



Fast convex optimization via a third-order in time evolution equation: TOGES-V an improved version of TOGES

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: TOGES-V an improved version of TOGES. 2020. hal-02910307

HAL Id: hal-02910307

<https://hal.science/hal-02910307>

Preprint submitted on 1 Aug 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: TOGES-V AN IMPROVED VERSION OF TOGES

HEDY ATTOUCH, ZAKI CHBANI, AND HASSAN RIAHI

ABSTRACT. In a Hilbert space setting \mathcal{H} , for convex optimization, we analyze the fast convergence properties as $t \rightarrow +\infty$ of the trajectories $t \mapsto u(t) \in \mathcal{H}$ generated by a third-order in time evolution system. The function $f : \mathcal{H} \rightarrow \mathbb{R}$ to minimize is supposed to be convex, continuously differentiable, with $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$. It enters into the dynamic through its gradient. Based on this new dynamical system, we improve the results obtained by [Attouch, Chbani, Riahi: Fast convex optimization via a third-order in time evolution equation, Optimization 2020]. As a main result, when the damping parameter α satisfies $\alpha > 3$, we show that $f(u(t)) - \inf_{\mathcal{H}} f = o(1/t^3)$ as $t \rightarrow +\infty$, as well as the convergence of the trajectories. We complement these results by introducing into the dynamic an Hessian driven damping term, which reduces the oscillations. In the case of a strongly convex function f , we show an autonomous evolution system of the third order in time with an exponential rate of convergence. All these results have natural extensions to the case of a convex lower semicontinuous function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Just replace f with its Moreau envelope.

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 \mathcal{C}^1 convex function with $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$. We take $t_0 > 0$ as the origin of time (this is justified by the singularity at the origin of the damping coefficient $\gamma(t) = \frac{\alpha}{t}$ which is used in the paper).

1. INTRODUCTION OF THE THIRD-ORDER DYNAMIC (TOGES-V)

In view of developing fast optimization methods, our study concerns the study of the convergence properties, as $t \rightarrow +\infty$, of the trajectories generated by the third-order in time evolution system

$$(\text{TOGES-V}) \quad \ddot{u}(t) + \frac{\alpha+7}{t}\dot{u}(t) + \frac{5(\alpha+1)}{t^2}u(t) + \nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) = 0.$$

As a main result, we will prove the following fast convergence result: assuming that the damping parameter satisfies $\alpha \geq 3$, then for any solution trajectory of the dynamic above

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

The above system is closely linked to the system (TOGES) introduced by the authors in [10] and that we recall below.

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

In [10, Theorem 2.1] the following result has been proved:

Suppose $\alpha \geq 3$. Then, for any trajectory of (TOGES) there exists a constant $C > 0$ such that, for

Date: July 4, 2020.

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

Key words and phrases. Accelerated gradient system; convex optimization; time rescaling; fast convergence; Lyapunov analysis.

all $t \geq t_0$

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

Note that the convergence rate of the values for $u(t)$ is only of order $1/t$, which is not completely satisfactory from the point of view of fast optimization (the classical first-order in time steepest descent does as well). By contrast, we will obtain that for any trajectory generated by (TOGES-V),

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

So, (TOGES-V) is a significant amelioration of (TOGES). As for (TOGES), the introduction of (TOGES-V) relies on time rescaling and change of variable technics. As a starting point for our study, we consider the second-order dynamic

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

introduced by Su-Boyd-Candès [34], and further studied by Attouch-Chbani-Peypouquet-Redont [12] and May [27]. The importance of this dynamic comes from the fact that the accelerated gradient method of Nesterov can be obtained as a temporal discretization by taking $\alpha = 3$. 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 α , $(\text{AVD})_{\alpha}$ for short. Let us briefly recall the convergence properties of this system:

- For $\alpha \geq 3$, each trajectory $x(\cdot)$ of $(\text{AVD})_{\alpha}$ satisfies the asymptotic convergence rate of the values $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/t^2)$ as $t \rightarrow +\infty$, see [5], [12], [27], [34].
- For $\alpha > 3$, it has been shown in [12] 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 [16] and [27] that the asymptotic convergence rate of the values is $o(1/t^2)$.

These rates are optimal, that is, they can be reached, or approached arbitrarily close. For further results concerning the system $(\text{AVD})_{\alpha}$ one can consult [2, 5, 6, 7, 12, 13, 16, 18, 19, 27, 34].

Let's make the time rescaling of $(\text{AVD})_{\alpha}$ given by

$$t = \tau(s) = s^{\frac{3}{2}}, \quad v(s) = x(\tau(s)).$$

After elementary computation, we obtain the rescaled dynamic (see [10], [8, Theorem 8.1] and [9, Corollary 3.4] for further details)

$$(1) \quad \ddot{v}(s) + \frac{3\alpha - 1}{2s}\dot{v}(s) + \frac{9}{4}s\nabla f(v(s)) = 0.$$

Since $\alpha \geq 3$ is equivalent to $\frac{3\alpha - 1}{2} \geq 4$, we obtain that, for $\alpha \geq 3$, for any solution trajectory $v(\cdot)$ of

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

we have

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

Let's go further, and make a change of the unknown function v . For all $t \geq t_0 > 0$, set

$$v(t) = \frac{1}{4t^3} \frac{d}{dt}(t^4 u(t)) = \frac{1}{4}t\dot{u}(t) + u(t).$$

Note that u is uniquely determined by v and the Cauchy data. Then

$$\dot{v}(t) = \frac{1}{4}t\ddot{u}(t) + \frac{5}{4}\dot{u}(t) \quad \text{and} \quad \ddot{v}(t) = \frac{1}{4}t\dddot{u}(t) + \frac{3}{2}\ddot{u}(t).$$

As a consequence, (1) becomes the following third-order evolution system (so doing, we have replaced $4f$ by f , which does not affect the convergence rates):

$$(\text{TOGES-V}) \quad \ddot{u}(t) + \frac{\alpha+7}{t}\ddot{u}(t) + \frac{5(\alpha+1)}{t^2}\dot{u}(t) + \nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) = 0.$$

While keeping a similar structure, this system differs notably from the system (TOGES) analyzed by the authors in [10], hence the suffix V, as Variant. But, as indicated below, these modifications of the dynamics coefficients have important consequences on the convergence rates. We take for granted the existence of global solutions in the classical sense of (TOGES-V), which is a direct consequence of the non-autonomous Cauchy-Lipschitz theorem, see for example [26, Proposition 6.2.1].

2. CONVERGENCE RESULTS

Let us state our main convergence result.

Theorem 2.1. *Let $u : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of the evolution system (TOGES-V).*

a) Suppose that $\alpha \geq 3$. Then, as $t \rightarrow +\infty$

$$(i) \quad f(u(t) + \tfrac{1}{4}t\dot{u}(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^3}\right);$$

$$(ii) \quad f(u(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^3}\right).$$

b) Suppose that $\alpha > 3$. Then, as $t \rightarrow +\infty$

$$(iii) \quad f(u(t) + \tfrac{1}{4}t\dot{u}(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right);$$

$$(iv) \quad f(u(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right);$$

(v) the trajectory converges weakly as $t \rightarrow +\infty$, let $u(t) \rightharpoonup u_\infty$, and $u_\infty \in \operatorname{argmin}_{\mathcal{H}} f$.

Proof. To obtain (i), just replace $v(t)$ by $u(t) + \frac{1}{4}t\dot{u}(t)$ in (3).

To prove (ii), let's start from the relation $\frac{d}{dt}(t^4 u(t)) = 4t^3 v(t)$. After integration from t to $t+h$ of this relation, we get

$$\begin{aligned} u(t+h) &= \left(\frac{t}{t+h}\right)^4 u(t) + \frac{1}{(t+h)^4} \int_t^{t+h} 4\tau^3 v(\tau) d\tau \\ &= \left(\frac{t}{t+h}\right)^4 u(t) + \left(1 - \left(\frac{t}{t+h}\right)^4\right) \frac{1}{(t+h)^4 - t^4} \int_t^{t+h} 4\tau^3 v(\tau) d\tau \\ &= \left(\frac{t}{t+h}\right)^4 u(t) + \left(1 - \left(\frac{t}{t+h}\right)^4\right) \frac{1}{(t+h)^4 - t^4} \int_{t^4}^{(t+h)^4} v(s^{1/4}) ds, \end{aligned}$$

where the last equality comes from the change of time variable $s = \tau^4$. According to the convexity of the function $F = f - \inf_{\mathcal{H}} f$, and using the Jensen inequality, we obtain

$$\begin{aligned} F(u(t+h)) &\leq \left(\frac{t}{t+h}\right)^4 F(u(t)) + \left(1 - \left(\frac{t}{t+h}\right)^4\right) F\left(\frac{1}{(t+h)^4 - t^4} \int_{t^4}^{(t+h)^4} v(s^{1/4}) ds\right) \\ &\leq \left(\frac{t}{t+h}\right)^4 F(u(t)) + \left(1 - \left(\frac{t}{t+h}\right)^4\right) \frac{1}{(t+h)^4 - t^4} \int_{t^4}^{(t+h)^4} F(v(s^{1/4})) ds. \end{aligned}$$

Using again the change of time variable $s = \tau^4$, we get

$$\int_{t^4}^{(t+h)^4} F(v(s^{1/4})) ds = \int_t^{t+h} 4\tau^3 F(v(\tau)) d\tau.$$

It follows

$$F(u(t+h)) \leq \left(\frac{t}{t+h}\right)^4 F(u(t)) + \left(1 - \left(\frac{t}{t+h}\right)^4\right) \frac{1}{(t+h)^4 - t^4} \int_t^{t+h} F(v(\tau)) 4\tau^3 d\tau.$$

Using the assertion (i), we obtain the existence of a constant $C > 0$ such that

$$\begin{aligned} F(u(t+h)) &\leq \left(\frac{t}{t+h}\right)^4 F(u(t)) + \left(1 - \left(\frac{t}{t+h}\right)^4\right) \frac{C}{(t+h)^4 - t^4} \int_t^{t+h} d\tau \\ (4) \quad &= \left(\frac{t}{t+h}\right)^4 F(u(t)) + \frac{Ch}{(t+h)^4}. \end{aligned}$$

Therefore,

$$\frac{1}{h} ((t+h)^4 F(u(t+h)) - t^4 F(u(t))) \leq C.$$

By letting $h \rightarrow 0^+$ in the above inequality, we get

$$\frac{d}{dt} (t^4 F(u(t))) \leq C = \frac{d}{dt} (Ct),$$

which implies that $t \mapsto t^4 F(u(t)) - Ct$ is nonincreasing. Consequently,

$$(5) \quad f(u(t)) - \inf_{\mathcal{H}} f = F(u(t)) \leq (t_0^4 F(u(t_0)) - Ct_0) \frac{1}{t^4} + \frac{C}{t^3}.$$

We conclude that $f(u(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^3}\right)$ as $t \rightarrow +\infty$.

iii) Take now $\alpha > 3$. We know that $v(t)$ converges weakly to some $v_\infty \in \operatorname{argmin}_{\mathcal{H}} f$. The relation

$$\frac{d}{dt} (t^4 u(t)) = 4t^3 v(t)$$

gives after integration

$$t^4 u(t) = t_0^4 u(t_0) + \int_{t_0}^t 4\tau^3 v(\tau) d\tau.$$

Equivalently

$$u(t) = \frac{t_0^4 u(t_0)}{t^4} + \frac{1}{t^4} \int_{t_0}^t 4\tau^3 v(\tau) d\tau.$$

Note that $\int_{t_0}^t 4\tau^3 d\tau \sim t^4$ as $t \rightarrow +\infty$. As a general rule, convergence implies ergodic convergence. Therefore, $u(t)$ converges weakly to $u_\infty = v_\infty \in \operatorname{argmin}_{\mathcal{H}} f$. Moreover we know that

$$(6) \quad f(v(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right) \text{ as } t \rightarrow +\infty,$$

which gives by a similar argument as above, as $t \rightarrow +\infty$

$$f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right) \text{ and } f(u(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right).$$

This completes the proof of Theorem 2.1. \square

Remark 2.2. Let us analyze in more detail the difference between the two dynamic systems of the third order (TOGES) and (TOGES-V). The key point is the choice of the parameter p in the relation

$$pt^{p-1}v(t) = \frac{d}{dt}(t^p u(t))$$

which connects the two variables v and u . In [10], we took $p = 1$ and thus get $v(t) = \frac{d}{dt}(tu(t))$. In this case, the convergence of the values for u is only of order $\mathcal{O}(1/t)$. In contrast in (TOGES-V), our choice is fixed at $p = 4$, which allows us to increase the speed of convergence to $\mathcal{O}(1/t^3)$. Thus, in Theorem 2.1, we improve the results of [10, Theorem 2.1]. Indeed, under similar conditions, the convergence rate of the values $f(u(t)) - \inf_{\mathcal{H}} f$ passes from $\mathcal{O}\left(\frac{1}{t}\right)$ to $\mathcal{O}\left(\frac{1}{t^3}\right)$.

3. 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 u is denoted by $\nabla^2 f(u)$. It belongs to $\mathcal{L}(\mathcal{H}, \mathcal{H})$, its action on $\xi \in \mathcal{H}$ is denoted by $\nabla^2 f(u)(\xi)$. The introduction of the Hessian driven damping into (TOGES-V)

$$(7) \quad \ddot{u}(t) + \frac{\alpha + 7}{t}\dot{u}(t) + \frac{5(\alpha + 1)}{t^2}u(t) + \nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) = 0,$$

leads us to consider the following system, called (TOGES-VH)

$$\ddot{u}(t) + \frac{\alpha + 7}{t}\dot{u}(t) + \frac{5(\alpha + 1)}{t^2}u(t) + \beta \nabla^2 f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right)\left(\frac{5}{4}\dot{u}(t) + \frac{1}{4}t\ddot{u}(t)\right) + \nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) = 0.$$

The nonnegative parameters α and β are respectively attached to the viscous damping and to the geometric damping driven by the Hessian. When $\beta = 0$, we recover the system (TOGES-V). With respect to (TOGES-V), the suffix "H" refers to Hessian. 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 [1], [11], [17], [23], [32]. In our context, we will confirm this fact on numerical examples. Note that $\nabla^2 f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right)\left(\frac{5}{4}\dot{u}(t) + \frac{1}{4}t\ddot{u}(t)\right)$ is exactly the time derivative of $\nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right)$. This plays a key role in the following developments. The following results could be obtained using arguments similar to those developed in the previous section. We choose to present an autonomous proof based on the decreasing properties of an ad hoc Lyapunov function. In doing so, we get another proof of the previous results (take $\beta = 0$), which is of independent interest. In addition, we will get explicit values of the constants that enter the convergence rates. As a main result, when $\beta > 0$, thanks to the geometric damping driven by the Hessian, we will obtain a result of rapid convergence of the gradients.

3.1. Convergence via Lyapunov analysis. The following convergence analysis is based on the decrease property of the function $t \mapsto E(t)$ defined by: given $z \in \operatorname{argmin}_{\mathcal{H}} f$

$$(8) \quad \begin{aligned} E(t) &:= 4(t^3 - 2\beta t^2) \left(f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) - \inf_{\mathcal{H}} f \right) \\ &\quad + \frac{1}{2} \left\| t^2 \left(\ddot{u}(t) + \beta \nabla f(u(t) + \frac{1}{4}t\dot{u}(t)) \right) + (\alpha + 5)t\dot{u}(t) + 4\alpha(u(t) - z) \right\|^2. \end{aligned}$$

It is convenient to work with the following condensed formulation of $E(t)$

$$(9) \quad E(t) := 4t\delta(t)(f(v(t)) - \inf_{\mathcal{H}} f) + \frac{1}{2}\|w(t)\|^2,$$

where

$$(10) \quad v(t) := u(t) + \frac{1}{4}t\dot{u}(t);$$

$$(11) \quad w(t) := t^2 \left(\ddot{u}(t) + \beta \nabla f(u(t) + \frac{1}{4}t\dot{u}(t)) \right) + (\alpha + 5)t\dot{u}(t) + 4\alpha(u(t) - z);$$

$$(12) \quad \delta(t) = t^2 \left(1 - \frac{2\beta}{t} \right).$$

Theorem 3.1. *Let $u : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of the evolution equation (TOGES-VH).*

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

$$\alpha > 3 \quad \text{and} \quad \beta \geq 0.$$

Then, for all $t \geq t_1 := \frac{2\beta(\alpha-2)}{\alpha-3}$

$$\begin{aligned} f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) - \inf_{\mathcal{H}} f &\leq \frac{(\alpha-2)E(t_1)}{4t^3}, \\ f(u(t)) - \inf_{\mathcal{H}} f &\leq (t_1^4 F(u(t_1)) - (\alpha-2)E(t_1)t_1) \frac{1}{t^4} + \frac{(\alpha-2)E(t_1)}{t^3}, \end{aligned}$$

with

$$\begin{aligned} E(t_1) &= 4(t_1^3 - 2\beta t_1^2) \left(f(u(t_1) + \frac{1}{4}t_1\dot{u}(t_1)) - \inf_{\mathcal{H}} f \right) \\ &\quad + \frac{1}{2} \|t_1^2 \left(\ddot{x}(t_1) + \beta \nabla f(u(t_1) + \frac{1}{4}t_1\dot{u}(t_1)) \right) + (\alpha + 5)t_1\dot{u}(t_1) + 4\alpha(u(t_1) - z) \|^2. \end{aligned}$$

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

The function $t \mapsto f(u(t)) + \frac{(\alpha-2)E(t_1)}{3t^3}$ is nonincreasing on $[t_1, +\infty[$.

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

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

Proof. Let us compute the time derivative of the energy function $E(\cdot)$ which is written as follows:

$$E(t) := 4t\delta(t)(f(v(t)) - \inf_{\mathcal{H}} f) + \frac{1}{2}\|w(t)\|^2.$$

According to the derivation chain rule, we get

$$(13) \quad \dot{E}(t) := 4t\delta(t)\langle \nabla f(v(t)), \dot{v}(t) \rangle + 4(\delta(t) + t\dot{\delta}(t))(f(v(t)) - \inf_{\mathcal{H}} f) + \langle w(t), \dot{w}(t) \rangle.$$

Let us compute $\dot{w}(t)$.

$$\begin{aligned}
\dot{w}(t) &= t^2 \left(\ddot{u}(t) + \beta \nabla^2 f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \left(\frac{5}{4} \dot{u}(t) + \frac{1}{4} t \ddot{u}(t) \right) \right) + 2t \left(\ddot{u}(t) + \beta \nabla f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \right) \\
&\quad + (\alpha + 5) \left(t \ddot{u}(t) + \dot{u}(t) \right) + 4\alpha \dot{u}(t) \\
&= \left(t^2 \ddot{u}(t) + (\alpha + 7) t \ddot{u}(t) + 5(\alpha + 1) \dot{u}(t) + t^2 \beta \nabla^2 f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \left(\frac{5}{4} \dot{u}(t) + \frac{1}{4} t \ddot{u}(t) \right) \right) \\
&\quad + 2t \beta \nabla f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \\
&= -t^2 \nabla f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) + 2t \beta \nabla f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) = -\delta(t) \nabla f(v(t))
\end{aligned}$$

where the two last inequalities follow from the equation (TOGES-VH) and the definition of $v(\cdot)$. Let us give an equivalent formulation of $w(t)$ using $v(t)$. This will be useful for the rest of the proof. According to the definition of $v(t)$, we have $t\ddot{u}(t) = 4\dot{v}(t) - 5\dot{u}(t)$. This gives

$$\begin{aligned}
w(t) &= t^2 \left(\ddot{u}(t) + \beta \nabla f \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \right) + (\alpha + 5) t \dot{u}(t) + 4\alpha(u(t) - z) \\
&= t(4\dot{v}(t) - 5\dot{u}(t)) + t^2 \beta \nabla f(v(t)) + (\alpha + 5) t \dot{u}(t) + 4\alpha(u(t) - z) \\
&= 4t\dot{v}(t) + t^2 \beta \nabla f(v(t)) + \alpha(t\dot{u}(t) + 4u(t)) - 4\alpha z \\
(14) \quad &= 4t\dot{v}(t) + t^2 \beta \nabla f(v(t)) + 4\alpha(v(t) - z).
\end{aligned}$$

Therefore

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

where the last inequality follows from the convexity of f . Combining (13) with (15) we obtain

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

After simplification we get

$$(17) \quad \dot{E}(t) + \beta t^2 \delta(t) \|\nabla f(v(t))\|^2 + 4 \left((\alpha - 1)\delta(t) - t\dot{\delta}(t) \right) (f(v(t)) - \inf_{\mathcal{H}} f) \leq 0.$$

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

$$\begin{aligned}
(\alpha - 1)\delta(t) - t\dot{\delta}(t) &= (\alpha - 1)(t^2 - 2\beta t) - t(2t - 2\beta) \\
&= (\alpha - 3)t^2 - 2\beta t(\alpha - 2).
\end{aligned}$$

Therefore,

$$(18) \quad \dot{E}(t) + \beta t^2 \delta(t) \|\nabla f(v(t))\|^2 + 4 \left((\alpha - 3)t^2 - 2\beta t(\alpha - 2) \right) (f(v(t)) - \inf_{\mathcal{H}} f) \leq 0.$$

For $\alpha > 3$ and $t \geq t_1 = \frac{2\beta(\alpha-2)}{\alpha-3}$ (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

$$4t^3 \left(1 - \frac{2\beta}{t} \right) (f(v(t)) - \inf_{\mathcal{H}} f) \leq E(t_1).$$

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

$$f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) - \inf_{\mathcal{H}} f \leq \frac{(\alpha-2)E(t_1)}{4t^3}.$$

Moreover by integrating (17) we obtain

$$\beta \int_{t_1}^{+\infty} t^2 \delta(t) \left\| \nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) \right\|^2 dt \leq E(t_1).$$

By a calculation similar to the one above, we obtain

$$(19) \quad \int_{t_1}^{+\infty} t^4 \left\| \nabla f\left(u(t) + \frac{1}{4}t\dot{u}(t)\right) \right\|^2 dt \leq \frac{(\alpha-2)E(t_1)}{\beta}.$$

Then, to pass from estimates on $y(t)$ to estimates on $x(t)$, we proceed as in Theorem 2.1. Precisely, define $C = (\alpha-2)E(t_1)$, and set $F = f - \inf f$. According to (i), we obtained $t^3 F(v(t)) \leq \frac{1}{4}C$.

On the basis of Jensen's inequality, we obtained in the proof of Theorem 2.1 that

$$(20) \quad \frac{d}{dt} (t^4 F(u(t))) \leq 4 \sup_t t^3 F(v(t)) \leq C = (\alpha-2)E(t_1).$$

This implies that $t \mapsto t^4 F(u(t)) - Ct$ is nonincreasing on $[t_1, +\infty[$. Consequently,

$$(21) \quad f(u(t)) - \inf_{\mathcal{H}} f \leq (t_1^4 F(u(t_1)) - Ct_1) \frac{1}{t^4} + \frac{C}{t^3}.$$

We conclude that $f(u(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^3}\right)$ as $t \rightarrow +\infty$. Let's return to (20). We have

$$t^4 \frac{d}{dt} (F(u(t))) + 4t^3 F(u(t)) \leq C.$$

Since $F(u(t)) \geq 0$, we get $\frac{d}{dt} (F(u(t))) \leq \frac{C}{t^4}$. Equivalently $\frac{d}{dt} (F(u(t)) + \frac{C}{3t^3}) \leq 0$. Therefore, $t \mapsto f(u(t)) + \frac{(\alpha-2)E(t_1)}{3t^3}$ is nonincreasing on $[t_1, +\infty[$. \square

3.2. Case $\alpha = 3$. In this case, equation (18) becomes $\dot{E}(t) \leq 8\beta t(f(v(t)) - \inf_{\mathcal{H}} f)$, which, after integration, gives

$$(22) \quad E(t) \leq E(t_0) + 8\beta \int_{t_0}^t \tau (f(v(\tau)) - \inf_{\mathcal{H}} f) d\tau.$$

Set $F(t) = f(v(t)) - \inf_{\mathcal{H}} f$. According to the definition of $E(t)$, we deduce from (22) that

$$(23) \quad 4t\delta(t)F(t) \leq E(t_0) + 8\beta \int_{t_0}^t \tau F(\tau) d\tau.$$

Let us apply the Gronwall lemma. After elementary computation, we get $F(t) \leq C/t^3$. Thus, the conclusions of Theorem 3.1 are still valid in the limiting case $\alpha = 3$.

4. NUMERICAL ILLUSTRATIONS

Let us compare the systems (TOGES), (TOGES-V), (TOGES-VH). Take $\alpha = 3$ for (TOGES) and (TOGES-V), so as to respect the condition $\alpha \geq 3$, and take $\beta = 1$ in (TOGES-VH). We get

$$(\text{TOGES}) \quad \ddot{u} + \frac{7}{t}\dot{u} + \frac{8}{t^2}\dot{u} + \nabla f(u + t\dot{u}) = 0,$$

$$(\text{TOGES-V}) \quad \ddot{u} + \frac{10}{t}\dot{u} + \frac{20}{t^2}\dot{u} + \nabla f\left(u + \frac{t}{4}\dot{u}\right) = 0,$$

$$(\text{TOGES-VH}) \quad \ddot{u} + \frac{10}{t}\dot{u} + \frac{20}{t^2}\dot{u} + \nabla f\left(u + \frac{t}{4}\dot{u}\right) + \frac{1}{4}\nabla^2 f\left(u(t) + \frac{t}{4}\dot{u}\right)(5\dot{u} + 4\ddot{u}) = 0.$$

Take $\mathcal{H} = \mathbb{R} \times \mathbb{R}$, and consider the three test functions:

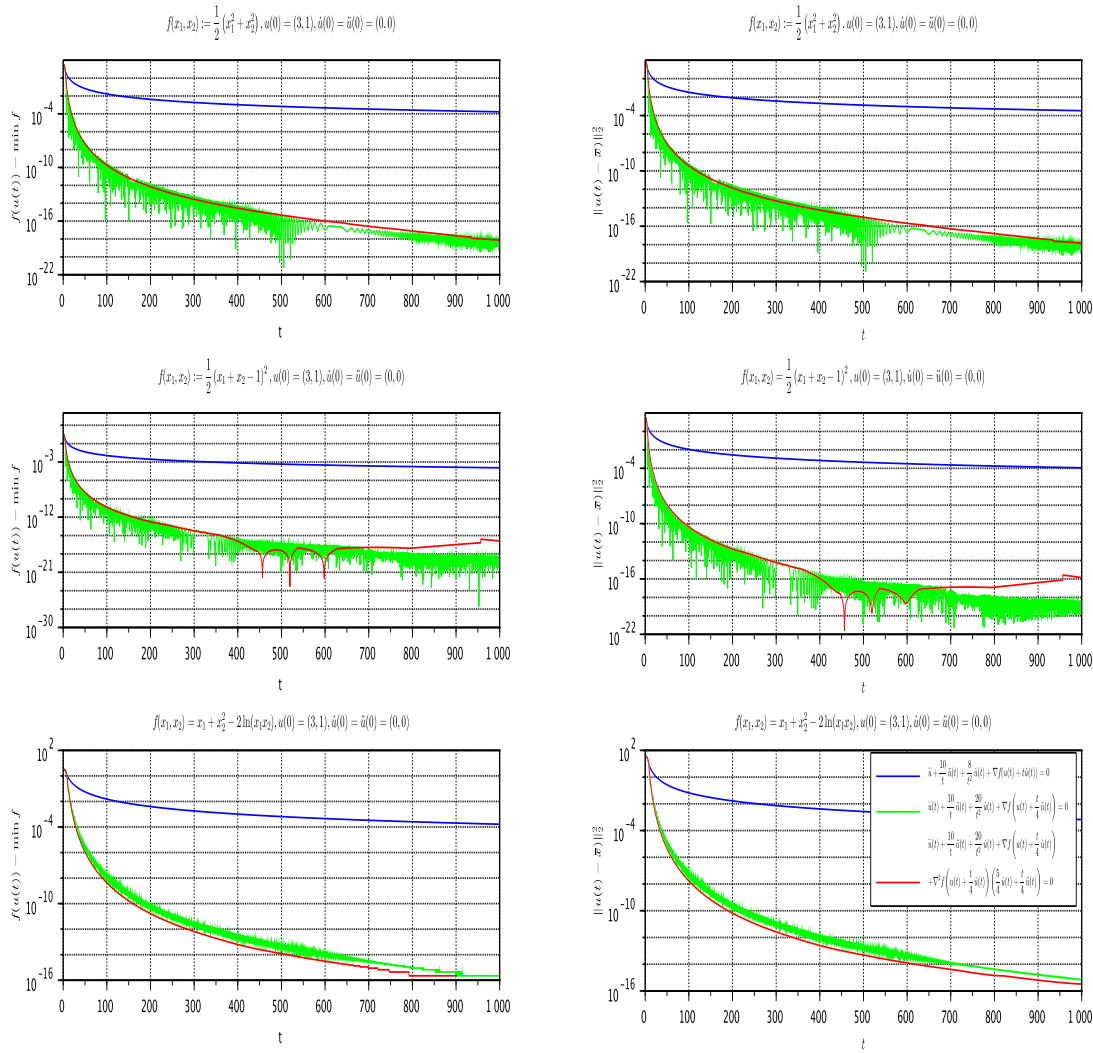


FIGURE 1. Convergence rates for (TOGES) (blue), (TOGES-V) (green), (TOGES-VH) (red)

- $f_1(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$,
- $f_2(x_1, x_2) = \frac{1}{2}(x_1 + x_2 - 1)^2$ for $(x_1, x_2) \in \mathbb{R} \times \mathbb{R}$
- $f_3(x_1, x_2) = (x_1 + x_2^2) - 2\ln(x_1 + 1)(x_2 + 1)$, for $(x, y) \in \mathbb{R}_+ \times \mathbb{R}_+$;

Given the initial conditions $u(0) = (3, 1)$, $\dot{u}(0) = \ddot{u}(0) = (0, 0)$, we have represented in Figure 1 for each function f_i ($i = 1, 2, 3$) the trajectories corresponding to (TOGES) (blue), (TOGES-V) (green) and (TOGES-VH) (red). Without ambiguity, we omit the index i for f_i . It appears clearly that, for all the convex functions proposed, (TOGES-V) is faster than (TOGES) in convergence of values. This confirms the best estimate of values $f(u(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/t^3)$ obtained in Theorem 2.1. In this numerical example, we confirm the interest of strengthening the system (TOGES-V) by adding the Hessian damping. As indicated in the first column of Figure 1, all the trajectories $u(\cdot)$ associated with (TOGES-VH) are more stable and neutralize the oscillations.

5. THE STRONGLY CONVEX CASE

When f is μ -strongly convex, we will show exponential convergence rates for the trajectories generated by the third-order autonomous evolution equation:

$$(24) \quad \ddot{u}(t) + 3\sqrt{\mu}\dot{u}(t) + 2\mu u(t) + \sqrt{\mu}\nabla f\left(u(t) + \frac{1}{\sqrt{\mu}}\dot{u}(t)\right) = 0.$$

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

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

and x^* is the unique global minimizer of f on \mathcal{H} . To condense the formulas, we write

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

Theorem 5.1. *Suppose that $f : \mathcal{H} \rightarrow \mathbb{R}$ is μ -strongly convex for some $\mu > 0$. Let $u(\cdot)$ be a solution trajectory of (24). Then, the following properties hold: For all $t \geq t_0$, we have*

$$(25) \quad f\left(u(t) + \frac{1}{\sqrt{\mu}}\dot{u}(t)\right) - \inf_{\mathcal{H}} f \leq \mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)};$$

$$(26) \quad f(u(t)) - \inf_{\mathcal{H}} f \leq \left(C\sqrt{\mu}t + e^{\sqrt{\mu}t_0}F(t_0) - C\sqrt{\mu}t_0\right)e^{-\sqrt{\mu}t};$$

$$(27) \quad \|u(t) + \frac{1}{\sqrt{\mu}}\dot{u}(t) - x^*\|^2 \leq \frac{e^{\sqrt{\mu}t_0}}{\sqrt{\mu}} \left(\sqrt{\mu}\|y(t_0) - x^*\|^2 + 2\mathcal{E}(t_0)(t - t_0)\right)e^{-\sqrt{\mu}t};$$

$$(28) \quad \|u(t) - x^*\|^2 \leq \frac{2}{\mu} \left(C\sqrt{\mu}t + e^{\sqrt{\mu}t_0}F(t_0) - C\sqrt{\mu}t_0\right)e^{-\sqrt{\mu}t}.$$

Proof. (a) Set $y(t) = u(t) + \frac{1}{\sqrt{\mu}}\dot{u}(t)$. Note that (24) can be equivalently written

$$(29) \quad \ddot{y}(t) + 2\sqrt{\mu}\dot{y}(t) + \nabla f(y(t)) = 0,$$

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

$$(30) \quad \mathcal{E}(t) = f(y(t)) - \inf_{\mathcal{H}} f + \frac{1}{2} \|\sqrt{\mu}(y(t) - x^*) + \dot{y}(t)\|^2.$$

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 (29), 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 (30) 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 this differential inequality, 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

$$(31) \quad f(y(t)) - \inf_{\mathcal{H}} f = f \left(u(t) + \frac{1}{\sqrt{\mu}}\dot{u}(t) \right) - \inf_{\mathcal{H}} f \leq \mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)},$$

and

$$\|\sqrt{\mu}(y(t) - x^*) + \dot{y}(t)\|^2 = \left\| \sqrt{\mu} \left(u(t) + \frac{1}{\sqrt{\mu}}\dot{u}(t) - x^* \right) + \dot{u}(t) + \frac{1}{\sqrt{\mu}}\ddot{u}(t) \right\|^2 \leq 2\mathcal{E}(t_0)e^{-\sqrt{\mu}(t-t_0)}.$$

(b) 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

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

(c) Let's now analyze the convergence rate of values for $f(u(t)) - \inf_{\mathcal{H}} f$. We start from (31), which can be equivalently written as follows: for all $t \geq t_0$

$$(32) \quad f(y(t)) - \inf_{\mathcal{H}} f \leq Ce^{-\sqrt{\mu}t},$$

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

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

Equivalently, $u(t+h) = e^{-\sqrt{\mu}h}u(t) + e^{-\sqrt{\mu}(t+h)} \int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau}y(\tau)d\tau$.

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

$$\begin{aligned} u(t+h) &= e^{-\sqrt{\mu}h}u(t) + e^{-\sqrt{\mu}(t+h)} \frac{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau}y(\tau)d\tau \\ (33) \quad &= e^{-\sqrt{\mu}h}u(t) + (1 - e^{-\sqrt{\mu}h}) \frac{1}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau}y(\tau)d\tau. \end{aligned}$$

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

$$(34) \quad \begin{aligned} f(u(t+h)) - \inf_{\mathcal{H}} f &\leq e^{-\sqrt{\mu}h}(f(u(t)) - \inf_{\mathcal{H}} f) \\ &+ (1 - e^{-\sqrt{\mu}h})(f - \inf_{\mathcal{H}} f) \left(\frac{1}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau}y(\tau)d\tau \right). \end{aligned}$$

Let us apply the Jensen inequality to majorize the last above expression. Set $s = e^{\sqrt{\mu}\tau}$, then

$$\int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau}y(\tau)d\tau = \int_{e^{\sqrt{\mu}t}}^{e^{\sqrt{\mu}(t+h)}} y \left(\frac{1}{\sqrt{\mu}} \ln s \right) ds$$

Let's apply Jensen's inequality. According to the convexity of $\Psi := f - \inf_{\mathcal{H}} f$, we get

$$\begin{aligned} \Psi \left(\frac{1}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau}y(\tau)d\tau \right) &= \Psi \left(\frac{1}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_{e^{\sqrt{\mu}t}}^{e^{\sqrt{\mu}(t+h)}} y \left(\frac{1}{\sqrt{\mu}} \ln s \right) ds \right) \\ &\leq \frac{1}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_{e^{\sqrt{\mu}t}}^{e^{\sqrt{\mu}(t+h)}} \Psi \left(\ln \left(\frac{1}{\sqrt{\mu}} \ln s \right) \right) ds \\ &= \frac{1}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} \sqrt{\mu}e^{\sqrt{\mu}\tau} \Psi(y(\tau))d\tau. \end{aligned}$$

Then, inequality (34) becomes

$$(35) \quad f(u(t+h)) - \inf_{\mathcal{H}} f \leq e^{-\sqrt{\mu}h}(f(u(t)) - \inf_{\mathcal{H}} f) + \frac{(1 - e^{-\sqrt{\mu}h})\sqrt{\mu}}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} e^{\sqrt{\mu}\tau}(f(y(\tau)) - \inf_{\mathcal{H}} f)d\tau.$$

According to the convergence rate of the values (32) which was obtained for y , we get

$$(36) \quad f(u(t+h)) - \inf_{\mathcal{H}} f \leq e^{-\sqrt{\mu}h}(f(u(t)) - \inf_{\mathcal{H}} f) + \frac{(1 - e^{-\sqrt{\mu}h})\sqrt{\mu}}{e^{\sqrt{\mu}(t+h)} - e^{\sqrt{\mu}t}} \int_t^{t+h} e^{\sqrt{\mu}\tau}Ce^{-\sqrt{\mu}\tau}d\tau.$$

Set $F(t) := f(u(t)) - \inf_{\mathcal{H}} f$. We can rewrite equivalently (36) as

$$F(t+h) - F(t) + (1 - e^{-\sqrt{\mu}h})F(t) \leq Ch\sqrt{\mu}e^{-\sqrt{\mu}t} \frac{1 - e^{-\sqrt{\mu}h}}{e^h - 1}.$$

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

$$(37) \quad F'(t) + \sqrt{\mu}F(t) \leq C\sqrt{\mu}e^{-\sqrt{\mu}t}.$$

Equivalently, $\frac{d}{dt}(e^{\sqrt{\mu}t}F(t) - C\sqrt{\mu}t) = e^{\sqrt{\mu}t}(F'(t) + F(t)) - C\sqrt{\mu} \leq 0$. Therefore, the function $t \mapsto e^{\sqrt{\mu}t}F(t) - C\sqrt{\mu}t$ is nonincreasing, which gives

$$e^{\sqrt{\mu}t}F(t) - C\sqrt{\mu}t \leq C_0 := e^{\sqrt{\mu}t_0}F(t_0) - C\sqrt{\mu}t_0.$$

We conclude that

$$f(u(t)) - \inf_{\mathcal{H}} f \leq (C\sqrt{\mu}t + C_0)e^{-\sqrt{\mu}t}.$$

(d) Relations (27), (28) follow immediately from (25), (26) and strong convexity of f . \square

5.2. Numerical illustration. The following numerical example illustrates the linear convergence rate of the third-order evolution equation (24) in the case of the strongly convex function $f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$. As we can see in the first table of Figure 2, in the case of strongly convex functions, the new system (24) (in green color) is more efficient than the two systems (TOGES-V) and (TOGES-VH). However, (TOGES-V) and (TOGES-VH) are quite fast and have the advantage that their rate of convergence is valid for any convex function. Linear convergence for the system (24) is illustrated in the second table of Figure 2, where we obtain the estimate $f(u(t)) - \min f \leq 10^{-20} \exp(-t^{0.47})$ for $100 \leq t_0 \leq t \leq 10^4$. The two systems (TOGES-V) and (TOGES-VH) give the following speed of convergence of values $f(u(t)) - \min f \leq 10^4 t^{-6}$. Finally, note that (24) behaves very similar to Polyak's accelerated dynamics for strongly convex functions (in pink color).

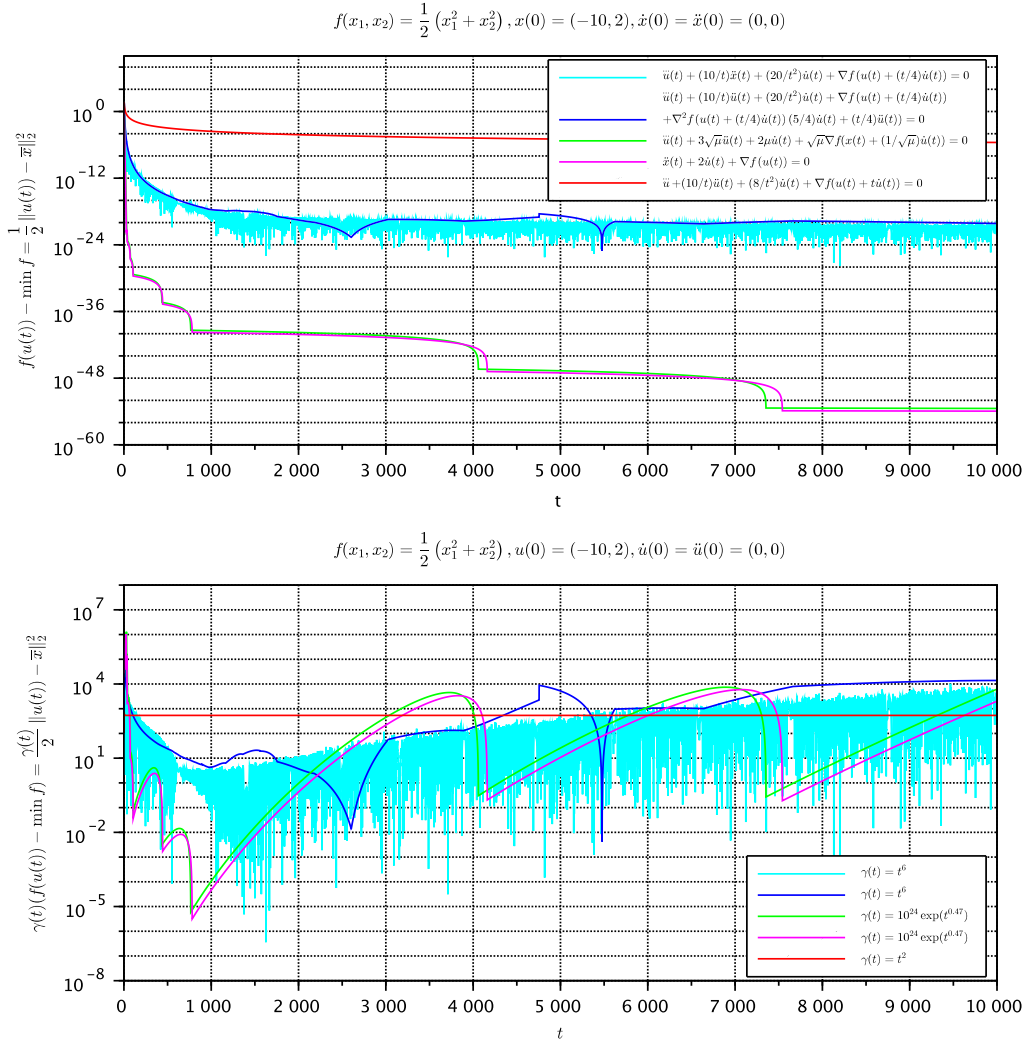


FIGURE 2. Speed of convergence of the values and trajectories associated with different systems for the strongly convex function $f(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$.

6. THE NONSMOOTH CASE

In this section, we assume that $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous and proper function such that $\inf_{\mathcal{H}} f > -\infty$. To reduce to the previous situation, where $f : \mathcal{H} \rightarrow \mathbb{R}$ is a \mathcal{C}^1 function whose gradient is Lipschitz continuous, the idea is to replace in (TOGES-V) f by its Moreau envelope $f_\lambda : \mathcal{H} \rightarrow \mathbb{R}$. Recall that f_λ is defined by: for all $x \in \mathcal{H}$,

$$f_\lambda(x) = \min_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{1}{2\lambda} \|x - \xi\|^2 \right\}.$$

The function f_λ is convex, of class $\mathcal{C}^{1,1}$, and satisfies $\inf_{\mathcal{H}} f_\lambda = \inf_{\mathcal{H}} f$, $\operatorname{argmin}_{\mathcal{H}} f_\lambda = \operatorname{argmin}_{\mathcal{H}} f$. We have

$$(38) \quad f_\lambda(x) = f(\operatorname{prox}_{\lambda f} x) + \frac{1}{2\lambda} \|x - \operatorname{prox}_{\lambda f} x\|^2,$$

where $\operatorname{prox}_{\lambda f}$ is the proximal mapping of index $\lambda > 0$ of f . Equivalently $\operatorname{prox}_{\lambda f} = (I + \lambda \partial f)^{-1}$ is the resolvent of index $\lambda > 0$ of the subdifferential of f (a maximally monotone operator). One can consult [3, section 17.2.1], [20], [24] for an in-depth study of the properties of the Moreau envelope in a Hilbert framework. So, replacing f by f_λ in (TOGES-V), and taking advantage of the fact that f_λ is continuously differentiable, we get

$$(TOGES\text{-}VR) \quad \ddot{u}(t) + \frac{\alpha + 7}{t} \dot{u}(t) + \frac{5(\alpha + 1)}{t^2} u(t) + \nabla f_\lambda \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) = 0, \quad t \in [t_0, +\infty[,$$

where the suffix R stands for Regularized. Note that $\lambda > 0$ is a fixed positive parameter, which can be chosen by the user at his convenience. Indeed, we do not intend to make λ tend to zero, as this would lead to a ill-posed limit equation. Based on the properties of the Moreau envelope, a direct adaptation of Theorem 2.1 gives the following convergence results for the regularized dynamic (TOGES-VR).

Theorem 6.1. *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous and proper function such that $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$. Let us fix $\lambda > 0$. Let $u : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of the evolution system (TOGES-VR).*

a) *Suppose that $\alpha \geq 3$. Then, as $t \rightarrow +\infty$*

$$(i) \quad f_\lambda \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) - \inf_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^3} \right); \quad f \left(\operatorname{prox}_{\lambda f} \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \right) - \inf_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^3} \right).$$

$$(ii) \quad f_\lambda(u(t)) - \inf_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^3} \right); \quad f \left(\operatorname{prox}_{\lambda f} u(t) \right) - \inf_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^3} \right).$$

b) *Suppose that $\alpha > 3$. Then, as $t \rightarrow +\infty$*

$$(iii) \quad f_\lambda \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) - \inf_{\mathcal{H}} f = o \left(\frac{1}{t^3} \right); \quad f \left(\operatorname{prox}_{\lambda f} \left(u(t) + \frac{1}{4} t \dot{u}(t) \right) \right) - \inf_{\mathcal{H}} f = o \left(\frac{1}{t^3} \right).$$

$$(iv) \quad f_\lambda(u(t)) - \inf_{\mathcal{H}} f = o \left(\frac{1}{t^3} \right); \quad f \left(\operatorname{prox}_{\lambda f} u(t) \right) - \inf_{\mathcal{H}} f = o \left(\frac{1}{t^3} \right).$$

(v) *the trajectory converges weakly as $t \rightarrow +\infty$, let $u(t) \rightharpoonup u_\infty$, and $u_\infty \in \operatorname{argmin}_{\mathcal{H}} f$.*

Proof. The proof consists in replacing f by f_λ in Theorem 2.1, and in using the property

$$(39) \quad f_\lambda(x) \geq f(\operatorname{prox}_{\lambda f} x),$$

which is a direct consequence of the definition (38) of the proximal mapping. \square

Remark 6.2. The temporal discretization of (TOGES-VR), explicit or implicit, naturally leads to relaxed proximal methods, see in the case of second order dynamics [4] and [15].

7. CONCLUSION, PERSPECTIVES

From the point of view of the design of first-order rapid optimization algorithms for convex optimization, the new system (TOGES-V) offers interesting perspectives. In its continuous form, this system offers the convergence rate $f(u(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^3}\right)$ as $t \rightarrow +\infty$, as well as the convergence of trajectories. This notably improves the convergence rate $\mathcal{O}\left(\frac{1}{t^2}\right)$ which is attached to the continuous version by Su-Boyd-Candès of the accelerated gradient method of Nesterov. Since the coefficient of the gradient is fixed in the dynamic, we can expect that the explicit temporal discretization (gradient methods), as well as the implicit temporal discretization (proximal methods), always benefit from similar convergence rates. This is a topic for further research. The system (TOGES-V) is flexible, and can be combined with the Hessian damping, which results in a significant reduction in oscillations. Finally, note that the system (TOGES-V) can be extended to the case of convex lower semicontinuous with extended real values. In this case, the corresponding algorithms are relaxed proximal algorithms, another promising topic.

REFERENCES

- [1] 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.
- [2] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov’s rule*, Math. Program., **180** (2020), 137–156. <https://doi.org/10.1007/s10107-018-1350-9>
- [3] H. Attouch, G. Buttazzo, G. Michaille, *Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization*. Second Edition, MOS/SIAM Series on Optimization, MO 17, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, (2014).
- [4] H. Attouch, A. Cabot, *Convergence rate of a relaxed inertial proximal algorithm for convex minimization*, Optimization Volume 69, 2020 - Issue 6.
- [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, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim., **29** (3) (2019), 2227–2256.
- [9] 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>.
- [10] H. Attouch, Z. Chbani, H. Riahi, *Fast convex optimization via a third-order in time evolution equation*, Optimization, <https://doi.org/10.1080/02331934.2020.1764953>
- [11] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, 2019. HAL-02193846.
- [12] 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.
- [13] 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). <https://doi.org/10.1051/cocv/2017083>.
- [14] 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).
- [15] 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).
- [16] 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.
- [17] H. Attouch, J. Peypouquet, P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, **261** (2016), 5734–5783.
- [18] 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.

- [19] 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>.
- [20] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CSM Books in Mathematics, Springer, 2011.
- [21] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., **2** (1) (2009), 183–202.
- [22] R. I. Bot, E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., **54** (2016), 1423–1443.
- [23] 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).
- [24] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).
- [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] A. Haraux, *Systèmes dynamiques dissipatifs et applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris, 1991.
- [27] 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.
- [28] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$* , Soviet Mathematics Doklady, **27** (1983), 372–376.
- [29] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004.
- [30] 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.
- [31] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).
- [32] 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.
- [33] W. Siegel, *Accelerated first-order methods: Differential equations and Lyapunov functions*, arXiv:1903.05671v1 [math.OC], 2019.
- [34] 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.
- [35] 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.
- [36] 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`