{\tau}(s)^2} \left( \ddot{y}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)}\dot{y}(s) \right) + \frac{\alpha}{\tau(s)} \frac{1}{\dot{\tau}(s)}\dot{y}(s) + \nabla f(y(s)) = 0.$$

Hence,  $y(\cdot)$  is solution of the rescaled equation

$$(3) \quad \ddot{y}(s) + \left( \frac{\alpha}{\tau(s)} \dot{\tau}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)} \right) \dot{y}(s) + \dot{\tau}(s)^2 \nabla f(y(s)) = 0.$$

The formula (1) becomes

$$(4) \quad f(y(s)) - \inf_{\mathcal{H}} f = \mathcal{O} \left( \frac{1}{\tau(s)^2} \right) \text{ as } s \rightarrow +\infty.$$

Hence, by making a fast time reparametrization, we can obtain arbitrarily fast convergence property of the values. The damping coefficient of (3) is equal to

$$\tilde{\gamma}(s) = \frac{\alpha}{\tau(s)} \dot{\tau}(s) - \frac{\ddot{\tau}(s)}{\dot{\tau}(s)} = \frac{\alpha \dot{\tau}(s)^2 - \tau(s) \ddot{\tau}(s)}{\tau(s) \dot{\tau}(s)}.$$

As a model example, take  $\tau(s) = s^p$ , where  $p$  is a positive parameter. Then  $\tilde{\gamma}(s) = \frac{\alpha_p}{s}$ , where  $\alpha_p = 1 + (\alpha - 1)p$ , and (3) writes

$$(5) \quad \ddot{y}(s) + \frac{\alpha_p}{s} \dot{y}(s) + p^2 s^{2(p-1)} \nabla f(y(s)) = 0.$$

a) Suppose that  $\alpha \geq 3$ . Given a trajectory  $x(\cdot)$  of  $(AVD)_\alpha$ , we obtain the following convergence rate of the values

$$(6) \quad f(y(s)) - \inf_{\mathcal{H}} f = \mathcal{O} \left( \frac{1}{s^{2p}} \right) \text{ as } s \rightarrow +\infty.$$

For  $p > 1$ , we have  $\alpha_p > \alpha$ , which gives damping features similar to  $(AVD)_\alpha$ . The only major difference is the coefficient  $s^{2(p-1)}$  in front of  $\nabla f(y(s))$  which explodes when  $s \rightarrow +\infty$ .

b) A similar time rescaling as above gives, for  $\alpha > 3$

$$(7) \quad f(y(s)) - \inf_{\mathcal{H}} f = o \left( \frac{1}{s^{2p}} \right) \text{ as } s \rightarrow +\infty.$$

## 2. FROM TIME RESCALING TO HIGHER ORDER IN TIME EVOLUTION SYSTEMS

Based on a rescaling approach, we introduce the system (TOGES) that supports our analysis.

### 2.1. Main result.

**Theorem 2.1.** *Let  $x : [t_0, +\infty[ \rightarrow \mathcal{H}$  be a solution trajectory of (TOGES) i.e.*

$$(8) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \frac{2\alpha - 6}{t^2} x(t) + \nabla f(x(t) + t\dot{x}(t)) = 0.$$

(i) *Suppose  $\alpha \geq 7$ . Then, there is a constant  $C > 0$  such that, for all  $t \geq t_0$*

$$\begin{aligned} f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f &\leq \frac{C}{t^3}; \\ f(x(t)) - \inf_{\mathcal{H}} f &\leq \frac{1}{t} \left( t_0 (f(x(t_0)) - \inf_{\mathcal{H}} f) + \frac{C}{2t_0^2} \right). \end{aligned}$$

*Moreover, we have an approximate descent method in the following sense:*

a) *The function  $t \mapsto t(f(x(t)) - \inf_{\mathcal{H}} f) + \frac{C}{2t^2}$  is nonincreasing on  $[t_0, +\infty[$ .*

b) *The function  $t \mapsto f(x(t)) + \frac{C}{3t^3}$  is nonincreasing on  $[t_0, +\infty[$ .*

(ii) *Suppose  $\alpha > 7$ . Then, as  $t \rightarrow +\infty$*

$$a) \quad f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f = o \left( \frac{1}{t^3} \right).$$

b)  $x(t)$  converges weakly in  $\mathcal{H}$ , and its limit belongs to  $S = \operatorname{argmin}_{\mathcal{H}} f$ .

*Proof.* (i) Take  $p = \frac{3}{2}$  in the formulas (5) and (6) obtained in the temporal rescaling analysis developed in section 1.2. We have  $\alpha_p = 1 + (\alpha - 1)p = \frac{1}{2}(3\alpha - 1)$ . When  $\alpha \geq 3$ , we obtain that, for any solution trajectory  $y(\cdot)$  of

$$(9) \quad \ddot{y}(t) + \frac{3\alpha - 1}{2t} \dot{y}(t) + \frac{9}{4} t \nabla f(y(t)) = 0,$$

we have

$$(10) \quad f(y(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^3}\right) \text{ as } t \rightarrow +\infty.$$

In fact, this convergence rate remains unchanged when replacing  $f$  with a homothetic  $cf$  with  $c > 0$ . So, for any solution trajectory  $y(\cdot)$  of

$$(11) \quad \ddot{y}(t) + \frac{3\alpha - 1}{2t} \dot{y}(t) + t \nabla f(y(t)) = 0,$$

we have that (10) is still verified. Set

$$y(t) := x(t) + t\dot{x}(t) = \frac{d}{dt}(tx(t)).$$

We have:

$$\dot{y}(t) = 2\dot{x}(t) + t\ddot{x}(t) \quad \text{and} \quad \ddot{y}(t) = 3\ddot{x}(t) + t\dddot{x}(t).$$

Replacing these expressions in (11), we obtain

$$(12) \quad t\ddot{x}(t) + 3\ddot{x}(t) + \frac{3\alpha - 1}{2t}(2\dot{x}(t) + t\ddot{x}(t)) + t\nabla f(x(t) + t\dot{x}(t)) = 0.$$

Equivalently,

$$(13) \quad t\ddot{x}(t) + \frac{3\alpha + 5}{2}\ddot{x}(t) + \frac{3\alpha - 1}{t}\dot{x}(t) + t\nabla f(x(t) + t\dot{x}(t)) = 0.$$

After division by  $t \geq t_0 > 0$ , we obtain

$$(14) \quad \ddot{x}(t) + \frac{3\alpha + 5}{2t}\ddot{x}(t) + \frac{3\alpha - 1}{t^2}\dot{x}(t) + \nabla f(x(t) + t\dot{x}(t)) = 0.$$

To simplify the formulation of the dynamic and of the results, let's make the parameter change  $\frac{3\alpha+5}{2} \mapsto \alpha$ . That makes  $3\alpha - 1$  becomes  $2\alpha - 6$ , and the condition  $\alpha \geq 3$  becomes  $\alpha \geq 7$ . Finally, we obtain that, under the condition  $\alpha \geq 7$ , for each solution trajectory  $x : [t_0, +\infty[ \rightarrow \mathcal{H}$  of

$$(15) \quad \ddot{x}(t) + \frac{\alpha}{t}\ddot{x}(t) + \frac{2\alpha - 6}{t^2}\dot{x}(t) + \nabla f(x(t) + t\dot{x}(t)) = 0,$$

we have

$$(16) \quad f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^3}\right) \text{ as } t \rightarrow +\infty.$$

That's the first statement of item (i). Let's now analyze the convergence rate of values for  $f(x(t)) - \inf_{\mathcal{H}} f$ . We start from (16), which can be equivalently written as follows

$$(17) \quad f(y(t)) - \inf_{\mathcal{H}} f \leq \frac{C}{t^3} \text{ as } t \rightarrow +\infty,$$

where  $C$  is a constant that can be made precise (we will return to its precise evaluation later). By integrating the relation  $\frac{d}{dt}(tx(t)) = y(t)$  from  $t$  to  $t + h$  ( $h$  is a positive parameter which is intended to go to zero) we get

$$(t + h)x(t + h) = tx(t) + \int_t^{t+h} y(\tau) d\tau.$$

Let us rewrite this relation in a barycentric form which is convenient to use a convexity argument:

$$x(t+h) = \left(1 - \frac{h}{t+h}\right)x(t) + \frac{h}{t+h} \frac{1}{h} \int_t^{t+h} y(\tau) d\tau.$$

According to the convexity of the function  $f - \inf_{\mathcal{H}} f$ , we obtain

$$f(x(t+h)) - \inf_{\mathcal{H}} f \leq \left(1 - \frac{h}{t+h}\right) (f(x(t)) - \inf_{\mathcal{H}} f) + \frac{h}{t+h} (f - \inf_{\mathcal{H}} f) \left( \frac{1}{h} \int_t^{t+h} y(\tau) d\tau \right).$$

Let us majorize the last expression above using Jensen's inequality. We obtain

$$f(x(t+h)) - \inf_{\mathcal{H}} f \leq \left(1 - \frac{h}{t+h}\right) (f(x(t)) - \inf_{\mathcal{H}} f) + \frac{1}{t+h} \int_t^{t+h} (f(y(\tau)) - \inf_{\mathcal{H}} f) d\tau.$$

According to the convergence rate of the values (17) which has been obtained for  $y$ , we obtain

$$f(x(t+h)) - \inf_{\mathcal{H}} f \leq \left(1 - \frac{h}{t+h}\right) (f(x(t)) - \inf_{\mathcal{H}} f) + \frac{1}{t+h} \int_t^{t+h} \frac{C}{\tau^3} d\tau.$$

Computing this last integral, we get

$$(18) \quad f(x(t+h)) - \inf_{\mathcal{H}} f \leq \left(1 - \frac{h}{t+h}\right) (f(x(t)) - \inf_{\mathcal{H}} f) + \frac{Ch(2t+h)}{2t^2(t+h)^3}.$$

Set

$$F(t) := f(x(t)) - \inf_{\mathcal{H}} f.$$

We can rewrite in an equivalent way (18) as

$$F(t+h) - F(t) + \frac{h}{t+h} F(t) \leq \frac{Ch(2t+h)}{2t^2(t+h)^3}.$$

Note that  $F$  is a  $\mathcal{C}^1$  function, as a composition of such functions. Therefore, dividing by  $h > 0$ , and letting  $h \rightarrow 0^+$  in the above inequality, we get

$$(19) \quad F'(t) + \frac{1}{t} F(t) \leq \frac{C}{t^4}.$$

Equivalently

$$(20) \quad \frac{d}{dt} tF(t) = tF'(t) + F(t) \leq \frac{C}{t^3} = -\frac{d}{dt} \frac{C}{2t^2}.$$

Therefore

$$(21) \quad \frac{d}{dt} \left( tF(t) + \frac{C}{2t^2} \right) \leq 0.$$

This gives that the function  $t \mapsto tF(t) + \frac{C}{2t^2}$  is nonincreasing. In particular

$$tF(t) + \frac{C}{2t^2} \leq t_0 F(t_0) + \frac{C}{2t_0^2},$$

which gives

$$f(x(t)) - \inf_{\mathcal{H}} f \leq \frac{1}{t} \left( t_0 (f(x(t_0)) - \inf_{\mathcal{H}} f) + \frac{C}{2t_0^2} \right).$$

From (19), we also get  $F'(t) \leq \frac{C}{t^4}$ , which gives

$$\frac{d}{dt} \left( F(t) + \frac{C}{3t^3} \right) \leq 0.$$

This gives that the function  $t \mapsto f(x(t)) + \frac{C}{3t^3}$  is nonincreasing.

ii) Take now  $\alpha > 3$ . We know that  $y(t)$  converges weakly to some  $y_\infty \in S$ . The relation  $y(t) := x(t) + t\dot{x}(t)$  writes equivalently

$$\frac{d}{dt}(tx(t)) = y(t).$$

Therefore

$$tx(t) = t_0x(t_0) + \int_{t_0}^t y(\tau)d\tau.$$

Equivalently

$$x(t) = \frac{t_0x(t_0)}{t} + \frac{1}{t} \int_{t_0}^t y(\tau)d\tau.$$

As a general rule, convergence implies ergodic convergence. Therefore,  $x(t)$  converges weakly to  $y_\infty \in S$ . Moreover we know that

$$(22) \quad f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right) \text{ as } t \rightarrow +\infty.$$

which gives

$$(23) \quad f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right) \text{ as } t \rightarrow +\infty.$$

□

**Remark 2.2.** Let us comment the "quasi-descent" property verified by the variable  $x(t)$ . For any  $t_0 \leq s \leq t$  we have

$$f(x(t)) + \frac{C}{3t^3} \leq f(x(s)) + \frac{C}{3s^3}$$

and hence

$$f(x(t)) \leq f(x(s)) + \frac{C}{3s^3} - \frac{C}{3t^3} \leq f(x(s)) + \frac{C}{3s^3}.$$

Thus, as soon as  $t \geq s$ ,  $f(x(t))$  cannot exceed  $f(x(s))$  by a quantity of order  $\frac{1}{s^3}$ . It means very small oscillations. Indeed, this results from the fact that  $x(t)$  can be obtained by averaging  $y(t)$ , with the result that the oscillations are almost eliminated.

### 3. A DIRECT ASYMPTOTIC CONVERGENCE ANALYSIS VIA LYAPUNOV ENERGY FUNCTION.

For numerical reasons, it is important to have an accurate estimate of the constant  $C$  that enters the convergence rate formula

$$f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \leq \frac{C}{t^3}$$

obtained in Theorem 2.1. It could be obtained by going back to the estimate obtained in [11] and do the same transformation as in the previous section. We prefer to establish them directly by developing a Lyapunov analysis on the (TOGES) system defined in (8). This will be useful to develop corresponding algorithmic results. It also makes the paper self-contained, and gives an other proof of Theorem 2.1. Indeed, our convergence results are based on the decrease property of the function  $t \mapsto E(t)$  defined by

$$(24) \quad E(t) := t^3 \left( f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2} \|t^2 \ddot{x}(t) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)\|^2$$

with  $z \in \operatorname{argmin} f$ .

**Theorem 3.1.** Let  $x : [t_0, +\infty[ \rightarrow \mathcal{H}$  be a solution trajectory of the (TOGES) system i.e.

$$(25) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \frac{2\alpha - 6}{t^2}x(t) + \nabla f(x(t) + t\dot{x}(t)) = 0.$$

(i) Suppose that the positive parameter  $\alpha$  satisfies the following condition:

$$(H_0) \quad \alpha \geq 7.$$

Then, for all  $t \geq t_0$ , the first statements of Theorem 2.1 are valid, the constant  $C$  being equal to <sup>2</sup>

$$E(t_0) = t_0^3 \{f(x(t_0) + t_0\dot{x}(t_0)) - \inf_{\mathcal{H}} f\} + \frac{1}{2} \|t_0^2\ddot{x}(t_0) + (\alpha - 2)t_0\dot{x}(t_0) + (\alpha - 4)(x(t_0) - z)\|^2.$$

(ii) Suppose moreover that  $\alpha > 7$ . Then,

$$\int_{t_0}^{\infty} s^2 \left( f(x(s) + t\dot{x}(s)) - \inf_{\mathcal{H}} f \right) dt \leq \frac{1}{\alpha - 7} E(t_0).$$

*Proof.* Let us rewrite the energy function  $E(\cdot)$  defined in (24) as follows:

$$(26) \quad E(t) := t^3(f(y(t)) - \inf_{\mathcal{H}} f) + \frac{1}{2} \|v(t)\|^2,$$

where  $y(t) := x(t) + t\dot{x}(t)$  and  $v(t) := t^2\ddot{x}(t) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)$ , with  $z \in \operatorname{argmin} f$ . First note that

$$\begin{aligned} \dot{v}(t) &= t^2\ddot{x}(t) + t\alpha\dot{x}(t) + (2\alpha - 6)\dot{x}(t) \\ &= -t^2\nabla f(y(t)), \end{aligned}$$

where the last inequality follows from the evolution equation (25) and the definition of  $y(\cdot)$ . According to the above result and  $\dot{y}(t) = t\ddot{x}(t) + 2\dot{x}(t)$ , we obtain

$$\begin{aligned} \dot{E}(t) &= 3t^2(f(y(t)) - \inf_{\mathcal{H}} f) + t^3\langle \nabla f(y(t)), \dot{y}(t) \rangle \\ &\quad - t^2\langle \nabla f(y(t)), t^2\ddot{x}(t) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z) \rangle \\ &= 3t^2(f(y(t)) - \inf_{\mathcal{H}} f) + t^3\langle \nabla f(y(t)), \dot{y}(t) \rangle - t^3\langle \nabla f(y(t)), t\ddot{x}(t) + 2\dot{x}(t) \rangle \\ &\quad - t^2\langle \nabla f(y(t)), (\alpha - 4)(x(t) + t\dot{x}(t) - z) \rangle. \end{aligned}$$

Since  $\dot{y}(t) = t\ddot{x}(t) + 2\dot{x}(t)$ , the expression above can be simplified to give

$$\dot{E}(t) - 3t^2(f(y(t)) - \inf_{\mathcal{H}} f) + (\alpha - 4)t^2\langle \nabla f(y(t)), y(t) - z \rangle \leq 0.$$

By convexity of  $f$  and definition of  $z$ , we have  $\langle \nabla f(y(t)), y(t) - z \rangle \geq f(y(t)) - f(z) = f(y(t)) - \inf_{\mathcal{H}} f$ . So, we finally obtain

$$(27) \quad \dot{E}(t) + (\alpha - 7)t^2(f(y(t)) - \inf_{\mathcal{H}} f) \leq 0.$$

According to the hypothesis  $(H_0)$ , we have  $\alpha \geq 7$ . Therefore,  $\dot{E}(t) \leq 0$ , which gives that  $E(\cdot)$  is a nonincreasing function. This implies that, for all  $t \geq t_0$

$$t^3 \left( f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \right) \leq E(t) \leq E(t_0),$$

which gives

$$(28) \quad f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \leq \frac{E(t_0)}{t^3}.$$

<sup>2</sup>The point  $z \in \operatorname{argmin} f$  can be chosen arbitrarily.



By integrating (27), we obtain

$$(\alpha - 7) \int_{t_0}^t s^2 (f(x(s) + s\dot{x}(s)) - \inf_{\mathcal{H}} f) ds \leq E(t_0) - E(t) \leq E(t_0).$$

This being true for any  $t \geq t_0$ , we obtain, when  $\alpha > 7$

$$\int_{t_0}^{\infty} s^2 \left( f(x(s) + s\dot{x}(s)) - \inf_{\mathcal{H}} f \right) ds \leq \frac{1}{\alpha - 7} E(t_0).$$

The corresponding convergence rates for the variable  $x(t)$  are obtained by following a similar argument as in the proof of Theorem 2.1.  $\square$

**3.1. Perturbations, errors.** For numerical reasons, it is important to study the stability with respect to perturbations of the evolution system (TOGES) which supports the related algorithms. In our context, this leads to consider the system

$$(29) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \frac{2\alpha - 6}{t^2} \dot{x}(t) + \nabla f(x(t) + t\dot{x}(t)) = e(t)$$

where the second member  $e(\cdot)$  takes account perturbations, errors. We use the same ingredients as in the Lyapunov analysis of the unperturbed case, that is

$$E(t) := t^3 (f(y(t)) - \inf_{\mathcal{H}} f) + \frac{1}{2} \|v(t)\|^2,$$

where  $y(t) := x(t) + t\dot{x}(t)$  and  $v(t) := t^2 \ddot{x}(t) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)$ , with  $z \in \operatorname{argmin} f$ . First note that

$$(30) \quad \begin{aligned} \dot{v}(t) &= t^2 \ddot{x}(t) + t\alpha \dot{x}(t) + (2\alpha - 6)\dot{x}(t) \\ &= -t^2 \nabla f(y(t)) + t^2 e(t). \end{aligned}$$

**Theorem 3.2.** *Let  $x : [t_0, +\infty[ \rightarrow \mathcal{H}$  be a solution trajectory of the perturbed (TOGES) system i.e.*

$$(31) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \frac{2\alpha - 6}{t^2} \dot{x}(t) + \nabla f(x(t) + t\dot{x}(t)) = e(t).$$

(i) *Suppose that the positive parameter  $\alpha$  satisfies the following condition:*

$$(H_0) \quad \alpha \geq 7,$$

*and that the perturbations, errors satisfy the integrability property*

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

*Then, for all  $t \geq t_0$*

$$\begin{aligned} f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f &\leq \frac{C}{t^3}; \\ f(x(t)) - \inf_{\mathcal{H}} f &\leq \frac{1}{t} \left( t_0 (f(x(t_0)) - \inf_{\mathcal{H}} f) + \frac{C}{2t_0^2} \right), \end{aligned}$$

*with*

$$C = E(t_0) + \left( \sqrt{2E(t_0)} + \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau \right) \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau.$$

*Proof.* According to (30), the derivation of  $E(\cdot)$  gives

$$\begin{aligned}\dot{E}(t) &= 3t^2(f(y(t)) - \inf_{\mathcal{H}} f) + t^3\langle \nabla f(y(t)), \dot{y}(t) \rangle + \langle t^2 e(t) - t^2 \nabla f(y(t)), v(t) \rangle \\ &= 3t^2(f(y(t)) - \inf_{\mathcal{H}} f) + t^3\langle \nabla f(y(t)), \dot{y}(t) \rangle - t^3\langle \nabla f(y(t)), t\ddot{x}(t) + 2\dot{x}(t) \rangle \\ &\quad - t^2\langle \nabla f(y(t)), (\alpha - 4)(x(t) + t\dot{x}(t) - z) \rangle + \langle t^2 e(t), v(t) \rangle.\end{aligned}$$

Since  $\dot{y}(t) = t\ddot{x}(t) + 2\dot{x}(t)$  the expression above can be simplified to give

$$\dot{E}(t) - 3t^2(f(y(t)) - \inf_{\mathcal{H}} f) + (\alpha - 4)t^2\langle \nabla f(y(t)), y(t) - z \rangle = \langle t^2 e(t), v(t) \rangle.$$

By convexity of  $f$  and definition of  $z$ , we have  $\langle \nabla f(y(t)), y(t) - z \rangle \geq f(y(t)) - f(z) = f(y(t)) - \inf_{\mathcal{H}} f$ . So, we finally obtain

$$(32) \quad \dot{E}(t) + (\alpha - 7)t^2(f(y(t)) - \inf_{\mathcal{H}} f) \leq \langle t^2 e(t), v(t) \rangle.$$

According to the hypothesis  $(H_0)$ , we have  $\alpha \geq 7$ . By Cauchy-Schwarz inequality, we deduce that

$$\dot{E}(t) \leq \|t^2 e(t)\| \|v(t)\|,$$

which, by integration gives

$$(33) \quad E(t) \leq E(t_0) + \int_{t_0}^t \|\tau^2 e(\tau)\| \|v(\tau)\| d\tau.$$

By definition of  $E(\cdot)$ , we have  $E(t) \geq \frac{1}{2}\|v(t)\|^2$ . Therefore,

$$\frac{1}{2}\|v(t)\|^2 \leq E(t_0) + \int_{t_0}^t \|\tau^2 e(\tau)\| \|v(\tau)\| d\tau.$$

According to Gronwall lemma (see [23, Lemme A.5]), we deduce that

$$\|v(t)\| \leq \sqrt{2E(t_0)} + \int_{t_0}^t \|\tau^2 e(\tau)\| d\tau.$$

According to the integrability assumption on the error terms, we get

$$\|v(t)\| \leq \sqrt{2E(t_0)} + \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau < +\infty.$$

Returning to (33), we obtain

$$(34) \quad E(t) \leq E(t_0) + \left( \sqrt{2E(t_0)} + \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau \right) \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau,$$

which gives that  $E(\cdot)$  is bounded from above. This implies that, for  $t \geq t_0$

$$t^3 \left( f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \right) \leq E(t_0) + \left( \sqrt{2E(t_0)} + \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau \right) \int_{t_0}^{+\infty} \|\tau^2 e(\tau)\| d\tau.$$

The end of the proof is similar to the unperturbed case.  $\square$

## 4. INERTIAL DYNAMICS FOR STRONGLY CONVEX FUNCTIONS

In this section, we assume that  $f$  is a strongly convex function. Recall that  $f : \mathcal{H} \rightarrow \mathbb{R}$  is  $\mu$ -strongly convex for some  $\mu > 0$ , if  $f - \frac{\mu}{2} \|\cdot\|^2$  is convex. Let us show that a judicious choice of the damping coefficients (adapted to  $\mu$ ) in our third order evolution system gives exponential convergence rate. This is consistent with similar type results obtained in [43, Theorem 2.2], [47, Theorem 1] for second order gradient based systems. Precisely, we consider the autonomous evolution system

$$(35) \quad \ddot{x}(t) + (1 + 2\sqrt{\mu})\dot{x}(t) + 2\sqrt{\mu}x(t) + \nabla f(x(t) + \dot{x}(t)) = 0.$$

We will develop a Lyapunov analysis based on the function  $t \mapsto \mathcal{E}(t)$  where, for all  $t \geq t_0$

$$\mathcal{E}(t) := f(x(t) + \dot{x}(t)) - \inf_{\mathcal{H}} f + \frac{1}{2} \|\sqrt{\mu}(x(t) + \dot{x}(t) - x^*) + \dot{x}(t) + \ddot{x}(t)\|^2,$$

and  $x^*$  is the unique minimizer of  $f$ . To condense the formulas, we write

$$F(t_0) = f(x(t_0)) - \inf_{\mathcal{H}} f \quad \text{and} \quad C = \mathcal{E}(t_0)e^{\sqrt{\mu}t_0}.$$

**Theorem 4.1.** *Suppose that  $f : \mathcal{H} \rightarrow \mathbb{R}$  is  $\mu$ -strongly convex for some  $\mu > 0$ . Let  $x(\cdot)$  be a solution trajectory of (35). Then, the following properties hold:*

(i) For all  $t \geq t_0$

$$(36) \quad f(x(t) + \dot{x}(t)) - \inf_{\mathcal{H}} f \leq \mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)}$$

$$(37) \quad \text{for } \mu \neq 1: \quad f(x(t)) - \inf_{\mathcal{H}} f \leq \frac{2\mathcal{E}(t_0)e^{\sqrt{\mu}t_0}}{1 - \sqrt{\mu}}e^{-\sqrt{\mu}t} + e^{-t} \left( e^{t_0}F(t_0) - \frac{2\mathcal{E}(t_0)e^{\sqrt{\mu}t_0}}{1 - \sqrt{\mu}}e^{(1-\sqrt{\mu})t_0} \right)$$

$$(38) \quad \text{for } \mu = 1 \quad f(x(t)) - \inf_{\mathcal{H}} f \leq (Ct + e^{t_0}F(t_0) - Ct_0)e^{-t}.$$

(ii) For all  $t \geq t_0$

$$(39) \quad \|x(t) + \dot{x}(t) - x^*\|^2 \leq \frac{e^{\sqrt{\mu}t_0}}{\sqrt{\mu}} \left( \sqrt{\mu}\|y(t_0) - x^*\|^2 + 2\mathcal{E}(t_0)(t - t_0) \right) e^{-\sqrt{\mu}t}.$$

$$(40) \quad \text{for } \mu \neq 1: \quad \|x(t) - x^*\|^2 \leq \frac{2}{\mu} \left( \frac{2\mathcal{E}(t_0)e^{\sqrt{\mu}t_0}}{1 - \sqrt{\mu}}e^{-\sqrt{\mu}t} + e^{-t} \left( e^{t_0}F(t_0) - \frac{2\mathcal{E}(t_0)e^{\sqrt{\mu}t_0}}{1 - \sqrt{\mu}}e^{(1-\sqrt{\mu})t_0} \right) \right)$$

$$(41) \quad \text{for } \mu = 1: \quad \|x(t) - x^*\|^2 \leq \frac{2}{\mu} \left( (Ct + e^{t_0}F(t_0) - Ct_0)e^{-t} \right).$$

*Proof.* Set  $y(t) = x(t) + \dot{x}(t)$ . Note that (35) can be equivalently written

$$(42) \quad \ddot{y}(t) + 2\sqrt{\mu}\dot{y}(t) + \nabla f(y(t)) = 0,$$

and that  $\mathcal{E} : [t_0, +\infty[ \rightarrow \mathbb{R}^+$  writes

$$(43) \quad \mathcal{E}(t) := f(y(t)) - \inf_{\mathcal{H}} f + \frac{1}{2} \|\sqrt{\mu}(y(t) - x^*) + \dot{y}(t)\|^2.$$

Set  $v(t) = \sqrt{\mu}(y(t) - x^*) + \dot{y}(t)$ . Derivation of  $\mathcal{E}(\cdot)$  gives

$$\frac{d}{dt}\mathcal{E}(t) := \langle \nabla f(y(t)), \dot{y}(t) \rangle + \langle v(t), \sqrt{\mu}\dot{y}(t) + \ddot{y}(t) \rangle.$$

According to (42), we obtain

$$\frac{d}{dt}\mathcal{E}(t) = \langle \nabla f(y(t)), \dot{y}(t) \rangle + \langle \sqrt{\mu}(y(t) - x^*) + \dot{y}(t), -\sqrt{\mu}\dot{y}(t) - \nabla f(y(t)) \rangle.$$

After developing and simplification, we obtain

$$\frac{d}{dt}\mathcal{E}(t) + \sqrt{\mu}\langle \nabla f(y(t)), y(t) - x^* \rangle + \mu\langle y(t) - x^*, \dot{y}(t) \rangle + \sqrt{\mu}\|\dot{y}(t)\|^2 = 0.$$

According to the strong convexity of  $f$ , we have

$$\langle \nabla f(y(t)), y(t) - x^* \rangle \geq f(y(t)) - f(x^*) + \frac{\mu}{2}\|y(t) - x^*\|^2.$$

By combining the two relations above, and using the formulation (43) of  $\mathcal{E}(t)$ , we obtain

$$\frac{d}{dt}\mathcal{E}(t) + \sqrt{\mu}\left(\mathcal{E}(t) + \frac{1}{2}\|\dot{y}\|^2\right) \leq 0.$$

Therefore,

$$\frac{d}{dt}\mathcal{E}(t) + \sqrt{\mu}\mathcal{E}(t) \leq 0.$$

By integrating the differential inequality above, we obtain, for all  $t \geq t_0$

$$\mathcal{E}(t) \leq \mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)}.$$

By definition of  $\mathcal{E}(t)$  and  $y(t)$ , we deduce that

$$(44) \quad f(y(t)) - \inf_{\mathcal{H}} f = f(x(t) + \dot{x}(t)) - \inf_{\mathcal{H}} f \leq \mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)},$$

and

$$\|\sqrt{\mu}(y(t) - x^*) + \dot{y}(t)\|^2 = \|\sqrt{\mu}(x(t) + \dot{x}(t) - x^*) + \dot{x}(t) + \ddot{x}(t)\|^2 \leq 2\mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)}.$$

(ii) Set  $C = \mathcal{E}(t_0)e^{\sqrt{\mu}t_0}$ . Developing the expression above, we obtain

$$\mu\|y(t) - x^*\|^2 + \|\dot{y}(t)\|^2 + \langle \dot{y}(t), 2\sqrt{\mu}(y(t) - x^*) \rangle \leq 2Ce^{-\sqrt{\mu}t}.$$

Note that

$$\langle \dot{y}(t), 2\sqrt{\mu}(y(t) - x^*) \rangle = \frac{d}{dt}(\sqrt{\mu}\|y(t) - x^*\|^2).$$

Combining the above results, we obtain

$$\sqrt{\mu}(\sqrt{\mu}\|y(t) - x^*\|^2) + \frac{d}{dt}(\sqrt{\mu}\|y(t) - x^*\|^2) \leq 2Ce^{-\sqrt{\mu}t}.$$

Set  $Z(t) := \sqrt{\mu}\|y(t) - x^*\|^2$ , then we have

$$\frac{d}{dt}(e^{\sqrt{\mu}t}Z(t) - 2Ct) = e^{\sqrt{\mu}t}\frac{d}{dt}Z(t) + \sqrt{\mu}e^{\sqrt{\mu}t}Z(t) - 2C \leq 0.$$

By integrating this differential inequality, elementary computation gives

$$Z(t) \leq e^{-\sqrt{\mu}(t-t_0)}Z(t_0) + 2C(t-t_0)e^{-\sqrt{\mu}t}.$$

Therefore

$$\|y(t) - x^*\|^2 \leq \frac{e^{\sqrt{\mu}t_0}}{\sqrt{\mu}}(\sqrt{\mu}\|y(t_0) - x^*\|^2 + 2\mathcal{E}(t_0)(t-t_0))e^{-\sqrt{\mu}t}.$$

(iii) Let's now analyze the convergence rate of values for  $f(x(t)) - \inf_{\mathcal{H}} f$ . We start from (44), which can be equivalently written as follows: for all  $t \geq t_0$

$$(45) \quad f(y(t)) - \inf_{\mathcal{H}} f \leq Ce^{-\sqrt{\mu}t},$$

where, we recall,  $C = \mathcal{E}(t_0)e^{\sqrt{\mu}t_0}$ . By integrating the relation  $\frac{d}{dt}(e^t x(t)) = e^t y(t)$  from  $t$  to  $t+h$  ( $h$  is a positive parameter which is intended to go to zero), we get

$$e^{t+h}x(t+h) = e^t x(t) + \int_t^{t+h} e^\tau y(\tau) d\tau.$$

Equivalently,

$$x(t+h) = e^{-h}x(t) + e^{-(t+h)} \int_t^{t+h} e^\tau y(\tau) d\tau.$$

Let us rewrite this relation in a barycentric form, which is convenient to use a convexity argument:

$$\begin{aligned} x(t+h) &= e^{-h}x(t) + e^{-(t+h)} \frac{e^{t+h} - e^t}{e^{t+h} - e^t} \int_t^{t+h} e^\tau y(\tau) d\tau \\ (46) \quad &= e^{-h}x(t) + (1 - e^{-h}) \frac{1}{e^{t+h} - e^t} \int_t^{t+h} e^\tau y(\tau) d\tau. \end{aligned}$$

According to the convexity of the function  $f - \inf_{\mathcal{H}} f$ , we obtain

$$(47) \quad f(x(t+h)) - \inf_{\mathcal{H}} f \leq e^{-h}(f(x(t)) - \inf_{\mathcal{H}} f) + (1 - e^{-h})(f - \inf_{\mathcal{H}} f) \left( \frac{1}{e^{t+h} - e^t} \int_t^{t+h} e^\tau y(\tau) d\tau \right).$$

Let us apply the Jensen inequality to majorize the last above expression. We obtain

$$(48) \quad f(x(t+h)) - \inf_{\mathcal{H}} f \leq e^{-h}(f(x(t)) - \inf_{\mathcal{H}} f) + \frac{1 - e^{-h}}{e^{t+h} - e^t} \int_t^{t+h} e^\tau (f(y(\tau)) - \inf_{\mathcal{H}} f) d\tau.$$

According to the convergence rate of the values (45) which has been obtained for  $y$  we obtain

$$(49) \quad f(x(t+h)) - \inf_{\mathcal{H}} f \leq e^{-h}(f(x(t)) - \inf_{\mathcal{H}} f) + \frac{1 - e^{-h}}{e^{t+h} - e^t} \int_t^{t+h} e^\tau C e^{-\sqrt{\mu}\tau} d\tau.$$

Set  $F(t) := f(x(t)) - \inf_{\mathcal{H}} f$ . We can rewrite equivalently (49) as

$$F(t+h) - F(t) + (1 - e^{-h})F(t) \leq C e^{-t} \frac{1 - e^{-h}}{e^h - 1} \int_t^{t+h} e^{(1-\sqrt{\mu})\tau} d\tau.$$

Note that  $F$  is a  $\mathcal{C}^1$  function, as a composition of such functions. Therefore, dividing by  $h > 0$ , and letting  $h \rightarrow 0^+$  in the above inequality, we get by elementary calculation

$$(50) \quad F'(t) + F(t) \leq C e^{-t} e^{(1-\sqrt{\mu})t} = C e^{-\sqrt{\mu}t}.$$

If  $\mu \neq 1$ , equivalently, we have

$$(51) \quad \frac{d}{dt} (e^t F(t)) = e^t (F'(t) + F(t)) \leq C e^{(1-\sqrt{\mu})t} = \frac{d}{dt} \left( \frac{C}{1 - \sqrt{\mu}} e^{(1-\sqrt{\mu})t} \right).$$

Therefore

$$(52) \quad \frac{d}{dt} \left( e^t F(t) - \frac{C}{1 - \sqrt{\mu}} e^{(1-\sqrt{\mu})t} \right) \leq 0.$$

This gives that the function  $t \mapsto e^t F(t) - \frac{C}{1 - \sqrt{\mu}} e^{(1-\sqrt{\mu})t}$  is nonincreasing. In particular

$$e^t F(t) - \frac{C}{1 - \sqrt{\mu}} e^{(1-\sqrt{\mu})t} \leq e^{t_0} F(t_0) - \frac{C}{1 - \sqrt{\mu}} e^{(1-\sqrt{\mu})t_0}$$

which gives

$$f(x(t)) - \inf_{\mathcal{H}} f \leq \frac{C}{1 - \sqrt{\mu}} e^{-\sqrt{\mu}t} + e^{-t} \left( e^{t_0} F(t_0) - \frac{C}{1 - \sqrt{\mu}} e^{(1-\sqrt{\mu})t_0} \right).$$

If  $\mu = 1$ , we get from (50),

$$\frac{d}{dt} (e^t F(t)) = e^t (F'(t) + F(t)) \leq C = \frac{d}{dt} (Ct),$$

which gives

$$f(x(t)) - \inf_{\mathcal{H}} f \leq (Ct + e^{t_0} F(t_0) - Ct_0) e^{-t}.$$

□

**Remark 4.2.** Let's justify the choice of  $\gamma = 2\sqrt{\mu}$  in theorem 4.1. Indeed, considering

$$\ddot{y}(t) + 2\gamma\dot{y}(t) + \nabla f(y(t)) = 0,$$

a similar demonstration to that described above can be made on the basis of the Lyapunov function

$$\mathcal{E}(t) := f(y(t)) - \inf_{\mathcal{H}} f + \frac{1}{2} \|\gamma(y(t) - x^*) + \dot{y}(t)\|^2.$$

Under the conditions  $\gamma \leq \sqrt{\mu}$  we obtain the exponential convergence rate

$$f(y(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(e^{-\gamma t}) \text{ as } t \rightarrow +\infty.$$

Taking  $\gamma = \sqrt{\mu}$  gives the best convergence rate, and the result of theorem 4.1.

## 5. THE THIRD ORDER DYNAMIC WITH THE HESSIAN DRIVEN DAMPING

In this section,  $f$  is a convex function which is twice continuously differentiable. The Hessian of  $f$  at  $x$  is denoted by  $\nabla^2 f(x)$ . It belongs to  $\mathcal{L}(\mathcal{H}, \mathcal{H})$ , its action on  $\xi \in \mathcal{H}$  is denoted by  $\nabla^2 f(x)(\xi)$ . The introduction of the Hessian driven damping into the third order system (TOGES) leads to consider

$$(53) \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \frac{2\alpha - 6}{t^2} \dot{x}(t) + \beta \nabla^2 f(x(t) + t\dot{x}(t))(2\dot{x}(t) + t\ddot{x}(t)) + \nabla f(x(t) + t\dot{x}(t)) = 0,$$

where  $\alpha$  and  $\beta$  are positive parameters. The Hessian driven damping has proved to be an efficient tool to control and attenuate the oscillations of inertial systems, which is a central issue for optimization purposes, see [2], [8], [16], [22], [42]. Note that  $\nabla^2 f(x(t) + t\dot{x}(t))(2\dot{x}(t) + t\ddot{x}(t))$  is exactly the time derivative of  $\nabla f(x(t) + t\dot{x}(t))$ . This plays a key role in the following developments.

**5.1. Convergence via Lyapunov analysis.** Our convergence results are based on the decrease property of the function  $t \mapsto E(t)$  defined by

$$(54) \quad \begin{aligned} E(t) &:= (t^3 - 2\beta t^2) \left( f(x(t) + t\dot{x}(t)) - m \right) \\ &+ \frac{1}{2} \|t^2(\ddot{x}(t) + \beta \nabla f(x(t) + t\dot{x}(t))) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)\|^2 \end{aligned}$$

with  $m = \inf_{\mathcal{H}} f$  and  $z \in \operatorname{argmin} f$ . When  $\beta = 0$ , we recover the Lyapunov function used in the proof of Theorem 3.1.

**Theorem 5.1.** *Let  $x : [t_0, +\infty[ \rightarrow \mathcal{H}$  be a solution trajectory of the evolution equation (53).*

(i) *Suppose that the parameters satisfy the following condition:*

$$\alpha > 7 \quad \text{and} \quad \beta \geq 0.$$

*Then, for all  $t \geq t_1 := \frac{2\beta(\alpha-6)}{\alpha-7}$*

$$\begin{aligned} f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f &\leq \frac{(\alpha - 6)E(t_1)}{t^3}; \\ f(x(t)) - \inf_{\mathcal{H}} f &\leq \frac{1}{t} \left( t_1 (f(x(t_1)) - \inf_{\mathcal{H}} f) + \frac{(\alpha - 6)E(t_1)}{2t_1^2} \right), \end{aligned}$$

with

$$\begin{aligned} E(t_1) &= (t_1^3 - 2\beta t_1^2) \left( f(x(t_1) + t_1 \dot{x}(t_1)) - \inf_{\mathcal{H}} f \right) \\ &+ \frac{1}{2} \|t_1^2(\ddot{x}(t_1) + \beta \nabla f(x(t_1) + t_1 \dot{x}(t_1))) + (\alpha - 2)t_1 \dot{x}(t_1) + (\alpha - 4)(x(t_1) - z)\|^2. \end{aligned}$$

Moreover, we have an approximate descent method in the following sense:

a) The function  $t \mapsto t(f(x(t)) - \inf_{\mathcal{H}} f) + \frac{(\alpha-6)E(t_1)}{2t^2}$  is nonincreasing on  $[t_1, +\infty[$ .

b) The function  $t \mapsto f(x(t)) + \frac{(\alpha-6)E(t_1)}{3t^3}$  is nonincreasing on  $[t_1, +\infty[$ .

(ii) In addition, for  $\beta > 0$

$$\int_{t_1}^{\infty} t^4 \|\nabla f(x(t) + t\dot{x}(t))\|^2 dt \leq \frac{(\alpha-6)E(t_1)}{\beta}.$$

*Proof.* Let us rewrite the energy function  $E(\cdot)$  defined in (54) as follows:

$$(55) \quad E(t) := t\delta(t)(f(y(t)) - \inf_{\mathcal{H}} f) + \frac{1}{2} \|v(t)\|^2,$$

where  $y(t) := x(t) + t\dot{x}(t)$ ,  $v(t) := t^2(\ddot{x}(t) + \beta \nabla f(x(t) + t\dot{x}(t))) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)$ , with  $z \in \operatorname{argmin} f$ , and  $\delta(t) = t^2 \left(1 - \frac{2\beta}{t}\right)$ . We have

$$(56) \quad \dot{E}(t) := t\delta(t)\langle \nabla f(y(t)), \dot{y}(t) \rangle + (\delta(t) + t\dot{\delta}(t))(f(y(t)) - \inf_{\mathcal{H}} f) + \langle v(t), \dot{v}(t) \rangle.$$

Let us compute  $\dot{v}(t)$  with the help of the derivation chain rule.

$$\begin{aligned} \dot{v}(t) &= t^2 \left( \ddot{x}(t) + \beta \nabla^2 f(x(t) + t\dot{x}(t))(2\dot{x}(t) + t\ddot{x}(t)) \right) + 2t \left( \ddot{x}(t) + \beta \nabla f(x(t) + t\dot{x}(t)) \right) \\ &+ (\alpha - 2)(t\ddot{x}(t) + \dot{x}(t)) + (\alpha - 4)\dot{x}(t) \\ &= \left( t^2 \ddot{x}(t) + \alpha t \ddot{x}(t) + (2\alpha - 6)\dot{x}(t) + t^2 \beta \nabla^2 f(x(t) + t\dot{x}(t))(2\dot{x}(t) + t\ddot{x}(t)) \right) + 2t \beta \nabla f(x(t) + t\dot{x}(t)) \\ &= -t^2 \nabla f(x(t) + t\dot{x}(t)) + 2t \beta \nabla f(x(t) + t\dot{x}(t)) \\ &= -\delta(t) \nabla f(y(t)) \end{aligned}$$

where the two last inequalities follow from the equation (53) and the definition of  $y(\cdot)$ . Therefore

$$\begin{aligned} \langle \dot{v}(t), v(t) \rangle &= -\delta(t) \langle \nabla f(y(t)), t^2(\ddot{x}(t) + \beta \nabla f(y(t)) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)) \rangle \\ &= -\delta(t) \langle \nabla f(y(t)), t^2 \ddot{x}(t) + \beta t^2 \nabla f(y(t)) \rangle + (\alpha - 4)(y(t) - z) + 2t\dot{x}(t) \\ &= -\beta t^2 \delta(t) \|\nabla f(y(t))\|^2 - t\delta(t) \langle \nabla f(y(t)), \dot{y}(t) \rangle - (\alpha - 4)\delta(t) \langle \nabla f(y(t)), y(t) - z \rangle \\ (57) \quad &\leq -\beta t^2 \delta(t) \|\nabla f(y(t))\|^2 - t\delta(t) \langle \nabla f(y(t)), \dot{y}(t) \rangle + (\alpha - 4)\delta(t)(f(z) - f(y(t))) \end{aligned}$$

where the last inequality follows from the convexity of  $f$  and  $\alpha \geq 4$ .

Combining (56) with (57) we obtain

$$(58) \quad \begin{aligned} \dot{E}(t) &\leq t\delta(t) \langle \nabla f(y(t)), \dot{y}(t) \rangle + (\delta(t) + t\dot{\delta}(t))(f(y(t)) - \inf_{\mathcal{H}} f) \\ &- \beta t^2 \delta(t) \|\nabla f(y(t))\|^2 - t\delta(t) \langle \nabla f(y(t)), \dot{y}(t) \rangle + (\alpha - 4)\delta(t)(f(z) - f(y(t))). \end{aligned}$$

After simplification we get

$$(59) \quad \dot{E}(t) + \beta t^2 \delta(t) \|\nabla f(y(t))\|^2 + \left( (\alpha - 4)\delta(t) - \delta(t) - t\dot{\delta}(t) \right) (f(y(t)) - \inf_{\mathcal{H}} f) \leq 0.$$

According to the definition of  $\delta(\cdot)$

$$\begin{aligned} (\alpha - 4)\delta(t) - \delta(t) - t\dot{\delta}(t) &= (\alpha - 5)\delta(t) - t\dot{\delta}(t) \\ &= (\alpha - 5)(t^2 - 2\beta t) - t(2t - 2\beta) \\ &= (\alpha - 7)t^2 - 2\beta t(\alpha - 6). \end{aligned}$$

Therefore, for  $\alpha > 7$  and  $t \geq t_1 = \frac{2\beta(\alpha-6)}{\alpha-7}$  (i.e.  $t$  sufficiently large), we have  $\dot{E}(t) \leq 0$ . So, for all  $t \geq t_1$  we have  $E(t) \leq E(t_1)$ , which by definition of  $E(\cdot)$  gives

$$t^3 \left(1 - \frac{2\beta}{t}\right) (f(y(t)) - \inf_{\mathcal{H}} f) \leq E(t_1).$$

Note that for  $t \geq t_1$  we have  $1 - \frac{2\beta}{t} \geq \frac{1}{\alpha-6}$ . Therefore, for  $t \geq t_1$

$$f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \leq \frac{(\alpha - 6)E(t_1)}{t^3}.$$

Moreover by integrating (59) we get

$$\beta \int_{t_1}^{\infty} t^2 \delta(t) \|\nabla f(x(t) + t\dot{x}(t))\|^2 dt \leq E(t_1).$$

which, by a similar argument as above, gives the claim. Finally, passing from the estimates on  $y(t)$  to the estimates on  $x(t)$  is obtained as in Theorem 3.1.  $\square$

## 6. CONVERGENCE OF AN IMPLICIT DISCRETIZATION OF THE THIRD ORDER EVOLUTION

Let us analyze the fast convergence properties of the proximal algorithms that can be obtained by temporal discretization of the third order evolution system (TOGES)

$$(60) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \frac{2\alpha - 6}{t^2}x(t) + \nabla f(x(t) + t\dot{x}(t)) = 0.$$

To introduce the algorithm, we first assume that  $f : \mathcal{H} \rightarrow \mathbb{R}$  is continuously differentiable. Then, when passing to the proximal algorithm, we will take a general convex lower semicontinuous function  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ . We consider the following implicit temporal discretization of equation (60), with fixed step size  $h = 1$ ,

$$(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}) + \frac{\alpha}{k-1}(x_{k+2} - 2x_{k+1} + x_k) + \frac{2\alpha - 6}{k(k-1)}(x_{k+2} - x_{k+1}) + \nabla f(y_{k+1}) = 0$$

where

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

Note that

$$y_k = (k+1)x_{k+1} - kx_k,$$

which makes convenient to introduce  $a_k := kx_k$ , and gives

$$y_k = a_{k+1} - a_k.$$

As a discrete version of the Lyapunov function considered in the continuous case

$$\begin{aligned} E(t) &:= t^3 \left( f(x(t) + t\dot{x}(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2} \|t^2 \ddot{x}(t) + (\alpha - 2)t\dot{x}(t) + (\alpha - 4)(x(t) - z)\|^2 \\ &= t^3 \left( f(y(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2} \|t\dot{y}(t) + (\alpha - 4)(y(t) - z)\|^2 \end{aligned}$$

where  $z \in \operatorname{argmin} f$ , we take

$$E_k := k(k-1)(k+1)(f(y_k) - \inf_{\mathcal{H}} f) + \frac{1}{2} \|v_k\|^2$$



where

$$v_k = (\alpha - 4)(y_k - z) + k(y_k - y_{k-1}).$$

Note that, after multiplication by  $k(k-1)$ , the algorithm can be equivalently formulated as

$$(61) \quad \begin{aligned} & k(k-1)(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}) + \alpha k(x_{k+2} - 2x_{k+1} + x_k) \\ & + (2\alpha - 6)(x_{k+2} - x_{k+1}) + k(k-1)\nabla f(y_{k+1}) = 0. \end{aligned}$$

Let's first verify that the algorithm is well posed. This results from the basic formula that will be established in the proof of following theorem

$$v_{k+1} - v_k = -k(k-1)\nabla f(y_{k+1}).$$

According to the definition of  $v_k$ , this gives

$$(\alpha - 4)(y_{k+1} - y_k) + (k+1)(y_{k+1} - y_k) - k(y_k - y_{k-1}) + k(k-1)\nabla f(y_{k+1}) = 0.$$

Equivalently,

$$y_{k+1} = \text{prox}_{\frac{k(k-1)}{k+\alpha-3}f} \left( y_k + \frac{k}{k+\alpha-3}(y_k - y_{k-1}) \right).$$

Hence  $(y_k)$  follows a proximal algorithm with large step size of order  $k$ . At this point, we now assume that  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a general convex lower semicontinuous function. Replacing the gradient by the subdifferential, the relation (61) becomes

$$(62) \quad \begin{aligned} & k(k-1)(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}) + \alpha k(x_{k+2} - 2x_{k+1} + x_k) \\ & + (2\alpha - 6)(x_{k+2} - x_{k+1}) + k(k-1)\xi_k = 0, \end{aligned}$$

with  $\xi_k \in \partial f(y_{k+1})$ . The variable  $(x_k)$  is obtained by averaging the variable  $(y_k)$ , as described below

$$x_k = \frac{1}{k} \left( x_1 + \sum_{i=1}^{k-1} y_i \right).$$

The previous results have been summarized in the following table.

<p>(TOPA): Third Order Proximal Algorithm</p> <hr style="border: 0.5px solid black;"/> <p><i>Initial data</i> : <math>y_0 \in \mathcal{H}</math>, <math>y_1 \in \mathcal{H}</math>, <math>x_1 \in \mathcal{H}</math></p> <p>For <math>k = 1, 2, \dots</math></p> $\begin{cases} y_{k+1} = \text{prox}_{\frac{k(k-1)}{k+\alpha-3}f} \left( y_k + \frac{k}{k+\alpha-3}(y_k - y_{k-1}) \right) \\ x_{k+1} = \frac{1}{k+1} \left( x_1 + \sum_{i=1}^k y_i \right). \end{cases}$
--

Note that the sequence of iterates  $(y_k)$  is governed by a large step proximal inertial algorithm, while  $(x_k)$  results from the average of  $(y_k)$ . We will successively study the convergence rate of the values, then the convergence of the iterates.

**6.1. Convergence rate of the values.** The following result will follow from a Lyapunov analysis which is based on the decreasing property of the sequence  $(E_k)$ .

**Theorem 6.1.** *Let  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex lower semicontinuous proper function such that  $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$ . Let  $(x_k)$  be a sequence generated by the algorithm (TOPA). Suppose that  $\alpha > 7$ . Then, as  $k \rightarrow +\infty$*

$$f\left((k+1)x_{k+1} - kx_k\right) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^3}\right),$$

and

$$f(x_k) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k}\right).$$

In addition, the sequence  $(x_k)$  satisfies a quasi-descent property: there exists a positive constant  $C$  such that for all  $k \geq 1$

$$f(x_{k+1}) \leq f(x_k) + \frac{C}{k^4}.$$

*Proof.* We first compute

$$\begin{aligned} v_{k+1} - v_k &= (\alpha - 4)(y_{k+1} - y_k) + (k+1)(y_{k+1} - y_k) - k(y_k - y_{k-1}) \\ &= (\alpha - 3)(y_{k+1} - y_k) + k\left((y_{k+1} - y_k) - (y_k - y_{k-1})\right) \\ &= (\alpha - 3)(a_{k+2} - 2a_{k+1} + a_k) + k(y_{k+1} - 2y_k + y_{k-1}) \\ &= (\alpha - 3)(a_{k+2} - 2a_{k+1} + a_k) + k(a_{k+2} - 3a_{k+1} + 3a_k - a_{k-1}) \\ &= (\alpha - 3)\left((k+2)x_{k+2} - 2(k+1)x_{k+1} + kx_k\right) \\ &\quad + k\left((k+2)x_{k+2} - 3(k+1)x_{k+1} + 3kx_k - (k-1)x_{k-1}\right) \\ &= k^2\left(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}\right) \\ &\quad + k\left((\alpha - 3)(x_{k+2} - 2x_{k+1} + x_k) + (2x_{k+2} - 3x_{k+1} + x_{k-1})\right) \\ &\quad + (\alpha - 3)(2x_{k+2} - 2x_{k+1}) \\ &= k^2\left(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}\right) \\ &\quad + k\left(\alpha(x_{k+2} - 2x_{k+1} + x_k) - 3(x_{k+2} - 2x_{k+1} + x_k) + (2x_{k+2} - 3x_{k+1} + x_{k-1})\right) \\ &\quad + (\alpha - 3)(2x_{k+2} - 2x_{k+1}) \\ &= k^2\left(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}\right) \\ &\quad + k\left(\alpha(x_{k+2} - 2x_{k+1} + x_k) - (x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1})\right) + (2\alpha - 6)(x_{k+2} - x_{k+1}) \\ &= (k^2 - k)\left(x_{k+2} - 3x_{k+1} + 3x_k - x_{k-1}\right) + k\alpha(x_{k+2} - 2x_{k+1} + x_k) \\ &\quad + (2\alpha - 6)(x_{k+2} - x_{k+1}) \\ &= -k(k-1)\xi_k. \end{aligned}$$

From this we deduce that

$$\begin{aligned} \langle v_{k+1} - v_k, v_{k+1} \rangle &= -k(k-1)\langle \xi_k, (\alpha - 4)(y_{k+1} - z) + (k+1)(y_{k+1} - y_k) \rangle \\ (63) \quad &\leq +k(k-1)(k+1)(f(y_k) - f(y_{k+1})) - k(k-1)(\alpha - 4)(f(y_{k+1}) - f(z)), \end{aligned}$$

where the last inequality follows from the convexity of  $f$ , and  $\xi_k \in \partial f(y_{k+1})$ .

By definition of  $E_k$ , we have

$$\begin{aligned} E_{k+1} - E_k &= k(k-1)(k+1)(f(y_{k+1}) - f(y_k)) + 3k(k+1)(f(y_{k+1}) - f(z)) + \frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2 \\ &\leq k(k-1)(k+1)(f(y_{k+1}) - f(y_k)) + 3k(k+1)(f(y_{k+1}) - f(z)) + \langle v_{k+1} - v_k, v_{k+1} \rangle \\ &\leq k(k-1)(k+1)(f(y_{k+1}) - f(y_k)) + 3k(k+1)(f(y_{k+1}) - f(z)) \\ &\quad + k(k-1)(k+1)(f(y_k) - f(y_{k+1})) - k(k-1)(\alpha-4)(f(y_{k+1}) - f(z)) \end{aligned}$$

where the above inequality follows from (63). After simplification, we get

$$E_{k+1} - E_k + \left( (\alpha-4)k(k-1) - 3k(k+1) \right) (f(y_{k+1}) - \inf_{\mathcal{H}} f) \leq 0.$$

We have  $(\alpha-4)k(k-1) - 3k(k+1) = (\alpha-7)k^2 - (\alpha-1)k$ . Therefore, for  $\alpha > 7$ , and  $k \geq \frac{\alpha-1}{\alpha-7}$ , we have  $E_{k+1} \leq E_k$ . So, for  $k \geq k_1 = 1 + \lceil \frac{\alpha-1}{\alpha-7} \rceil$ , we have

$$f(x_{k+1} + k(x_{k+1} - x_k)) - \inf_{\mathcal{H}} f \leq \frac{E_{k_1}}{k(k-1)(k+1)}$$

which gives the claim. Let us now consider the sequence  $(x_k)$ . From

$$y_k = (k+1)x_{k+1} - kx_k,$$

we get

$$x_{k+1} = \frac{1}{k+1}y_k + \frac{k}{k+1}x_k.$$

According to the convexity of  $f$ , we deduce that

$$f(x_{k+1}) - \inf_{\mathcal{H}} f \leq \frac{1}{k+1}(f(y_k) - \inf_{\mathcal{H}} f) + \frac{k}{k+1}(f(x_k) - \inf_{\mathcal{H}} f).$$

Set  $F_k := f(x_k) - \inf_{\mathcal{H}} f$ . We have

$$F_{k+1} - F_k + \frac{1}{k+1}F_k \leq \frac{1}{k+1}(f(y_k) - \inf_{\mathcal{H}} f).$$

After multiplication by  $(k+1)$  we get

$$(k+1)F_{k+1} - kF_k \leq f(y_k) - \inf_{\mathcal{H}} f.$$

According to the estimate obtained for  $(y_k)$  we infer

$$(k+1)F_{k+1} - kF_k \leq \frac{C}{k^3}.$$

By summing the above inequalities, we obtain,  $\sup_k kF_k < +\infty$ , that is

$$f(x_k) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k}\right).$$

In addition,

$$f(x_{k+1}) \leq f(x_k) + \frac{C}{k^4},$$

which gives that a quasi-descent property for the sequence  $(x_k)$ .  $\square$

**6.2. Convergence of the iterates.** As a key ingredient of the following convergence analysis, we use that the sequence  $(y_k)$  is governed by the proximal algorithm

$$(64) \quad y_{k+1} = \text{prox}_{\frac{k(k-1)}{k+\alpha-3}f} \left( y_k + \frac{k}{k+\alpha-3}(y_k - y_{k-1}) \right).$$

This inertial proximal algorithm involves large proximal steps ( $\frac{k(k-1)}{k+\alpha-3} \rightarrow +\infty$  as  $k \rightarrow +\infty$ ). This last property comes from the temporal rescaling technique which underlies the algorithm. This kind of situation was first considered by Güler in [28]. Indeed, (64) is a particular case of the Inertial Proximal algorithm

$$(IP)_{\alpha_k, \beta_k} \quad y_{k+1} = \text{prox}_{\beta_k f}(y_k + \alpha_k(y_k - y_{k-1})),$$

with general parameters  $\alpha_k$  and  $\beta_k$ . A systematic study of  $(IP)_{\alpha_k, \beta_k}$  was developed by the authors in [13]. Let us recall the convergence result for  $(IP)_{\alpha_k, \beta_k}$ , obtained in [13, Theorem 7]. It makes use of the sequence  $(t_k)$  defined by

$$(65) \quad \alpha_k = \frac{t_k - 1}{t_{k+1}}.$$

**Theorem 6.2.** *Assume that*

*i) There exists  $m < 1$ , and  $k_0 \in \mathbb{N}$  such that*

$$(K_{1, \alpha_k, \beta_k}^+) \quad t_{k+1}^2 \beta_k - t_k^2 \beta_{k-1} \leq m t_{k+1} \beta_k \quad \text{for every } k \geq k_0.$$

*ii)  $\alpha_k \in [0, 1]$  for every  $k \geq 1$ ;*

*iii) The sequence  $(\beta_k)$  is non-decreasing;*

*iv)  $\sum_{k=1}^{\infty} t_k \beta_{k-1} = +\infty$ .*

*Then, any sequence  $(y_k)$  generated by the algorithm  $(IP)_{\alpha_k, \beta_k}$  converges weakly in  $\mathcal{H}$ , and its limit belongs to  $\text{argmin} f$ .*

By particularizing this general result to our setting, we obtain the following theorem.

**Theorem 6.3.** *Let  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex lower semicontinuous proper function such that  $\text{argmin}_{\mathcal{H}} f \neq \emptyset$ . Let  $(x_k)$  be a sequence generated by the algorithm (TOPA). Suppose that  $\alpha > 7$ . Then, as  $k \rightarrow +\infty$ , the sequence  $(x_k)$  converge weakly in  $\mathcal{H}$ , and its limit belongs to  $\text{argmin}_{\mathcal{H}} f$ . In addition, the sequence  $(y_k)$  converges weakly to the same limit.*

*Proof.* Let us verify that the conditions of Theorem 6.2 are satisfied. Specifically, (64) is obtained from  $(IP)_{\alpha_k, \beta_k}$  by taking the extrapolation and the proximal parameters respectively equal to

$$\alpha_k = \frac{k}{k+\alpha-3} \quad \text{and} \quad \beta_k = \frac{k(k-1)}{k+\alpha-3}.$$

First, verify that

$$(66) \quad t_k = \frac{k+\alpha-4}{\alpha-4}.$$

This is a consequence of the fact that this expression for  $t_k$  satisfies the relation (65), *i.e.*

$$\frac{t_k - 1}{t_{k+1}} = \frac{k}{k+\alpha-3}.$$

From this, we can easily verify that the conditions *ii)*, *iii)*, *iv)* are satisfied. The only non-trivial point to verify is the condition  $(K_{1, \alpha_k, \beta_k}^+)$ . Replacing  $t_k$  and  $\beta_k$  by their values, we have to show the

existence of  $m < 1$  such that (for  $k$  large enough)

$$\left(\frac{k + \alpha - 3}{\alpha - 4}\right)^2 \frac{k(k-1)}{k + \alpha - 3} - \left(\frac{k + \alpha - 4}{\alpha - 4}\right)^2 \frac{(k-1)(k-2)}{k + \alpha - 4} \leq m \frac{k + \alpha - 3}{\alpha - 4} \frac{k(k-1)}{k + \alpha - 3}.$$

After simplification, we get

$$\frac{k + \alpha - 3}{(\alpha - 4)^2} k(k-1) - \frac{k + \alpha - 4}{(\alpha - 4)^2} (k-1)(k-2) \leq m \frac{k(k-1)}{\alpha - 4}.$$

Equivalently

$$(k + \alpha - 3)k - (k + \alpha - 4)(k - 2) \leq m(\alpha - 4)k.$$

After development and simplification, we end up with

$$3k + 2(\alpha - 4) \leq m(\alpha - 4)k,$$

which is clearly satisfied for some  $m < 1$ , and  $k$  large enough, when  $\alpha > 7$ .

So, by Theorem 6.2 we obtain that the sequence  $(y_k)$  converges weakly in  $\mathcal{H}$ , and its limit belongs to  $\operatorname{argmin}_{\mathcal{H}} f$ . To pass from the convergence of  $(y_k)$  to the convergence of  $(x_k)$ , we use the second line of algorithm (IP) $_{\alpha_k, \beta_k}$

$$x_{k+1} = \frac{1}{k+1} \left( x_1 + \sum_{i=1}^k y_i \right).$$

According to the fact that convergence implies ergodic convergence to the same limit, we conclude that the sequence  $(x_k)$  converges weakly in  $\mathcal{H}$ , and this limit belongs to  $\operatorname{argmin}_{\mathcal{H}} f$ . In addition, it has the same limit as the sequence  $(y_k)$ .  $\square$

**6.3. Convergence rate of velocities and accelerations towards zero.** We will prove the following result.

**Theorem 6.4.** *Let  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex lower semicontinuous proper function such that  $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$ . Let  $(x_k)$  be a sequence generated by the algorithm (TOPA). Suppose that  $\alpha > 7$ . Then, as  $k \rightarrow +\infty$ ,*

- (i)  $\|y_k - y_{k-1}\| = o\left(\frac{1}{k}\right)$ , and  $\sum_k k \|y_k - y_{k-1}\|^2 < +\infty$ ;
- (ii)  $\|x_k - x_{k-1}\| = \mathcal{O}\left(\frac{1}{k}\right)$ ;
- (iii)  $\|x_{k+1} - 2x_k + x_{k-1}\| = \mathcal{O}\left(\frac{1}{k^2}\right)$ .

*Proof.* As in the proof of Theorem 6.3, we rely on the general results concerning the algorithm (IP) $_{\alpha_k, \beta_k}$  obtained by the authors in [13]. In [13, Theorems 5-6] (see also [11, Proposition 3.3]) it is proved that, under the assumptions of Theorem 6.2,

$$\|y_k - y_{k-1}\|^2 = o\left(\frac{1}{t_k^2}\right), \text{ and } \sum_k t_k \|y_k - y_{k-1}\|^2 < +\infty.$$

We already verified in the proof of Theorem 6.3 that, in our situation, the assumptions of Theorem 6.2 are satisfied. According to (66), the sequence  $(t_k)$  is of order  $k$ . This gives the announced convergence

rate of velocities for  $(y_k)$ . According to the definition of  $x_k$  in  $(\text{IP})_{\alpha_k, \beta_k}$ , we have

$$\begin{aligned} x_k - x_{k-1} &= \frac{1}{k} \left( x_1 + \sum_{i=1}^{k-1} y_i \right) - \frac{1}{k-1} \left( x_1 + \sum_{i=1}^{k-2} y_i \right) \\ &= -\frac{1}{k(k-1)} x_1 + \frac{1}{k} \left( \sum_{i=1}^{k-2} y_i + y_{k-1} \right) - \frac{1}{k-1} \sum_{i=1}^{k-2} y_i \\ &= -\frac{1}{k(k-1)} x_1 + \frac{1}{k} y_{k-1} - \frac{1}{k(k-1)} \sum_{i=1}^{k-2} y_i. \end{aligned}$$

Since the sequence  $(y_k)$  converges weakly, it is bounded in  $\mathcal{H}$ . The announced convergence rate towards zero of  $\|x_k - x_{k-1}\|$  follows immediately from the inequality above.

Let us now analyze the convergence rate of the accelerations to zero. According to the estimate obtained above in item (i), and  $y_k = (k+1)x_{k+1} - kx_k$ , we obtain

$$\|((k+1)x_{k+1} - kx_k) - (kx_k - (k-1)x_{k-1})\| = o\left(\frac{1}{k}\right).$$

Equivalently,

$$\|k(x_{k+1} - 2x_k + x_{k-1}) + ((x_{k+1} - x_k) + (x_k - x_{k-1}))\| = o\left(\frac{1}{k}\right).$$

By the triangle inequality, and according to the estimate obtained above in item (ii), we obtain

$$\begin{aligned} \|k(x_{k+1} - 2x_k + x_{k-1})\| &\leq o\left(\frac{1}{k^2}\right) + \frac{1}{k} \|x_{k+1} - x_k\| + \frac{1}{k} \|x_k - x_{k-1}\| \\ &\leq o\left(\frac{1}{k^2}\right) + \frac{1}{k} \mathcal{O}\left(\frac{1}{k}\right) \\ &\leq \mathcal{O}\left(\frac{1}{k^2}\right), \end{aligned}$$

which gives the announced convergence rate of the accelerations.  $\square$

## 7. CONCLUSION, PERSPECTIVES

Based on time rescaling and change of variable techniques, we introduced the third-order in time gradient-based evolution system (TOGES) with the convergence rate of the values of the order  $\mathcal{O}\left(\frac{1}{t^3}\right)$ . Then, based on a direct Lyapunov analysis of this system, we retrieved these results and enriched them by considering several related situations: the strongly convex case, the introduction of errors and/or perturbations, the introduction of the Hessian damping. For a general convex lower semicontinuous function with extended real values, we have obtained similar convergence rates for the associated proximal algorithms. Many questions remain to be explored. It would be important to analyze the convergence rate of the corresponding gradient algorithms, and of the proximal-gradient methods for structured convex optimization. It would also be interesting to deepen the understanding of the effect of average that governs the state of our dynamics, and its impact on the rate of convergence. Natural extensions would consider similar issues for nonconvex optimization and monotone inclusions.

## REFERENCES

- [1] F. Álvarez, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM J. Control Optim., **38** (4) (2000), 1102-1119.

- [2] F. Álvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., **81** (8) (2002), 747–779.
- [3] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov’s rule*, Mathematical Programming, Series A, Springer, 2018, in press.
- [4] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, A. Rondepierre, *Convergence rates of an inertial gradient descent algorithm under growth and flatness condition*, 2019.
- [5] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, **263** (2017), 5412–5458.
- [6] H. Attouch, A. Cabot, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., **28** (1) (2018), 849–874.
- [7] H. Attouch, A. Cabot, Z. Chbani, H. Riahi, *Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient*, Evolution Equations and Control Theory, **7** (3) (2018), 353–371.
- [8] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, 2019. HAL-02193846.
- [9] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. Ser. B **168** (2018), 123–175.
- [10] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case  $\alpha \leq 3$* , ESAIM: COCV, **25** (2) (2019).
- [11] H. Attouch, Z. Chbani, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim., **29** (3) (2019), 2227–2256.
- [12] H. Attouch, Z. Chbani, H. Riahi, *Fast convex optimization via time scaling of damped inertial gradient dynamics*, to appear in PAFA, <https://hal.archives-ouvertes.fr/hal-02138954>.
- [13] H. Attouch, Z. Chbani, H. Riahi, *Convergence rates of inertial proximal algorithms with general extrapolation and proximal coefficients*, HAL-02021322, accepted in Vietnam J. Math., (2019).
- [14] H. Attouch, J. Peypouquet, *Convergence rate of proximal inertial algorithms associated with Moreau envelopes of convex functions*, Splitting Algorithms, Modern Operator Theory, and Applications, eds. H. Bauschke, R. Burachik, R. Luke, Springer (2019).
- [15] H. Attouch, J. Peypouquet, *The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than  $1/k^2$* , SIAM J. Optim., **26** (3) (2016), 1824–1834.
- [16] H. Attouch, J. Peypouquet, P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, **261** (2016), 5734–5783.
- [17] J.-F. Aujol, Ch. Dossal, *Stability of over-relaxations for the Forward-Backward algorithm, application to FISTA*, SIAM J. Optim., **25** (4) (2015), 2408–2433.
- [18] J.-F. Aujol, Ch. Dossal, *Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for  $b > 0$* , 2017, <https://hal.inria.fr/hal-01547251v2>.
- [19] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CSM Books in Mathematics, Springer, 2011.
- [20] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., **2** (1) (2009), 183–202.
- [21] R. I. Bot, E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., **54** (2016), 1423–1443.
- [22] R. I. Bot, E. R. Csetnek, S. C. László *Tikhonov regularization of a second order dynamical system with Hessian driven damping*, arXiv:1911.12845v1 [math.OC] (2019).
- [23] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d’évolution*, Lecture Notes 5, North Holland, (1972).
- [24] A. Cabot, H. Engler, S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation* Transactions of the American Mathematical Society, **361** (2009), 5983–6017.
- [25] A. Chambolle, Ch. Dossal, *On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm*, Journal of Optimization Theory and Applications, **166** (2015), 968–982.
- [26] M. Ghisi, M. Gobino, A. Haraux, *The remarkable effectiveness of time-dependent damping terms for second order evolution equations*, SIAM J. Control Optim., **54** (3) (2016), 1266–1294.
- [27] O. Güler, *On the convergence of the proximal point algorithm for convex optimization*, SIAM J. Control Optim., **29** (1991), 403–419.
- [28] O. Güler, *New proximal point algorithms for convex minimization*, SIAM J. Optim., **2** (4) (1992), 649–664.
- [29] A. Haraux, *Systèmes dynamiques dissipatifs et applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris, 1991.

- [30] B. Martinet, *Brève communication. Régularisation d'inéquations variationnelles par approximations successives*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, **4** (3) (1970), 154–158.
- [31] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Mathematics, **41** (3) (2017), 681–685.
- [32] Y. Nesterov, *A method of solving a convex programming problem with convergence rate  $O(1/k^2)$* , Soviet Mathematics Doklady, **27** (1983), 372–376.
- [33] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004.
- [34] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., **73** (1967), 591–597.
- [35] N. Parikh, S. Boyd, *Proximal algorithms*, Foundations and trends in optimization, volume 1, (2013), 123–231.
- [36] J. Peypouquet, *Convex optimization in normed spaces: theory, methods and examples*, Springer, 2015.
- [37] J. Peypouquet, S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. Convex Anal, **17** (3-4) (2010), 1113–1163.
- [38] B.T. Polyak, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. Math. Phys., **4** (1964), 1–17.
- [39] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).
- [40] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., **14** (5) (1976), 877–898.
- [41] M. Schmidt, N. Le Roux, F. Bach, *Convergence rates of inexact proximal-gradient methods for convex optimization*, NIPS'11 - 25 th Annual Conference on Neural Information Processing Systems, Dec 2011, Grenada, Spain. (2011) HAL inria-00618152v3.
- [42] B. Shi, S. S. Du, M. I. Jordan, W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, arXiv:submit/2440124[cs.LG] 21 Oct 2018.
- [43] W. Siegel, *Accelerated first-order methods: Differential equations and Lyapunov functions*, arXiv:1903.05671v1 [math.OC], 2019.
- [44] W. J. Su, S. Boyd, E. J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*. Neural Information Processing Systems, **27** (2014), 2510–2518.
- [45] S. Villa, S. Salzo, L. Baldassarres, A. Verri, *Accelerated and inexact forward-backward*, SIAM J. Optim., **23** (3) (2013), 1607–1633.
- [46] A. Wibisono, A.C. Wilson, M.I. Jordan, *A variational perspective on accelerated methods in optimization*, Proceedings of the National Academy of Sciences, **113** (47) (2016), E7351–E7358.
- [47] A. C. Wilson, B. Recht, M. I. Jordan, *A Lyapunov analysis of momentum methods in optimization*. arXiv preprint arXiv:1611.02635, 2016.

IMAG, UMR 5149 CNRS, UNIVERSITÉ MONTPELLIER, 34095 MONTPELLIER CEDEX 5, FRANCE  
*Email address:* hedy.attouch@umontpellier.fr

CADI AYYAD UNIVERSITY, FACULTY OF SCIENCES SEMLALIA, MATHEMATICS, 40000 MARRAKECH, MORROCO.  
*Email address:* chbaniz@uca.ac.ma

CADI AYYAD UNIVERSITY, FACULTY OF SCIENCES SEMLALIA, MATHEMATICS, 40000 MARRAKECH, MORROCO.  
*Email address:* h-riahi@uca.ac.ma