FAST CONVEX OPTIMIZATION VIA CLOSED-LOOP TIME SCALING OF GRADIENT DYNAMICS [*],[**]

Hedy Attouch[1], Radu Ioan Boţ[2] and Dang-Khoa Nguyen[3]

Abstract. In a Hilbert setting, for convex differentiable optimization, we develop a general framework for adaptive accelerated gradient methods. They are based on damped inertial dynamics where the coefficients are designed in a closed-loop way. Specifically, the damping is a feedback control of the velocity, or of the gradient of the objective function. For this, we develop a closed-loop version of the time scaling and averaging technique introduced by the authors. We thus obtain autonomous inertial dynamics which involve vanishing viscous damping and implicit Hessian driven damping. By simply using the convergence rates for the continuous steepest descent and Jensen's inequality, without the need for further Lyapunov analysis, we show that the trajectories have several remarkable properties at once: they ensure fast convergence of values, fast convergence of the gradients towards zero, and they converge to optimal solutions. Our approach leads to parallel algorithmic results, that we study in the case of proximal algorithms. These are among the very first general results of this type obtained using autonomous dynamics. Since the proposed numerical methods are based on proximal techniques, the results can be extended to a broader class, specifically to the problem of minimizing a proper, lower semicontinuous, and convex function. Numerical experiments are conducted to demonstrate the efficiency of the proposed methods.

2020 Mathematics Subject Classification. 37N40, 46N10, 65B99, 65K05, 65K10, 90C25.

September 10, 2025.

1. Introduction

In a real Hilbert space $\mathcal{H}$, we develop a dynamic approach to the rapid resolution of convex optimization problems which relies on inertial dynamics whose damping is designed as a closed-loop control. We consider the minimization problem

$$\min \left\{ f(x) : x \in \mathcal{H} \right\}, \tag{1}$$

[1] IMAG, Univ. Montpellier, CNRS, Montpellier, France
e-mail: `hedy.attouch@umontpellier.fr`

[2] Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
e-mail: `radu.bot@univie.ac.at`

[3] Faculty of Mathematics and Computer Science, University of Science, Ho Chi Minh City, Vietnam
and Vietnam National University, Ho Chi Minh City, Vietnam
e-mail: `ndkhoa@hcmus.edu.vn`

where, throughout the paper, we make the following assumptions on the function f to be minimized

$$(\mathcal{A}) \begin{cases} f : \mathcal{H} \to \mathbb{R} \text{ is a convex function of class } \mathcal{C}^1; \ S = \operatorname{argmin}_{\mathcal{H}} f \neq \emptyset; \\ \nabla f \text{ is Lipschitz continuous on the bounded sets of } \mathcal{H}. \end{cases} \tag{2}$$

Our study is part of the close links between dissipative dynamical systems and optimization algorithms, the latter being obtained by temporal discretization of the continuous dynamics. Our study comes as a natural extension of the authors' previous work [4] where the technique of time scaling and averaging was used in an open-loop way, giving rise to non-autonomous damped inertial dynamics with fast convergence properties. In the present paper, we take advantage of the simplicity and flexibility of this technique to develop it in a closed-loop way. This will give rise to autonomous damped inertial dynamics with fast convergence properties. Recall that the low-resolution ODE obtained by Su, Boyd, and Candès [36] of the accelerated gradient method of Nesterov, together with the corresponding high-resolution ODE [8], [34] (which involves an additional Hessian driven damping term) are non-autonomous dynamics, the coefficient of viscous friction being of the form α/t. Our study therefore opens a new path in the field of first-order adaptive optimization methods.

1.1. Time scale and averaging: the open-loop approach

Let us briefly explain the time scaling and averaging method in the open-loop case on a model example (see [4] for more details). Then we will look at how to develop a corresponding closed-loop approach. As the basic starting dynamic, we consider the continuous steepest descent

$$(\text{SD}) \quad \dot{z}(s) + \nabla f(z(s)) = 0, \tag{3}$$

for which we have the classical convergence result

$$f(z(s)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{s}\right) \text{ as } s \to +\infty.$$

Then, we make the change of time variable $s = \tau(t)$ in (SD), where $\tau(\cdot)$ is an increasing function from $\mathbb{R}_+$ to $\mathbb{R}_{++}$, continuously differentiable, and satisfying $\lim_{t \to +\infty} \tau(t) = +\infty$. Setting $y(t) := z(\tau(t))$, we get

$$\dot{y}(t) + \dot{\tau}(t) \nabla f(y(t)) = 0. \tag{4}$$

The convergence rate becomes

$$f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\tau(t)}\right) \text{ as } t \to +\infty. \tag{5}$$

Taking $\tau(\cdot)$ which grows faster than the identity, makes the solution trajectories unchanged but travelled faster. The price to pay is that (4) is a non-autonomous dynamic in which the coefficient in front of the gradient term tends to infinity as $t \to +\infty$. This prevents from using gradient methods to discretize it. Recall that for fixed-stepsize gradient methods the step size has to be less than or equal to twice the inverse of the Lipschitz constant of the gradient. To overcome this difficulty we come with the second step of our method which is averaging. Let us attach to $y(\cdot)$ the new function $x : [t_0, +\infty[\to \mathcal{H}$ defined by

$$\dot{x}(t) + \frac{1}{\dot{\tau}(t)}(x(t) - y(t)) = 0, \tag{6}$$

with $x(t_0) = x_0$ given in $\mathcal{H}$. We shall explain further the averaging interpretation. Equivalently

$$y(t) = x(t) + \dot{\tau}(t)\dot{x}(t). \tag{7}$$

By temporal derivation of (7) we get

$$\dot{y}(t) = \dot{x}(t) + \ddot{\tau}(t)\dot{x}(t) + \dot{\tau}(t)\ddot{x}(t). \tag{8}$$

Replacing $y(t)$ and $\dot{y}(t)$ as given by (7) and (8) in (4), we get

$$\ddot{x}(t) + \frac{1 + \ddot{\tau}(t)}{\dot{\tau}(t)}\dot{x}(t) + \nabla f\Big(x(t) + \dot{\tau}(t)\dot{x}(t)\Big) = 0. \tag{9}$$

In doing so, we passed from the first-order differential equation (4) to the second-order differential equation (9), with the advantage that now the coefficient in front of the gradient is fixed. Let us now particularize the time scale $\tau(\cdot)$. Taking

$$\tau(t) = \frac{t^2}{2(\alpha - 1)}, \tag{10}$$

gives $\frac{1+\ddot{\tau}(t)}{\dot{\tau}(t)} = \frac{\alpha}{t}$, and the corresponding dynamic with implicit Hessian driven damping

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\Big(x(t) + \frac{t}{\alpha - 1}\dot{x}(t)\Big) = 0. \tag{11}$$

In this dynamic, the Hessian driven damping appears in an implicit form. This type of dynamic was initiated in [1], see also [28] for a related autonomous system in the case of a strongly convex function f. The rationale justifying the use of the term "implicit" comes from the observation that by a Taylor expansion (as $t \to +\infty$ we have $t\dot{x}(t) \to 0$ which justifies the use of Taylor expansion), we have

$$\nabla f\Big(x(t) + \frac{t}{\alpha - 1}\dot{x}(t)\Big) \approx \nabla f(x(t)) + \frac{t}{\alpha - 1}\nabla^2 f(x(t))\dot{x}(t),$$

thus making the Hessian damping appear indirectly in (11). Because of its important role in attenuating the oscillations, several recent studies have been devoted to inertial dynamics combining the asymptotic vanishing damping with the geometric Hessian-driven damping (coined sometimes Newton-type inertial dynamics); see e.g., [2, 8–11, 18, 19, 27, 34]. In turn, the corresponding algorithms, among which IGAHD enjoys several favorable properties, introduce a correction term in the Nesterov accelerated gradient method (see [30, 31]) which reduces the oscillatory aspects.

Note that in (11) the coefficient of the Hessian damping is proportional to the inverse of the viscosity damping. Thus asymptotically when the viscous damping tends towards zero, and therefore can cause many small oscillations to appear, the coefficient of the Hessian driven damping tends towards infinity, and therefore has an effective effect on the attenuation of the oscillations. This is the situation considered by Attouch, Boţ and Nguyen in [4], who obtained convergence rates comparable to those associated with the Nesterov accelerated gradient method. A major advantage of this approach is that there is no need to do a Lyapunov analysis, we only use the classical convergence rate for the continuous steepest descent. Moreover, the convergence of the trajectories is a direct consequence of the known results for the steepest descent.

In the continuous-time models considered above, the system parameters are given as explicit functions of time – for instance, terms such as $\frac{\alpha}{t}$ or $\frac{t^2}{\alpha-1}$ appear – thus giving rise to what are known as open-loop systems. By contrast, in what follows we shall introduce and study dynamical systems whose coefficients depend directly on the evolving trajectory itself. Such systems are referred to as closed-loop systems, the terminology being self-explanatory.

1.2. Closed-loop control

The idea is to exploit the time scaling and averaging method and the fact that (SD) provides several quantities which are increasing and converge to $+\infty$ as $t \to +\infty$, so which are eligible for time scaling. This will enable us to perform time scaling and averaging in a closed-loop way. Indeed, in (SD), the velocity and the norm of the gradient are monotonically decreasing to zero. So, the idea is to use their inverse for defining the time scaling. Specifically, in a first result we are going to define the derivative of the time scaling $\tau(\cdot)$ as a function of the inverse of the speed. This means acceleration of the time scaling when the speed decreases. Following this approach, we will obtain in Theorem 3.2 the following model result.

Theorem 1.1. *Suppose that $f\colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let us choose the positive parameters according to $q > 0$, $p \geq 1$, and $\gamma > 1$. Let $x\colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following system*

$$\begin{cases} \ddot{x}(t) + \dfrac{(1+\gamma)\dot{\tau}(t)^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)}\dot{x}(t) + \gamma\dfrac{\dot{\tau}(t)^2}{\tau(t)}\nabla f\left(x(t) + \dfrac{1}{\gamma}\dfrac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right) &= 0 \\[4mm] \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^{t}[\lambda(r)]^{\frac{1}{q}}\,dr\right)^q &= 0 \\[4mm] [\lambda(t)]^p\,|\dot{\tau}(t)|^{p-1}\left\|\nabla f\left(x(t) + \dfrac{1}{\gamma}\dfrac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right)\right\|^{p-1} &= 1. \end{cases} \tag{12}$$

Then, we have the fast convergence of values: as $t \to +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^{1+q-\frac{1}{p}}}\right). \tag{13}$$

Moreover, the solution trajectory $x(t)$ converges weakly as $t \to +\infty$, and its limit belongs to $S = \operatorname{argmin} f$.

Throughout this paper, convergence of the trajectory is always to be understood in the weak sense. Consequently, the expression "its limit" consistently refers to the weak limit of the trajectory.

As a special case of Theorem 1.1, take $p = 1$, $q = 2$. Then, the last equation of (12) gives $\lambda(t) \equiv 1$. According to this, the second equation of (12) gives $\tau(t) = \frac{t^2}{4}$, and we find a case with time scaling in an open-loop form. After elementary calculation, the first equation of (12) is written as

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \frac{\alpha-1}{2}\nabla f\left(x(t) + \frac{t}{\alpha-1}\dot{x}(t)\right) = 0,$$

with $\alpha = 2\gamma + 1 > 3$, and the convergence rate of the values becomes

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right). \tag{14}$$

We therefore recover the results obtained by the authors in the case of the open loop, giving the optimal convergence rates for general convex differentiable optimization. This inertial formulation may seem at first glance complicated. Indeed it is equivalent to the first-order system in time and space

$$\begin{cases} \dot{y}(t) + \dot{\tau}(t)\nabla f(y(t)) & = 0 \\ \dot{x}(t) + \gamma\dfrac{\dot{\tau}(t)}{\tau(t)}x(t) - \gamma\dfrac{\dot{\tau}(t)}{\tau(t)}y(t) & = 0 \\ \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^{t}[\lambda(r)]^{\frac{1}{q}}\,dr\right)^q & = 0 \\ [\lambda(t)]^p\,\|\dot{y}(t)\|^{p-1} & = 1, \end{cases} \tag{15}$$

whose temporal discretization provides corresponding optimization algorithms, see Theorem 5.1.

1.3. Link with the existing literature

Contrary to the rich literature that has been devoted to non-autonomous damped inertial methods and their links with the fast first-order optimization algorithms for general convex optimization (in particular the Nesterov accelerated gradient method), only a small number of papers have been devoted to these questions, based on autonomous methods. Indeed the heavy ball method of Polyak only provides the asymptotic convergence rate $1/t$ for general convex functions. The idea is therefore to see if we can mimic the fast convergence properties of the Su, Boyd, and Candès dynamic model (see [36]) of the Nesterov accelerated gradient method, using autonomous dynamics. A natural idea is to design the damping term, on which is based the optimization properties of the system, in a closed-loop way. In this direction, we can mention the following contributions.

a) Our study has a natural link with works devoted to regularized Newton methods for solving monotone inclusions (and (1) in particular). Given a general maximally monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$, to overcome the ill-posed character of the continuous Newton method, in line with [13], Attouch, Redont and Svaiter have studied in [12] the following closed-loop dynamic version of the Levenberg-Marquardt method

$$\begin{cases} v(t) \in A(x(t)) \\ \|v(t)\|^{\gamma}\dot{x}(t) + \beta\dot{v}(t) + v(t) = 0. \end{cases}$$

When $\gamma > 1$, they showed the well-posedness of the above system, and analyzed its convergence properties. When $A = \nabla f$ this system writes

$$\|\nabla f(x(t))\|^{\gamma}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0.$$

Thus, its inertial version

$$\ddot{x}(t) + \|\nabla f(x(t))\|^{\gamma}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0$$

falls within the framework of our study with the damping equal to a closed-loop control of the norm of the gradient. The techniques developed in [12] are particularly useful for studying the well-posedness of dynamics with implicit features.

b) Although significantly different, our approach has several points in common with the article by Lin and Jordan [23]. In this article, the authors study the closed-loop dynamical system

$$
\begin{cases}
\dot{y}(t) + \dot{\tau}(t)\nabla f(x(t)) &= 0 \\
\dot{x}(t) + \frac{\dot{\tau}(t)}{\tau(t)}(x(t) - y(t)) + \frac{[\dot{\tau}(t)]^2}{\tau(t)}\nabla f(x(t)) &= 0 \\
\tau(t) - \frac{1}{4}\left(\int_0^t \sqrt{\lambda(r)}dr + c\right)^2 &= 0 \\
[\lambda(t)]^p \|\nabla f(x(t))\|^{p-1} &= \theta,
\end{cases}
\tag{16}
$$

where $c > 0$ and $0 < \theta < 1$. The corresponding second-order in time damped inertial system writes as follows

$$
\ddot{x}(t) + \frac{2[\dot{\tau}(t)]^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)}\dot{x}(t) + \frac{[\dot{\tau}(t)]^2}{\tau(t)}\nabla^2 f(x(t))\dot{x}(t) + \frac{\dot{\tau}(t)(\dot{\tau}(t) + \ddot{\tau}(t))}{\tau(t)}\nabla f(x(t)) = 0.
\tag{17}
$$

In the above system, the Hessian driven damping comes in an explicit way because of the structure of the first equation which differs from the structure of the continuous steepest descent. In contrast, in our approach, the first equation is the rescaled continuous steepest descent, and the Hessian driven damping comes implicitly. Let us highlight some advantages of our approach.

• Our system is introduced in a natural way by using the time scaling and averaging method. This makes unnecessary to perform a Lyapunov analysis for the inertial system. It has already been done for the continuous steepest descent. This results in a significantly simplified mathematical analysis.

• Our dynamic model contains an additional parameter q which, when $q = 2$, gives the setting of Lin and Jordan, and which, when judiciously tuned, gives better convergence rates.

• Our approach provides the weak convergence of the trajectories to optimal solutions.

We shall return later to the precise comparison between the two systems.

c) In [3], Attouch, Boţ and Csetnek study the convergence properties of the Autonomous Damped Inertial Gradient Equation

$$
\text{(ADIGE)} \qquad \ddot{x}(t) + \mathcal{G}\Big(\dot{x}(t), \nabla f(x(t)), \nabla^2 f(x(t))\Big) + \nabla f(x(t)) = 0,
$$

where the damping term $\mathcal{G}\Big(\dot{x}(t), \nabla f(x(t)), \nabla^2 f(x(t))\Big)$ acts as a closed-loop control. They pay particular attention to the role played by the parameter $r > 1$ in the asymptotic convergence analysis of the dynamic

$$
\ddot{x}(t) + \|\dot{x}(t)\|^{r-2}\dot{x}(t) + \nabla f(x(t)) = 0.
$$

They show that the case $r = 2$ separates the weak damping ($r > 2$) from the strong damping ($r < 2$), which emphasizes the importance of this case. These questions have also been considered by Haraux and Jendoubi in [21].

d) In [35], Song, Jiang, and Ma develop an interesting technique for accelerating high-order algorithms under general Hölder continuity assumption. Their continuous-time framework reduces to an inertial system without Hessian-driven damping in the first-order setting, which has been proven to be an inaccurate surrogate. Although underlying their approach, the acceleration via time scaling, the averaging technique, and the closed-loop tuning of the coefficients are not clearly identified.

e) Recently, Maier, Castera, and Ochs proposed in [26] the following closed-loop dynamical system

$$
\ddot{x}(t) + \sqrt{E(t)}\dot{x}(t) + \nabla f(x(t)) = 0,
$$

where $E(t) = f(x(t)) - \inf_{\mathcal{H}} f + \frac{1}{2}\|\dot{x}(t)\|^2$. They established that the function values along the generated trajectory converge at a rate of $o\left(\frac{1}{t^{2-\delta}}\right)$ as $t \to +\infty$, for arbitrary $\delta > 0$. Our approach differs fundamentally from that in [26]: the coefficients in our closed-loop systems depend on the derivative of the trajectory and on the gradient of f along the trajectory, but not on the function value itself. Moreover, while we are able to prove convergence of the trajectory, this property remains unknown for the system studied in [26]. A further limitation of the latter approach, as also acknowledged by the authors, is that the continuous-time dynamics requires prior knowledge of the optimal value of the problem.

1.4. Organization of the paper

After a general presentation of the article in the introduction, we provide in Section 2 a general estimate of the time scaling for the continuous steepest descent when it is defined in a closed-loop way. This is crucial for the rest of the paper. Then we specialize these results to situations of particular interest, and examine in details the case of closed-loop systems induced respectively by velocities, and then by gradients. In Section 3, which is the main part of the paper, we develop the next important step in our approach, which is the averaging operation. This provides accelerated damped inertial dynamics that are autonomous and with fast convergence properties. Finally, in Section 4 we analyze the fast convergence properties of proximal algorithms which come naturally from the temporal discretization of the continuous dynamics.

2. CLOSED-LOOP TIME SCALING OF THE STEEPEST DESCENT

2.1. Formulation of the closed-loop time scaling

Given $t_0 \geq 0$, $q > 0$, and $p \geq 1$, the time scale function $\tau \colon [t_0, +\infty[\to \mathbb{R}_{++}$ is defined by

$$
\begin{cases}
\dot{y}(t) + \dot{\tau}(t) \nabla f(y(t)) & = 0 \\
\tau(t) - \dfrac{1}{q^q} \left(t_0 + \displaystyle\int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr \right)^q & = 0 \\
[\lambda(t)]^p [\mathcal{G}(y(t))]^{p-1} & = 1,
\end{cases}
\tag{18}
$$

where $\mathcal{G}(\cdot)$ is a given positive, continuous function that depends on the information of the trajectory $y(\cdot)$. This general formalism allows us to unify the various situations coming from different choices of the time scaling as a feedback control of the state of the system. For example, $\mathcal{G}$ may be a function of $y, \dot{y}, f(y), \nabla f(y)$ and/or any mixture combination of them. Then the function $\lambda(\cdot)$, for which we assume only positive values, is continuous and it links the coefficient of ∇f, namely $\dot{\tau}(\cdot)$, with the solution trajectory $y(\cdot)$.
As a useful result, note that for every $t \geq t_0$, it holds

$$
\begin{aligned}
\dot{\tau}(t) &= \frac{1}{q^{q-1}} \left(\int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr + t_0 \right)^{q-1} [\lambda(t)]^{\frac{1}{q}} \\
&= [\tau(t)]^{\frac{q-1}{q}} [\lambda(t)]^{\frac{1}{q}} > 0.
\end{aligned}
\tag{19}
$$

Moreover, the relations (18) allow us to cover the open-loop case. In particular, when $p = 1$ it holds $\lambda(t) = 1$ for every $t \geq t_0$. This yields for every $q > 0$

$$
\tau(t) = \left(\frac{t}{q} \right)^q.
$$

Taking further $q := 1$, then $\tau(t)$ becomes the *regular time* in variable t, namely $\tau(t) = t$ for every $t \geq t_0$. Let us specify the interpretation of (18) as a steepest descent dynamic which is rescaled in time in a closed-loop way.

Proposition 2.1. *Suppose that $f \colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $t_0 \geq 0$, $q > 0$, $p \geq 1$ and $y \colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the system* (18). *Suppose that*

$$
\lim_{s \to +\infty} \tau(s) = +\infty.
$$

Then, $y(\cdot)$ is a solution trajectory of a time rescaled continuous steepest descent (SD), *as described below: Let $s_0 = \tau(t_0)$ and $z \colon [s_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following system*

$$
\dot{z}(s) + \nabla f(z(s)) = 0.
\tag{20}
$$

Then, we have

$$
y(t) = z(\tau(t)) \qquad \forall t \geq t_0,
$$

and there exists a continuously differentiable function $\sigma \colon [s_0, +\infty[\to \mathbb{R}_{++}$ such that

$$
z(s) = y(\sigma(s)) \qquad \forall s \geq s_0.
$$

6

Proof. We already interpreted how to go from a solution trajectory $z(\cdot)$ of (SD) to the closed-loop system above via the time scaling function $\tau(\cdot)$. Let us now show the reverse direction. Let $y\colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of (18). We have that λ is continuous and positive on $[t_0, +\infty[$, therefore τ is a monotonically increasing function, hence injective. On the other hand, we have $t_0 = \tau(t_0) = \left(\frac{t_0}{q}\right)^q$. Since by assumption $\lim_{t \to +\infty} \tau(t) = +\infty$, this means τ is a continuous function whose image contains $[s_0, +\infty[$, hence surjective. Combining these premises, we have shown that τ is a bijection, which means it is invertible.

Set $\sigma \equiv \tau^{-1}$ and make the change of time variable $t := \sigma(s)$ in (18). Let us define

$$z(s) = y(\sigma(s)) = y\left(\tau^{-1}(s)\right).$$

Then by the chain rule, we have

$$\dot{z}(s) = \dot{y}\left(\tau^{-1}(s)\right) \frac{1}{\dot{\tau}(\tau^{-1}(s))} = \dot{y}(\sigma(s)) \frac{1}{\dot{\tau}(\sigma(s))}.$$

This leads to

$$\dot{z}(s) + \nabla f(z(s)) = 0.$$

In other words, $z\colon [s_0, +\infty[\to \mathcal{H}$ is a solution trajectory of (SD). $\qquad\square$

The above assertion allows us to transfer the convergence results of (SD) to some closed-loop systems.

In particular, given a time scaling function $\tau(\cdot)$ as described above, by making the change of time variable $s := \tau(t)$, we obtain the following results from Theorem 8.1 in the appendix applied to the unperturbed continuous steepest descent system.

$$\int_{t_0}^{+\infty} \frac{\tau(t)}{\dot{\tau}(t)} \|\dot{y}(t)\|^2 \, dt < +\infty, \tag{21}$$

$$\int_{t_0}^{+\infty} \tau(t)\dot{\tau}(t) \|\nabla f(y(t))\|^2 \, dt < +\infty, \tag{22}$$

$$f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\tau(t)}\right) \quad \text{as } t \to +\infty, \tag{23}$$

$$\|\nabla f(y(t))\| = o\left(\frac{1}{\tau(t)}\right) \quad \text{as } t \to +\infty. \tag{24}$$

2.2. Lower bound estimate of the time scaling $\tau(t)$

As key ingredient of our approach, the next step is to establish a lower bound for $\tau(t)$ in terms of t. This will reflect the acceleration of our dynamic via time scaling and allow us to achieve fast convergence rates. For this, we will need the following technical lemma, which can be seen as a nonlinear Grönwall result.

Lemma 2.1. *Suppose that there exists $C_0 > 0$ and $b > a \geq 0$ such that*

$$\int_{t_0}^{t} [\tau(r)]^a [\lambda(r)]^{-b} \, dr \leq C_0 < +\infty \quad \forall t \geq t_0, \tag{25}$$

where τ is defined in terms of λ as in (18), namely, $\tau(t) = \frac{1}{q^q}\left(t_0 + \int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr\right)^q$. Then, there exists $C_1 > 0$ such that

$$\tau(t) \geq C_1 (t - t_0)^{\frac{qb+1}{b-a}} \quad \forall t \geq t_0. \tag{26}$$

Proof. Let $t \geq t_0$ be fixed. By applying the Hölder inequality and (25), we get

$$\int_{t_0}^{t} [\tau(r)]^{\frac{a}{qb+1}} \, dr \leq \left(\int_{t_0}^{t} [\tau(r)]^a [\lambda(r)]^{-b} \, dr\right)^{\frac{1}{qb+1}} \left(\int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr\right)^{\frac{qb}{qb+1}}$$

$$\leq C_0^{\frac{1}{qb+1}} \left(t_0 + \int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr\right)^{\frac{qb}{qb+1}} = \left(C_0 q^{qb}\right)^{\frac{1}{qb+1}} [\tau(t)]^{\frac{b}{qb+1}}. \tag{27}$$

7

If $a = 0$, then (26) follows immediately from (27). From now on suppose that $a > 0$, so that the inequality (27) can be rewritten as

$$\int_{t_0}^{t} [\tau(r)]^{\frac{a}{qb+1}} \, dr \leq \left(C_0 q^{qb}\right)^{\frac{1}{qb+1}} \left([\tau(t)]^{\frac{a}{qb+1}}\right)^{\frac{b}{a}}. \tag{28}$$

The arguments are now adapted from [23], which is inspired by the proof of Bihari-LaSalle inequality. Let

$$C_{q,b} := \left(C_0 q^{qb}\right)^{\frac{1}{qb+1}} > 0 \qquad \text{and} \qquad A(t) := \int_{t_0}^{t} [\tau(r)]^{\frac{a}{qb+1}} \, dr \qquad \forall t \geq t_0,$$

so that (28) becomes

$$A(t) \leq C_{q,b} \left[\dot{A}(t)\right]^{\frac{b}{a}} \quad \forall t \geq t_0$$

or, equivalently,

$$C_{q,b}^{-\frac{a}{b}} \leq [A(t)]^{-\frac{a}{b}} \dot{A}(t) \quad \forall t \geq t_0.$$

Integrating from t_0 to t, we obtain

$$C_{q,b}^{-\frac{a}{b}} (t - t_0) \leq \left(1 - \frac{a}{b}\right) \left[[A(t)]^{1-\frac{a}{b}} - [A(t_0)]^{1-\frac{1}{b}}\right]$$

$$\leq [A(t)]^{1-\frac{a}{b}} \leq \left[C_{q,b} [\tau(t)]^{\frac{b}{qb+1}}\right]^{1-\frac{a}{b}}$$

$$= C_{q,b}^{\frac{b-a}{b}} [\tau(t)]^{\frac{b-a}{qb+1}},$$

where the last inequality comes from (27). Since $b > a$, the conclusion follows. $\qquad \square$

Let us now particularize our results to some model situations.

2.3. Closed-loop control of (SD) via the velocity

Theorem 2.1. *Suppose that $f \colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $q > 0$, $p \geq 1$ and $y \colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following system*

$$\begin{cases} \dot{y}(t) + \dot{\tau}(t) \nabla f(y(t)) & = 0 \\ \tau(t) - \frac{1}{q^q} \left(t_0 + \int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr\right)^q & = 0 \\ [\lambda(t)]^p \|\dot{y}(t)\|^{p-1} & = 1. \end{cases} \tag{29}$$

Then, the following statements are satisfied:

(i) (convergence of values) $\quad f(y(t)) - \inf_{\mathcal{H}} f = o\left(t^{-\left(1+q-\frac{1}{p}\right)}\right)$ *as $t \to +\infty$.*

(ii) (convergence of gradients towards zero) $\quad \|\nabla f(y(t))\| = o\left(t^{-\left(1+q-\frac{1}{p}\right)}\right)$ *as $t \to +\infty$.*

(iii) (integral estimate of the velocities) $\quad \displaystyle\int_{t_0}^{+\infty} t^{\left(1+\frac{1}{q}-\frac{1}{pq}\right)} \|\dot{y}(t)\|^{2+\frac{p-1}{pq}} \, dt < +\infty.$

(iv) The solution trajectory $y(t)$ converges weakly as $t \to +\infty$, and its limit belongs to $\mathcal{S} = \arg\min f$.

Proof. When $p = 1$, we recover the open loop case with the time scaling function $\tau(t) = \left(\frac{t}{q}\right)^q$. The result is a direct consequence of Theorem 8.1. Therefore, from now on we only consider the case $p > 1$. Recall that from (21) we have

$$\int_{t_0}^{+\infty} \frac{\tau(t)}{\dot{\tau}(t)} \|\dot{y}(t)\|^2 \, dt < +\infty. \tag{30}$$

By using successively the definition of λ, and relation (19), we obtain

$$\frac{\tau(t)}{\dot{\tau}(t)} \|\dot{y}(t)\|^2 = \frac{\tau(t)}{\dot{\tau}(t)} [\lambda(t)]^{-\frac{2p}{p-1}} = [\tau(t)]^{\frac{1}{q}} [\lambda(t)]^{-\frac{1}{q}-\frac{2p}{p-1}} \quad \forall t \geq t_0.$$

8

According to the two above results we get

$$\int_{t_0}^{+\infty} [\tau(r)]^{\frac{1}{q}} [\lambda(r)]^{-\frac{1}{q}-\frac{2p}{p-1}} \, dr < +\infty.$$

We are now in position to apply Lemma 2.1 with $p > 1$, $a := \frac{1}{q}$ and $b := \frac{1}{q} + \frac{2p}{p-1}$. We have

$$\frac{qb+1}{b-a} = \frac{2 + \frac{2pq}{p-1}}{\frac{2p}{p-1}} = \frac{p-1+pq}{p} = 1 + q - \frac{1}{p},$$

and therefore there exists some constant $C_1 > 0$ such that

$$\tau(t) \geq C_1 (t - t_0)^{1+q-\frac{1}{p}} \quad \forall t \geq t_0. \tag{31}$$

This leads to $\lim_{t \to +\infty} \tau(t) = +\infty$. Therefore, according to Proposition 2.1, we can extract the results from Theorem 8.1 and the corresponding formulas (21), (22), (23). Specifically, we obtain

(i) for the values

$$f(y(t)) - \inf_{\mathcal{H}} f \;=\; o\left(\frac{1}{\tau(t)}\right) = o\left(\frac{1}{t^{1+q-\frac{1}{p}}}\right),$$

(ii) for the gradients

$$\|\nabla f(y(t))\| \;=\; o\left(\frac{1}{\tau(t)}\right) = o\left(\frac{1}{t^{1+q-\frac{1}{p}}}\right).$$

(iii) for the velocities: we start from (30), $i.e. \int_{t_0}^{+\infty} \frac{\tau(t)}{\dot{\tau}(t)} \|\dot{y}(t)\|^2 \, dt < +\infty$, that we evaluate as follows:

$$\begin{aligned}
\frac{\tau(t)}{\dot{\tau}(t)} \|\dot{y}(t)\|^2 &= \frac{\tau(t)}{\tau(t)^{\frac{q-1}{q}} \lambda(t)^{\frac{1}{q}}} \|\dot{y}(t)\|^2 \\
&= \frac{\tau(t)^{\frac{1}{q}}}{\lambda(t)^{\frac{1}{q}}} \|\dot{y}(t)\|^2 \\
&= \tau(t)^{\frac{1}{q}} \|\dot{y}(t)\|^{2+\frac{p-1}{pq}} \quad \forall t \geq t_0.
\end{aligned}$$

According to (31) we deduce that

$$\int_{t_0}^{+\infty} t^{\left(1+\frac{1}{q}-\frac{1}{pq}\right)} \|\dot{y}(t)\|^{2+\frac{p-1}{pq}} \, dt < +\infty.$$

(iv) Let us finally examine the convergence of the solution trajectories. We know that the solution trajectory of the continuous steepest descent converges weakly when $t \to +\infty$, and its limit belong to $S = \operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$; see Theorem 8.1 in appendix. With our notation we therefore have that $z(s)$ converges weakly when $s \to +\infty$. Since $\tau(t) \to +\infty$ as $t \to +\infty$, we immediately deduce that $y(t) = z(\tau(t))$ converges weakly as $t \to +\infty$, and its limit belong to $S = \operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$. This completes the proof. $\qquad \square \qquad \square$

2.4. Closed-loop control of (SD) via the norm of gradient

We develop an analysis parallel to that of the previous section, replacing speed control with gradient control.

Theorem 2.2. *Suppose that $f: \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $q \geq \frac{1}{2}$, $p \geq 1$ and $y: [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following system*

$$\begin{cases}
\dot{y}(t) + \dot{\tau}(t) \nabla f(y(t)) & = 0 \\
\tau(t) - \frac{1}{q^q} \left(t_0 + \int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}} \, dr\right)^q & = 0 \\
[\lambda(t)]^p \|\nabla f(y(t))\|^{p-1} & = 1.
\end{cases} \tag{32}$$

Then, the following statements are satisfied:

9

(i) (convergence of values) $f\left(y\left(t\right)\right) - \inf_{\mathcal{H}} f = o\left(t^{-pq}\right)$ as $t \to +\infty$.

(ii) (convergence of gradients towards zero) $\left\|\nabla f\left(y\left(t\right)\right)\right\| = o\left(t^{-pq}\right)$ as $t \to +\infty$.

(iii) (integral estimate of the gradients) $\displaystyle\int_{t_0}^{+\infty} t^{pq\left(2-\frac{1}{q}\right)} \left\|\nabla f\left(y\left(t\right)\right)\right\|^{2+\frac{p-1}{pq}} dt < +\infty$.

(iv) The solution trajectory $y(t)$ converges weakly as $t \to +\infty$, and its limit belongs to $\mathcal{S} = \arg\min f$.

Proof. Again, we only consider the case $p > 1$. We know from (22) that

$$\int_{t_0}^{+\infty} \tau\left(t\right) \dot{\tau}\left(t\right) \left\|\nabla f\left(y\left(t\right)\right)\right\|^2 dt < +\infty.$$

By using successively the definition of λ, and the relation (19), we obtain

$$\tau\left(t\right) \dot{\tau}\left(t\right) \left\|\nabla f\left(y\left(t\right)\right)\right\|^2 = \tau\left(t\right) \dot{\tau}\left(t\right) \left[\lambda\left(t\right)\right]^{-\frac{2p}{p-1}} = \left[\tau\left(t\right)\right]^{2-\frac{1}{q}} \left[\lambda\left(t\right)\right]^{\frac{1}{q}-\frac{2p}{p-1}} \quad \forall t \geq t_0.$$

Therefore

$$\int_{t_0}^{+\infty} \left[\tau\left(t\right)\right]^{2-\frac{1}{q}} \left[\lambda\left(t\right)\right]^{\frac{1}{q}-\frac{2p}{p-1}} dt < +\infty.$$

Let us apply Lemma 2.1 with $a := 2 - \frac{1}{q}$ and $b = \frac{2p}{p-1} - \frac{1}{q}$. We have $b > a$, $a \geq 0$ for $q \geq \frac{1}{2}$, and

$$\frac{qb+1}{b-a} = \frac{\frac{2pq}{p-1}}{\frac{2p}{p-1}-2} = pq.$$

Therefore

$$\tau\left(t\right) \geq C_1 \left(t-t_0\right)^{pq} \quad \forall t \geq t_0. \tag{33}$$

This gives $\lim_{t\to+\infty} \tau\left(t\right) = +\infty$. According to Proposition 2.1, we can extract the results from Theorem 8.1 and the corresponding formulas (21), (22), (23). Specifically, we obtain

(i) for the values

$$f\left(y\left(t\right)\right) - \inf_{\mathcal{H}} f \quad = \quad o\left(\frac{1}{\tau(t)}\right) = o\left(\frac{1}{t^{pq}}\right);$$

(ii) for the gradients

$$\left\|\nabla f\left(y\left(t\right)\right)\right\| \quad = \quad o\left(\frac{1}{\tau(t)}\right) = o\left(t^{-pq}\right);$$

(iii) for the integral estimate of the gradients: we start from (30)

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

that we evaluate as follows:

$$\begin{aligned}
\tau\left(t\right) \dot{\tau}\left(t\right) \left\|\nabla f\left(y\left(t\right)\right)\right\|^2 &= \tau\left(t\right) \left[\tau\left(t\right)\right]^{\frac{q-1}{q}} \left[\lambda\left(t\right)\right]^{\frac{1}{q}} \left\|\nabla f\left(y\left(t\right)\right)\right\|^2 \\
&= \tau(t)^{2-\frac{1}{q}} \lambda(t)^{\frac{1}{q}} \left\|\nabla f\left(y\left(t\right)\right)\right\|^2 \\
&= \tau(t)^{2-\frac{1}{q}} \left\|\nabla f(y(t))\right\|^{2-\frac{p-1}{pq}} \quad \forall t \geq t_0.
\end{aligned}$$

According to (33) we deduce that

$$\int_{t_0}^{+\infty} t^{pq\left(2-\frac{1}{q}\right)} \left\|\nabla f\left(y\left(t\right)\right)\right\|^{2+\frac{p-1}{pq}} dt < +\infty.$$

(iv) The convergence of the solution trajectory follows from an argument similar to that of the previous section. This completes the proof. $\qquad\square$

Remark 2.1. a) We thus achieved our first goal which was to accelerate the convergence properties of the continuous steepest descent using closed-loop time scaling. For example, concerning the convergence rate of the values, we passed from the convergence rate $1/t$ for the steepest descent to $1/t^{\left(1+q-\frac{1}{p}\right)}$ when the closed-loop control acts on the velocity, and $1/t^{pq}$ in the case of the gradient. Clearly, by playing with

the parameters p and q we can get arbitrary fast convergence results. The same observation holds for the convergence of the gradients towards zero.

b) By introducing a time scale function $\tau(\cdot)$ which grows faster than the identity (*i.e.* $\tau(t) \geq t$) either in open-loop or closed-loop, we have thus accelerated the continuous steepest descent dynamic. The price to pay is that we no longer have an autonomous dynamic in (4), with as major drawback the fact that the coefficient in front of the gradient term tends towards infinity as $t \to +\infty$. This prevents from using gradient methods to discretize it, as for fixed-stepsize gradient methods the step size has to be less than or equal to twice the inverse of the Lipschitz constant of the gradient. To overcome this, we come with the second step of our method which is averaging.

3. Accelerated gradient systems with closed-loop control of the damping

3.1. General results concerning time scale and averaging

We will prove the following general result which puts forward a damped inertial dynamics which comes by time scale and averaging of the continuous steepest descent. Then we will specialize it and consider time scale obtained in a closed-loop way, and thus cover the two model situations.

Theorem 3.1. *Suppose that* $f \colon \mathcal{H} \to \mathbb{R}$ *satisfies* $(\mathcal{A})$. *Let* $\gamma > 1$, *and let* $\tau \colon [t_0, +\infty[\to \mathbb{R}_{++}$ *be an increasing function, continuously differentiable, such that* $\lim_{t \to +\infty} \tau(t) = +\infty$. *Let* $x \colon [t_0, +\infty[\to \mathcal{H}$ *be a solution trajectory of the following second-order differential equation*

$$\ddot{x}(t) + \frac{(1+\gamma)\dot\tau(t)^2 - \tau(t)\ddot\tau(t)}{\tau(t)\dot\tau(t)}\dot{x}(t) + \gamma\frac{\dot\tau(t)^2}{\tau(t)}\nabla f\left(x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot\tau(t)}\dot{x}(t)\right) = 0. \tag{34}$$

Then, we have the convergence rate of the values: as $t \to +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\tau(t)}\right), \tag{35}$$

and $x(t)$ *converges weakly as* $t \to +\infty$, *and its limit belongs to* $S = \arg\min f$.

Proof. For simplicity, we assume that the system has the Cauchy data $(x(t_0), \dot{x}(t_0)) = (x_0, 0) \in \mathcal{H} \times \mathcal{H}$.

a) We first prove that the trajectory $x(\cdot)$ can be seen to be obtained from the trajectory of the continuous steepest descent via time scale and averaging. We start from $y(\cdot)$ solution of

$$\dot{y}(t) + \dot\tau(t)\nabla f(y(t)) = 0. \tag{36}$$

According to the time scale analysis developed in (5) we have

$$f(y(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\tau(t)}\right) \text{ as } t \to +\infty.$$

This means there exists a positive function ε which satisfies $\lim_{t \to +\infty} \varepsilon(t) = 0$ and

$$f(y(t)) - \inf_{\mathcal{H}} f = \frac{\varepsilon(t)}{\tau(t)} \quad \forall t \geq t_0. \tag{37}$$

Let us define the time averaging process as the transformation from y to x according to the formula

$$\dot{x}(t) + \gamma\frac{\dot\tau(t)}{\tau(t)}x(t) = \gamma\frac{\dot\tau(t)}{\tau(t)}y(t), \tag{38}$$

where $\gamma > 1$. Equivalently

$$y(t) = x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot\tau(t)}\dot{x}(t). \tag{39}$$

By taking the derivative of $y(\cdot)$ we get

$$\dot{y}(t) = \dot{x}(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot\tau(t)}\ddot{x}(t) + \frac{1}{\gamma}\frac{\dot\tau(t)^2 - \tau(t)\ddot\tau(t)}{\dot\tau(t)^2}\dot{x}(t). \tag{40}$$

Replacing $\dot{y}(t)$ by this expression in the constitutive rescaled steepest descent equation (36), we get

$$\dot{x}(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot\tau(t)}\ddot{x}(t) + \frac{1}{\gamma}\frac{\dot\tau(t)^2 - \tau(t)\ddot\tau(t)}{\dot\tau(t)^2}\dot{x}(t) + \dot\tau(t)\nabla f\left(x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot\tau(t)}\dot{x}(t)\right) = 0.$$

11

Equivalently

$$\frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\ddot{x}(t) + \frac{(1+\gamma)\dot{\tau}(t)^2 - \tau(t)\ddot{\tau}(t)}{\gamma\dot{\tau}(t)^2}\dot{x}(t) + \dot{\tau}(t)\nabla f\left(x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right) = 0.$$

After multiplication by $\gamma\frac{\dot{\tau}(t)}{\tau(t)}$ we get

$$\ddot{x}(t) + \frac{(1+\gamma)\dot{\tau}(t)^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)}\dot{x}(t) + \gamma\frac{\dot{\tau}(t)^2}{\tau(t)}\nabla f\left(x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right) = 0. \tag{41}$$

b) Let us now come to the corresponding estimate of the convergence rates with $x(t)$ instead of $y(t)$. The idea is to express x as an average of y, and then conclude thanks to Jensen's inequality. Set

$$b(t) = \frac{\dot{\tau}(t)}{\tau(t)} \geq 0 \tag{42}$$

$$B(t) = \int_{t_0}^t b(u)du = \int_{t_0}^t \frac{\dot{\tau}(u)}{\tau(u)}du = \ln\left(\frac{\tau(t)}{\tau(t_0)}\right). \tag{43}$$

Therefore

$$e^{B(t)} = \frac{\tau(t)}{\tau(t_0)}. \tag{44}$$

In order to express x in terms of y, we need to integrate the first-order linear differential equation (38) which is written equivalently as follows

$$\dot{x}(t) + \gamma b(t)x(t) = \gamma b(t)y(t).$$

After multiplying by $e^{\gamma B(t)}$, we get equivalently

$$e^{\gamma B(t)}\dot{x}(t) + \gamma b(t)e^{\gamma B(t)}x(t) = \gamma b(t)e^{\gamma B(t)}y(t),$$

that is,

$$\frac{d}{dt}\left(e^{\gamma B(t)}x(t)\right) = \gamma b(t)e^{\gamma B(t)}y(t).$$

Integrating from t_0 to t we get

$$e^{\gamma B(t)}x(t) = e^{\gamma B(t_0)}x(t_0) + \gamma\int_{t_0}^t b(u)e^{\gamma B(u)}y(u)du.$$

According to $e^{\gamma B(t_0)} = e^0 = 1$ we get

$$\begin{aligned} x(t) &= e^{-\gamma B(t)}x(t_0) + \gamma e^{-\gamma B(t)}\int_{t_0}^t b(u)e^{\gamma B(u)}y(u)du \\ &= e^{-\gamma B(t)}y(t_0) + \gamma e^{-\gamma B(t)}\int_{t_0}^t b(u)e^{\gamma B(u)}y(u)du, \end{aligned} \tag{45}$$

where the last equality follows from the choice of the Cauchy data $y(t_0) = x(t_0)$ and $\dot{x}(t_0) = 0$. Then, observe that $x(t)$ can be simply written as follows

$$x(t) = \int_{t_0}^t y(u)\, d\mu_t(u), \tag{46}$$

where μ_t is the positive Radon measure on $[t_0, t]$ defined by

$$\mu_t = e^{-\gamma B(t)}\delta_{t_0} + \gamma b(u)e^{\gamma(B(u)-B(t))}du. \tag{47}$$

Precisely, in (47), δ_{t_0} is the Dirac measure at t_0, and $b(u)e^{B(u)-B(t)}du$ is the measure with density $b(u)e^{B(u)-B(t)}$ with respect to the Lebesgue measure on $[t_0, t]$. According to

$$\gamma e^{-\gamma B(t)}\int_{t_0}^t b(u)e^{\gamma B(u)}du = 1 - e^{-\gamma B(t)},$$

we have that μ_t is a positive Radon measure on $[t_0, t]$ whose total mass is equal to 1. It is therefore a probability measure, and $x(t)$ is obtained by **averaging** the trajectory $y(\cdot)$ on $[t_0, t]$ with respect to μ_t. From

there, let us show how to deduce fast convergence properties for the so defined trajectory $x(\cdot)$. According to the convexity of f, and **Jensen's inequality**, we deduce that

$$
\begin{aligned}
f\left(\int_{t_0}^t y(u)\, d\mu_t(u)\right) - \inf_{\mathcal{H}} f &= (f - \inf_{\mathcal{H}} f)\left(\int_{t_0}^t y(u) d\mu_t(u)\right) \\
&\leq \int_{t_0}^t (f(y(u)) - \inf_{\mathcal{H}} f)\, d\mu_t(u) \\
&= \int_{t_0}^t \frac{\varepsilon(u)}{\tau(u)} d\mu_t(u),
\end{aligned}
$$

where the last inequality above comes from (37). According to the definition of μ_t (see (47)) and the formulation of $x(t)$ (see (46)), we deduce that

$$
f(x(t)) - \inf_{\mathcal{H}} f \;\leq\; \frac{\varepsilon(t_0)}{\tau(t_0)} e^{-\gamma B(t)} + \gamma e^{-\gamma B(t)} \int_{t_0}^t \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du.
$$

Equivalently,

$$
\tau(t)\left(f(x(t)) - \inf_{\mathcal{H}} f\right) \leq \varepsilon(t_0)\left(\frac{\tau(t)}{\tau(t_0)}\right)^{1-\gamma} + \gamma \tau(t)\, e^{-\gamma B(t)} \int_{t_0}^t \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du. \tag{48}
$$

Since $\gamma > 1$ and $\lim_{t \to +\infty} \tau(t) = +\infty$, it holds

$$
\limsup_{t \to +\infty} \tau(t)\left(f(x(t)) - \inf_{\mathcal{H}} f\right) \leq \gamma \limsup_{t \to +\infty} \tau(t)\, e^{-\gamma B(t)} \int_{t_0}^t \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du.
$$

It is therefore enough to show that

$$
\limsup_{t \to +\infty}\left(\gamma \tau(t)\, e^{-\gamma B(t)} \int_{t_0}^t \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du\right) \leq 0.
$$

In order to prepare for integration by parts, note that

$$
\gamma b(u) e^{\gamma B(u)} = \frac{d}{du}\left(e^{\gamma B(u)}\right) \quad \text{and} \quad \frac{\dot{\tau}(u)}{[\tau(u)]^{2-\gamma}} = \frac{d}{du}\left(\frac{1}{\gamma - 1}\frac{1}{[\tau(t)]^{1-\gamma}}\right).
$$

Given an arbitrary $\eta > 0$ we consider $T_\eta > t_0$ such that $\varepsilon(u) \leq \eta$ for every $u \geq T_\eta$. Therefore, for every $t \geq T_\eta$, by integration by parts and by taking into consideration the relations (42)-(44), we get

$$
\gamma \tau(t)\, e^{-\gamma B(t)} \int_{t_0}^t \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du
$$

$$
= \gamma \tau(t)\, e^{-\gamma B(t)}\left(\int_{t_0}^{T_\eta} \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du + \int_{T_\eta}^t \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du\right)
$$

$$
\leq \tau(t)\, e^{-\gamma B(t)}\left(\gamma \int_{t_0}^{T_\eta} \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du + \eta\gamma \int_{T_\eta}^t \frac{1}{\tau(u)} b(u) e^{\gamma B(u)}\, du\right)
$$

$$
= \tau(t)\, e^{-\gamma B(t)}\left(\gamma \int_{t_0}^{T_\eta} \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du + \frac{\eta}{\tau(t)} e^{\gamma B(t)} - \frac{\eta}{\tau(T_\eta)} e^{\gamma B(T_\eta)} + \eta \int_{T_\eta}^t \frac{\dot{\tau}(u)}{[\tau(u)]^2} e^{\gamma B(u)}\, du\right)
$$

$$
= \tau(t)\, e^{-\gamma B(t)}\left(\gamma \int_{t_0}^{T_\eta} \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du + \frac{\eta}{\tau(t)} e^{\gamma B(t)} - \frac{\eta}{\tau(T_\eta)} e^{\gamma B(T_\eta)} + \frac{\eta}{[\tau(t_0)]^\gamma} \int_{T_\eta}^t \frac{\dot{\tau}(u)}{[\tau(u)]^{2-\gamma}}\, du\right)
$$

$$
= \tau(t)\, e^{-\gamma B(t)}\left(\gamma \int_{t_0}^{T_\eta} \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du + \frac{\eta}{\tau(t)} e^{\gamma B(t)} - \frac{\eta}{\tau(T_\eta)} e^{\gamma B(T_\eta)} + \frac{\eta [\tau(t_0)]^{-\gamma}}{\gamma - 1}\left(\frac{1}{[\tau(t)]^{1-\gamma}} - \frac{1}{[\tau(T_\eta)]^{1-\gamma}}\right)\right)
$$

$$
\leq \left(\gamma \int_{t_0}^{T_\eta} \frac{\varepsilon(u)}{\tau(u)} b(u) e^{\gamma B(u)}\, du\right) \tau(t)\, e^{-\gamma B(t)} + \eta + \frac{\eta}{\gamma - 1}
$$

$$
\leq C\, [\tau(t)]^{1-\gamma} + \frac{\eta\gamma}{\gamma - 1}.
$$

Since $\lim_{t\to+\infty}\tau(t)=+\infty$, and $\gamma>1$, we obtain

$$\limsup_{t\to+\infty}\left(\gamma\tau(t)e^{-\gamma B(t)}\int_{t_0}^{t}\frac{\varepsilon(u)}{\tau(u)}b(u)e^{\gamma B(u)}du\right)\leq\frac{\eta\gamma}{\gamma-1}. \tag{49}$$

This being true for every $\eta>0$, we infer

$$f(x(t))-\inf_{\mathcal{H}}f=o\left(\frac{1}{\tau(t)}\right). \tag{50}$$

c) For trajectories convergence, we take advantage of the fact that the solution trajectory $z(\cdot)$ of the continuous steepest descent converges weakly towards a solution $x_*\in S$. Since $\lim_{t\to+\infty}\tau(t)=+\infty$, this immediately implies that $y(t)=z(\tau(t))$ converges weakly to x_* as $s\to+\infty$. In other words, for each $v\in\mathcal{H}$

$$\langle y(t),v\rangle\to\langle x_*,v\rangle\ \text{ as }t\to+\infty.$$

To pass from the convergence of y to that of x, we use the interpretation of x as an average of y. The convergence then results from the general property which says that convergence entails ergodic convergence. Let us make this precise. Using again that $\lim_{t\to+\infty}\tau(t)=+\infty$, we have

$$x(t)\sim\gamma e^{-\gamma B(t)}\int_{t_0}^{t}b(u)e^{\gamma B(u)}y(u)\,du=\frac{\gamma}{[\tau(t)]^{\gamma}}\int_{t_0}^{t}\dot\tau(u)[\tau(u)]^{\gamma-1}y(u)du.$$

After elementary calculus, we just need to prove that if $a(\cdot)$ is a positive real-valued function which verifies $\lim_{u\to+\infty}a(u)=0$, then $\lim_{t\to+\infty}A(t)=0$, where

$$A(t)=\frac{\gamma}{[\tau(t)]^{\gamma}}\int_{t_0}^{t}\dot\tau(u)[\tau(u)]^{\gamma-1}a(u)du.$$

Given an arbitrary $\eta>0$, let us take T_η such that $t_0<T_\eta$ and $a(u)\leq\eta$ for $u\geq T_\eta$. For $t>T_\eta$, we have

$$\begin{aligned}A(t)&=\frac{\gamma}{[\tau(t)]^{\gamma}}\int_{t_0}^{T_\eta}\dot\tau(u)[\tau(u)]^{\gamma-1}a(u)du+\frac{\gamma}{[\tau(t)]^{\gamma}}\int_{T_\eta}^{t}\dot\tau(u)[\tau(u)]^{\gamma-1}a(u)du\\&\leq\frac{\gamma}{[\tau(t)]^{\gamma}}\int_{t_0}^{T_\eta}\dot\tau(u)[\tau(u)]^{\gamma-1}a(u)du+\eta\tau(t_0).\end{aligned}$$

Letting t converge to $+\infty$ we get

$$\limsup_{t\to+\infty}A(t)\leq\eta\tau(t_0).$$

This being true for any $\eta>0$, we infer that $\lim_{t\to+\infty}A(t)=0$, which completes the proof. $\qquad\square$

Remark 3.1. a) By taking $\gamma:=\frac{\alpha-1}{2}$ and $\tau(t):=\frac{t^2}{2(\alpha-1)}$, equation (41) becomes (see [4])

$$\ddot x(t)+\frac{\alpha}{t}\dot x(t)+\nabla f\left(x(t)+\frac{t}{\alpha-1}\dot x(t)\right)=0.$$

We have $\gamma>1\iff\alpha>3$, which is in accordance with the convergence results attached to Nesterov method.

b) The assumption of zero initial velocity is made solely to simplify the proof. For general Cauchy data $(x(t_0),\dot x(t_0))=(x_0,x_1)\in\mathcal{H}\times\mathcal{H}$, the formula (46) takes the form

$$x(t)=\int_{t_0}^{t}y(u)\,d\mu_t(u)+\xi(t),$$

where $\xi(t)=-\frac{e^{-\gamma B(t)}}{\gamma}\frac{\tau(t_0)}{\dot\tau(t_0)}x_1$ does not vanish since $x_1\neq0$. Nevertheless, the presence of this perturbation term does not affect the conclusion, as it can be handled in the same manner as in [4, Theorem 2].

3.2. **Damped inertial system via closed-loop control of the velocity**

Let us now examine the model situation where the time scaling is defined in a closed-loop way as a feedback control of the velocity. Completing this construction with the averaging process, as described as

above, we get that $(x, y)\colon [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ is a solution trajectory of the following algebraic-differential system

$$\begin{cases} \dot{y}(t) + \dot{\tau}(t)\,\nabla f\,(y(t)) & = 0 \\[2mm] \dot{x}(t) + \gamma\dfrac{\dot{\tau}(t)}{\tau(t)}x(t) - \gamma\dfrac{\dot{\tau}(t)}{\tau(t)}y(t) & = 0 \\[2mm] \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}}\,dr\right)^q & = 0 \\[2mm] [\lambda(t)]^p\,\|\dot{y}(t)\|^{p-1} & = 1. \end{cases}$$

(51)

By specializing Theorem 3.1 to this situation we get the following result.

Theorem 3.2. *Suppose that $f\colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $q > 0$, $p \geq 1$, $\gamma > 1$ and $x\colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following system*

$$\begin{cases} \ddot{x}(t) + \dfrac{(1+\gamma)\dot{\tau}(t)^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)}\dot{x}(t) + \gamma\dfrac{\dot{\tau}(t)^2}{\tau(t)}\nabla f\left(x(t) + \dfrac{1}{\gamma}\dfrac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right) & = 0 \\[3mm] \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}}\,dr\right)^q & = 0 \\[3mm] [\lambda(t)]^p\,|\dot{\tau}(t)|^{p-1}\left\|\nabla f\left(x(t) + \dfrac{1}{\gamma}\dfrac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right)\right\|^{p-1} & = 1. \end{cases}$$

(52)

Then, we have the fast convergence of values: as $t \to +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\dfrac{1}{t^{1+q-\frac{1}{p}}}\right).$$

(53)

Moreover, the solution trajectory $x(t)$ converges weakly as $t \to +\infty$, and its limit belongs to $S = \arg\min f$.

Proof. We showed in the proof of Theorem 3.1 how to pass from (51) to (52). Conversely, let $x(\cdot)$ be a solution trajectory of the damped inertial dynamic (52). Let us show that by setting

$$y(t) = \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t) + x(t),$$

then $(x, y)\colon [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ is a solution trajectory of

$$\begin{cases} \dot{y}(t) + \dot{\tau}(t)\,\nabla f\,(y(t)) & = 0 \\[2mm] \dot{x}(t) + \gamma\dfrac{\dot{\tau}(t)}{\tau(t)}x(t) - \gamma\dfrac{\dot{\tau}(t)}{\tau(t)}y(t) & = 0 \\[2mm] \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^{t} [\lambda(r)]^{\frac{1}{q}}\,dr\right)^q & = 0 \\[2mm] [\lambda(t)]^p\,\|\dot{y}(t)\|^{p-1} & = 1. \end{cases}$$

(54)

Indeed, by taking the time derivative of $y(\cdot)$, as given by the second equation of (54), we get

$$\dot{y}(t) = \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\ddot{x}(t) + \frac{1}{\gamma}\left(1 + \gamma - \frac{\tau(t)\ddot{\tau}(t)}{[\dot{\tau}(t)]^2}\right)\dot{x}(t)$$

$$= \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\left(\ddot{x}(t) + \frac{(1+\gamma)[\dot{\tau}(t)]^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)}\dot{x}(t)\right)$$

$$= -\dot{\tau}(t)\,\nabla f\left(x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right) = -\dot{\tau}(t)\,\nabla f\,(y(t)).$$

This gives the first equation in (54) and

$$[\lambda(t)]^p\,\|\dot{y}(t)\|^{p-1} = [\lambda(t)]^p\,|\dot{\tau}(t)|^{p-1}\left\|\nabla f\left(x(t) + \frac{1}{\gamma}\frac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right)\right\|^{p-1} = 1.$$

This shows the equivalence of the two systems. According to Theorem 2.1, and formula (31), there exists a constant $C_1 > 0$ such that

$$\tau(t) \geq C_1 (t - t_0)^{1 + q - \frac{1}{p}}. \tag{55}$$

Therefore $\lim_{t \to +\infty} \tau(t) = +\infty$. According to Theorem 3.1 we deduce

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^{1 + q - \frac{1}{p}}}\right), \tag{56}$$

and the convergence of the trajectory. $\qquad\square$

3.3. **Damped inertial system via closed-loop control of the gradient**

We proceed in parallel to the previous section to obtain the following result.

Theorem 3.3. *Suppose that $f \colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $q > 0$, $p \geq 1$, $\gamma > 1$, and $x \colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following system*

$$\begin{cases} \ddot{x}(t) + \dfrac{(1 + \gamma)\dot{\tau}(t)^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)} \dot{x}(t) + \gamma \dfrac{\dot{\tau}(t)^2}{\tau(t)} \nabla f\left(x(t) + \dfrac{1}{\gamma} \dfrac{\tau(t)}{\dot{\tau}(t)} \dot{x}(t)\right) & = 0 \\[2ex] \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^t [\lambda(r)]^{\frac{1}{q}}\, dr\right)^q & = 0 \\[2ex] [\lambda(t)]^p \left\|\nabla f\left(x(t) + \dfrac{1}{\gamma} \dfrac{\tau(t)}{\dot{\tau}(t)} \dot{x}(t)\right)\right\|^{p-1} & = 1. \end{cases} \tag{57}$$

Then, we have the fast convergence of values: as $t \to +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^{pq}}\right). \tag{58}$$

Moreover, the solution trajectory $x(t)$ converges weakly as $t \to +\infty$, and its limit belongs to $S = \operatorname{argmin} f$.

Proof. Let $x(\cdot)$ be a solution trajectory of the damped inertial dynamic (57). Let us show show that by setting

$$y(t) = \frac{1}{\gamma} \frac{\tau(t)}{\dot{\tau}(t)} \dot{x}(t) + x(t),$$

then $(x, y) \colon [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ is a solution trajectory of

$$\begin{cases} \dot{y}(t) + \dot{\tau}(t)\nabla f(y(t)) & = 0 \\[1.5ex] \dot{x}(t) + \gamma \dfrac{\dot{\tau}(t)}{\tau(t)} x(t) - \gamma \dfrac{\dot{\tau}(t)}{\tau(t)} y(t) & = 0 \\[1.5ex] \tau(t) - \dfrac{1}{q^q}\left(t_0 + \displaystyle\int_{t_0}^t [\lambda(r)]^{\frac{1}{q}}\, dr\right)^q & = 0 \\[1.5ex] [\lambda(t)]^p \|\nabla f(y(t))\|^{p-1} & = 1. \end{cases} \tag{59}$$

Indeed, by the same argument as for the velocity case, we get

$$\dot{y}(t) = -\dot{\tau}(t)\nabla f(y(t)).$$

This gives the first equation in (59) and

$$[\lambda(t)]^p \|\nabla f(y(t))\|^{p-1} = [\lambda(t)]^p \left\|\nabla f\left(x(t) + \frac{1}{\gamma} \frac{\tau(t)}{\dot{\tau}(t)} \dot{x}(t)\right)\right\|^{p-1} = 1.$$

This shows the equivalence of the two systems. According to Theorem 2.2, and formula (33), there exists a constant $C_1 > 0$ such that

$$\tau(t) \geq C_1 (t - t_0)^{pq}. \tag{60}$$

Therefore from Theorem 3.1 we deduce, as $t \to +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^{pq}}\right), \tag{61}$$

and the convergence of the trajectory. $\qquad\square$

3.4. Comparison with the Lin-Jordan approach

In [23], the authors study the second-order closed-loop dynamical system

$$
\begin{cases}
\ddot{x}(t) + \left(\dfrac{2\dot{\tau}(t)}{\tau(t)} - \dfrac{\ddot{\tau}(t)}{\dot{\tau}(t)} \right) \dot{x}(t) + \dfrac{[\dot{\tau}(t)]^2}{\tau(t)} \nabla^2 f(x(t)) \dot{x}(t) + \dfrac{\dot{\tau}(t)\,[\dot{\tau}(t) + \ddot{\tau}(t)]}{\tau(t)} \nabla f(x(t)) & = 0 \\[2mm]
\tau(t) - \dfrac{1}{4} \left(\int_0^t \sqrt{\lambda(t)}\,dr + c \right)^2 & = 0 \\[2mm]
[\lambda(t)]^p \, \|\nabla f(x(t))\|^{p-1} & = \theta,
\end{cases}
\tag{62}
$$

whose first-order reformulation reads

$$
\begin{cases}
\dot{y}(t) + \dot{\tau}(t)\,\nabla f(x(t)) & = 0 \\[2mm]
\dot{x}(t) + \dfrac{\dot{\tau}(t)}{\tau(t)}(x(t) - y(t)) + \dfrac{[\dot{\tau}(t)]^2}{\tau(t)} \nabla f(x(t)) & = 0 \\[2mm]
\tau(t) - \dfrac{1}{4} \left(\int_0^t \sqrt{\lambda(t)}\,dr + c \right)^2 & = 0 \\[2mm]
[\lambda(t)]^p \, \|\nabla f(x(t))\|^{p-1} & = \theta,
\end{cases}
\tag{63}
$$

where $c > 0$ and $0 < \theta < 1$ are given parameters. See also [25] and [24] for some extensions to monotone equations and monotone inclusions, respectively.

a) In [23], the authors obtained the following convergence rate of function values

$$
f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{\frac{3p+1}{2}}} \right) \quad \text{as } t \to +\infty.
$$

Note that the last two equations in (63) are nothing else than those in (32) with $q := 2$. For comparison, in our approach the convergence rate of the values obtained in Theorem 3.3 when $q = 2$ is

$$
f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t^{2p}} \right),
$$

which is better for every $p > 1$.

b) Let us now compare the convergence estimates of the gradients. In [23], the authors obtain the integral estimate

$$
\int_{t_0}^{+\infty} t^{\frac{3p+1}{2}} \, \|\nabla f(x(t))\|^{\frac{p+1}{p}} \, dt < +\infty,
$$

which leads to

$$
\inf_{t_0 \le \sigma \le t} \|\nabla f(x(\sigma))\| = \mathcal{O}\left(t^{-\frac{3p}{2}} \right) \quad \text{as } t \to +\infty.
$$

In our approach, the right variable to consider is $y(t)$, instead of $x(t)$. According to (22) we have

$$
\int_{t_0}^{+\infty} \tau(t)\,\dot{\tau}(t)\,\|\nabla f(y(t))\|^2 \, dt < +\infty.
$$

Since $q = 2$, according to (19) we have

$$
\dot{\tau}(t) = [\tau(t)]^{\frac{1}{2}} [\lambda(t)]^{\frac{1}{2}}.
$$

Therefore

$$
\tau(t)\,\dot{\tau}(t)\,\|\nabla f(y(t))\|^2 \;=\; \tau(t)^{\frac{3}{2}} [\lambda(t)]^{\frac{1}{2}} \|\nabla f(y(t))\|^2 = \tau(t)^{\frac{3}{2}} \|\nabla f(y(t))\|^{2 - \frac{p-1}{2p}}.
$$

Since $\tau(t) \ge Ct^{2p}$, we deduce that

$$
\int_{t_0}^{+\infty} t^{3p} \, \|\nabla f(y(t))\|^{\frac{3p+1}{2p}} \, dt < +\infty,
$$

which leads to

$$
\inf_{t_0 \le \sigma \le t} \|\nabla f(y(\sigma))\| = \mathcal{O}\left(t^{-2p} \right) \quad \text{as } t \to +\infty.
$$

17

Again, our approach gives a better convergence rate than [23]. Let us also specify that our analysis provides the convergence of the trajectories, which is an open question for [23]. Moreover, since our approach is consistent with the steepest continuous descent, it can naturally be extended to the non-smooth case, and to the case of cocoercive operators, as it was done in the open-loop case in [4].

3.5. The limiting case $\gamma = 1$

Our previous results are valid under the assumption $\gamma > 1$. It is a natural question to examine the limiting case $\gamma = 1$. Close examination of the proof of the theorem reveals a slight change in the integration procedure and a logarithm factor appears. The corresponding result obtained is written as follows.

Theorem 3.4. *Suppose that $f \colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $x \colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of the following second-order differential equation*

$$\ddot{x}(t) + \frac{2\left[\dot{\tau}(t)\right]^2 - \tau(t)\ddot{\tau}(t)}{\tau(t)\dot{\tau}(t)}\dot{x}(t) + \frac{\left[\dot{\tau}(t)\right]^2}{\tau(t)}\nabla f\left(x(t) + \frac{\tau(t)}{\dot{\tau}(t)}\dot{x}(t)\right) = 0 \tag{64}$$

where $\tau : [t_0, +\infty[\to \mathbb{R}_{++}$ is an increasing function, continuously differentiable, and satisfying $\lim_{t \to +\infty} \tau(t) = +\infty$.
Then, we have the convergence rate of the values: as $t \to +\infty$

$$f\left(x(t)\right) - \inf_{\mathcal{H}} f = o\left(\frac{\ln\left(\tau\left(t\right)\right)}{\tau\left(t\right)}\right), \tag{65}$$

and the solution trajectory $x(t)$ converges weakly as $t \to +\infty$, and its limits belongs to $S = \operatorname{argmin} f$.
Suppose moreover that there exists some $\theta > 0$ and $C_1 > 0$ such that for t sufficiently large

$$(\mathcal{A})_{\text{asymp}} \qquad \tau(t) \geq C_1\left(t - t_0\right)^\theta. \tag{66}$$

Then, we have the fast convergence of values: as $t \to +\infty$

$$f\left(x(t)\right) - \inf_{\mathcal{H}} f = o\left(\frac{\ln\left(t\right)}{t^\theta}\right). \tag{67}$$

When specialized to the closed-loop control of the velocity, we obtain

$$f\left(x(t)\right) - \inf_{\mathcal{H}} f = o\left(\frac{\ln\left(t\right)}{t^{1+q-\frac{1}{p}}}\right), \tag{68}$$

and in the case of the closed-loop control of the gradient

$$f\left(x(t)\right) - \inf_{\mathcal{H}} f = o\left(\frac{\ln\left(t\right)}{t^{pq}}\right). \tag{69}$$

So, the convergence rates are in this limiting case a little worse because of the logarithm term.

4. Associated proximal algorithms

4.1. A proximal-explicit discretization

In the following, we present a numerical approach based on a *proximal-explicit* temporal discretization of the closed-loop systems investigated in this paper. By proximal-explicit we mean that the function f is evaluated using a proximal step while the step size sequence $(\lambda_k)_{k \geq 0}$ and the time scaling sequence $(\tau_k)_{k \geq 0}$ are computed explicitly. This makes our numerical scheme much easier implementable than the numerical algorithm proposed in [23] as well as the large-step A HPE approach by Monteiro and Svaiter [29] which are in fact approximations of a *proximal-implicit* discrete time method. We restrict ourselves to the case $q = 1$, which gives $\dot{\tau}(t) = \lambda(t)$. In this case, the continuous time closed-loop dynamical system is written as follows

$$\begin{cases} \dot{y}(t) + \lambda(t)\nabla f\left(y(t)\right) &= 0 \\ \left[\lambda(t)\right]^p \left[\mathcal{G}\left(y(t)\right)\right]^{p-1} &= 1. \end{cases} \tag{70}$$

Let us describe the general structure of the algorithm which is obtained by a proximal-explicit discretization of the continuous system (70).

18

Given y_k, y_{k-1} in $\mathcal{H}$, we first define λ_k by

$$[\lambda_k]^p \left[\mathcal{G}\left(y_k, y_{k-1}\right)\right]^{p-1} = 1.$$

and consider then an implicit finite difference scheme for the first equation of (70)

$$y_{k+1} - y_k + \lambda_k \nabla f\left(y_{k+1}\right) = 0. \tag{71}$$

This gives the following algorithm, called (PEAS) for Proximal Explicit Algorithm with Adaptive Step Size. Note that $(\lambda_k)_{k \geq 0}$ is computed explicitly in terms of $(y_k)_{k \geq 0}$. In other words, the definition of the sequence

---

Algorithm 1: Proximal-explicit algorithm with adaptive step size (PEAS)

    Input: $y_0 \neq y_{-1} \in \mathcal{H}$
1 for $k = 0, 1, \cdots$ **do**
2 $\lambda_k := \left[\mathcal{G}\left(y_k, y_{k-1}\right)\right]^{-\frac{p-1}{p}}$
3 $y_{k+1} := \operatorname{prox}_{\lambda_k f}\left(y_k\right)$
4 end

$(\lambda_k)_{k \geq 0}$ is decoupled from the computation of $(y_k)_{k \geq 0}$. This is different from the method in [23], which ultimately leads to the large-step A HPE approach by Monteiro and Svaiter in [29].

Let us now specify the link between λ_k and τ_k. We start from the relation (recall that we take $q = 1$)

$$\dot{\tau}\left(t\right) = \lambda\left(t\right). \tag{72}$$

Then, for every $k \geq 0$ we discretize (72) as follows

$$\tau_{k+1} - \tau_k = \lambda_k \Longleftrightarrow \tau_{k+1} = \lambda_k + \tau_k \tag{73}$$

with the convention $\lambda_0 := t_0$ and $\tau_0 := 0$, which then yields $\tau_k = \sum_{i=0}^{k-1} \lambda_i$. Drawing inspiration from continuous analysis, we will first show that the function value $f\left(y_k\right) - \inf_{\mathcal{H}} f$ attains the $o\left(\frac{1}{\tau_{k+1}}\right)$ rate of convergence, and the sequence $(y_k)_{k \geq 0}$ converges weakly to a solution. Then, as a crucial result, we will derive a lower bound of τ_{k+1} in terms of k.

The following result emphasizes that the rate of convergence and summability results holds for $(y_k)_{k \geq 0}$ for arbitrary step sizes λ_k that satisfy $\sum_{k \geq 0} \lambda_k = +\infty$. The proof is an adaptation of [4, Theorem 4.1].

Theorem 4.1. *Let* $y_0 \in \mathcal{H}$, $(\lambda_k)_{k \geq 0}$ *be a given positive sequence satisfying* $\displaystyle\sum_{k \geq 0} \lambda_k = +\infty$, $\tau_0 = 0$ *and* $\tau_k = \sum_{i=0}^{k-1} \lambda_i$ *for every* $k \geq 1$. *Then, for any sequence* $(y_k)_{k \geq 0}$ *generated by the proximal algorithm*

$$y_{k+1} := \operatorname{prox}_{\lambda_k f}\left(y_k\right) \quad \forall k \geq 0, \tag{74}$$

the following properties are satisfied:

(i) *(summability of function values)* $\displaystyle\sum_{k \geq 0} \lambda_k \left(f\left(y_{k+1}\right) - \inf_{\mathcal{H}} f\right) < +\infty$;

(ii) *(summability of gradients)* $\displaystyle\sum_{k \geq 0} \tau_k \lambda_k \left\|\nabla f\left(y_{k+1}\right)\right\|^2 < +\infty$;

(iii) *(summability of velocities)* $\displaystyle\sum_{k \geq 0} \frac{\tau_k}{\lambda_k} \left\|y_{k+1} - y_k\right\|^2 < +\infty$;

(iv) *(convergence of function values)* $f\left(y_{k+1}\right) - \inf_{\mathcal{H}} f = o\left(\dfrac{1}{\tau_{k+1}}\right)$ *as* $k \to +\infty$;

(v) *(convergence of gradient)* $\left\|\nabla f\left(y_{k+1}\right)\right\| = o\left(\dfrac{1}{\sqrt{\sum_{l=0}^{k} \tau_l \lambda_l}}\right)$ *as* $k \to +\infty$;

(vi) *the sequence of iterates* $(y_k)_{k \geq 0}$ *converges weakly as* $k \to +\infty$, *and its limit belongs to* $S = \operatorname{argmin}_{\mathcal{H}} f$.

19

Proof. Let $k \geq 0$ be fixed. Take $z_* \in S = \operatorname{argmin} f$. According to (71) and the convexity of f, we deduce that

$$\frac{1}{2}\|y_{k+1} - z_*\|^2 = \frac{1}{2}\|y_k - z_*\|^2 + \langle y_{k+1} - z_*, y_{k+1} - y_k \rangle - \frac{1}{2}\|y_{k+1} - y_k\|^2$$

$$= \frac{1}{2}\|y_k - z_*\|^2 - \lambda_k \langle y_{k+1} - z_*, \nabla f(y_{k+1}) \rangle - \frac{1}{2}\lambda_k^2 \|\nabla f(y_{k+1})\|^2$$

$$\leq \frac{1}{2}\|y_k - z_*\|^2 - \lambda_k (f(y_{k+1}) - \inf_{\mathcal{H}} f) - \frac{1}{2}\lambda_k^2 \|\nabla f(y_{k+1})\|^2. \tag{75}$$

Statement (i) follows from [14, Lemma 5.31]. In addition, the limit $\lim_{k \to +\infty} \|y_k - z_*\| \in \mathbb{R}$ exists, which means that the first condition of the discrete Opial's lemma is fulfilled.

On the other hand, the sequence $(f(y_k) - \inf_{\mathcal{H}} f)_{k \geq 0}$ is nonincreasing. Precisely, we have for every $k \geq 0$

$$(f(y_k) - \inf_{\mathcal{H}} f) - (f(y_{k+1}) - \inf_{\mathcal{H}} f) \geq \langle \nabla f(y_{k+1}), y_k - y_{k+1} \rangle = \lambda_k \|\nabla f(y_{k+1})\|^2 \geq 0. \tag{76}$$

According to [6, Lemma 22] we get

$$f(y_{k+1}) - \inf_{\mathcal{H}} f = o\left(\frac{1}{\sum_{i=0}^{k} \lambda_i}\right),$$

which proves (iv).

Let $k \geq 1$. Multiplying both sides of (76) by $\tau_k = \sum_{i=0}^{k-1} \lambda_i > 0$, then adding the result into (75), we get

$$\tau_{k+1}(f(y_{k+1}) - \inf_{\mathcal{H}} f) + \frac{1}{2}\|y_{k+1} - z_*\|^2 \leq \tau_k(f(y_k) - \inf_{\mathcal{H}} f) + \frac{1}{2}\|y_k - z_*\|^2$$

$$- \frac{1}{2}\lambda_k^2 \|\nabla f(y_{k+1})\|^2 - \tau_k \lambda_k \|\nabla f(y_{k+1})\|^2.$$

This implies

$$\sum_{k \geq 0} \lambda_k^2 \|\nabla f(y_{k+1})\|^2 < +\infty \quad \text{and} \quad \sum_{k \geq 1} \tau_k \lambda_k \|\nabla f(y_{k+1})\|^2 < +\infty,$$

which yields (ii). From (71), we infer (iii). To deduce (v), it suffices to show that the sequence $(\|\nabla f(y_k)\|)_{k \geq 0}$ is nonincreasing. Indeed, it follows from (74) and the cocoercivity of ∇f that

$$\frac{1}{2}\|\nabla f(y_{k+1})\|^2 = \frac{1}{2}\|\nabla f(y_k)\|^2 + \langle \nabla f(y_{k+1}), \nabla f(y_{k+1}) - \nabla f(y_k) \rangle - \frac{1}{2}\|\nabla f(y_{k+1}) - \nabla f(y_k)\|^2$$

$$= \frac{1}{2}\|\nabla f(y_k)\|^2 - \frac{1}{\lambda_k}\langle y_{k+1} - y_k, \nabla f(y_{k+1}) - \nabla f(y_k) \rangle - \frac{1}{2}\|\nabla f(y_{k+1}) - \nabla f(y_k)\|^2$$

$$\leq \frac{1}{2}\|\nabla f(y_k)\|^2.$$

Taking into account also (ii), we obtain (v).

Finally, according to the assumption $\sum_{k \geq 0} \lambda_k = +\infty$, and (iv), we have that $\lim_{k \to +\infty} f(y_k) = \inf_{\mathcal{H}} f$. Since f is convex and lower semicontinuous, the second condition of Opial's lemma is also fulfilled. This gives the weak convergence of the sequence $(y_k)_{k \geq 0}$ to an element in $S = \operatorname{argmin} f$. $\qquad\square$

Then, we give a statement which can be seen as a discrete counterpart of Lemma 2.1. The result is more complex not only because we are in the discrete setting, but also because it allows an explicit choice of the stepsize, as we will see later.

Lemma 4.1. *Let $(\lambda_k)_{k \geq 0}$ be a positive sequence and $(\tau_k)_{k \geq 0}$ such that $\tau_0 = 0$ and $\tau_k = \sum_{i=0}^{k-1} \lambda_i$ for all $k \geq 1$. Suppose that there exist $C_2 > 0$ and $a, b, c \geq 0$ such that $b + c > a$ and*

$$\sum_{k \geq 0} \tau_k^a \lambda_k^{-b} \lambda_{k+1}^{-c} \leq C_2 < +\infty$$

Then, there exists $C_3 > 0$ such that for every $k \geq 1$ it holds

$$\tau_{k+1} \geq C_3 k^{\frac{b+c+1}{b+c-a}}. \tag{77}$$

Proof. By applying the Hölder inequality twice we get for all $k \geq 0$

$$\sum_{i=0}^{k} \tau_i^{\frac{a}{b+c+1}} \leq \left(\sum_{i=0}^{k} \tau_i^a \lambda_i^{-b} \lambda_{i+1}^{-c} \right)^{\frac{1}{b+c+1}} \left(\sum_{i=0}^{k} \lambda_i \right)^{\frac{b}{b+c+1}} \left(\sum_{i=0}^{k} \lambda_{i+1} \right)^{\frac{c}{b+c+1}}$$

$$\leq C_2^{\frac{1}{b+c+1}} \left(\sum_{i=0}^{k+1} \lambda_i \right)^{\frac{b+c}{b+c+1}} = C_2^{\frac{1}{b+c+1}} \tau_{k+2}^{\frac{b+c}{b+c+1}}. \tag{78}$$

If $a = 0$ then (77) follows immediately. From now on we suppose that $a > 0$. Inequality (78) becomes

$$\sum_{i=0}^{k} \tau_i^{\frac{a}{b+c+1}} \leq C_2^{\frac{1}{b+c+1}} \left(\tau_{k+2}^{\frac{a}{b+c+1}} \right)^{\frac{b+c}{a}} \quad \forall k \geq 0. \tag{79}$$

Following the continuous counterpart, let us define

$$C_{b+c} := C_2^{\frac{1}{b+c+1}} > 0 \qquad \text{and} \qquad A_k := \sum_{i=0}^{k} \tau_i^{\frac{a}{b+c+1}} \qquad \forall k \geq 0$$

so that (79) becomes

$$A_k \leq C_{b+c} \left(A_{k+2} - A_{k+1} \right)^{\frac{b+c}{a}} \quad \forall k \geq 0.$$

From here,

$$C_{b+c}^{-\frac{a}{b+c}} \leq A_k^{-\frac{a}{b+c}} \left(A_{k+2} - A_{k+1} \right) \quad \forall k \geq 1. \tag{80}$$

For convenience, we define the following function $\psi \colon \mathbb{R}_{++} \to \mathbb{R}_{++}$ as $\psi(r) := r^{-\frac{a}{b+c}}$. It is clear that

$$\frac{d}{dr} \left(\frac{b+c}{b+c-a} r^{1-\frac{a}{b+c}} \right) = \psi(r) \qquad \text{and} \qquad \dot{\psi}(r) = -\frac{a}{b+c} r^{-\frac{a}{b+c}-1} < 0.$$

Since $(A_k)_{k \geq 0}$ is increasing, this means $\psi(A_{k+2}) \leq \psi(r) \leq \psi(A_k)$ for every $A_k \leq r \leq A_{k+2}$.

Let $k \geq 1$ fixed. We consider two separate cases.

Case 1: $\psi(A_k) \leq 2\psi(A_{k+2})$. Then (80) leads to

$$C_{b+c}^{-\frac{a}{b+c}} \leq A_k^{-\frac{a}{b+c}} \left(A_{k+2} - A_k \right) = \psi(A_k) \left(A_{k+2} - A_k \right)$$

$$\leq 2\psi(A_{k+2}) \left(A_{k+2} - A_k \right) = 2\psi(A_{k+2}) \int_{A_k}^{A_{k+2}} 1 \, dr$$

$$\leq 2 \int_{A_k}^{A_{k+2}} \psi(r) \, dr = 2 \frac{b+c}{b+c-a} \left(A_{k+2}^{1-\frac{a}{b+c}} - A_k^{1-\frac{a}{b+c}} \right).$$

Case 2: $\psi(A_k) > 2\psi(A_{k+2})$. This is equivalent to $A_{k+2} > 2^{\frac{b+c}{a}} A_k$. Since $b+c > a$, we can deduce further

$$A_{k+2}^{1-\frac{a}{b+c}} > 2^{\frac{b+c}{a}-1} A_k^{1-\frac{a}{b+c}}.$$

Consequently,

$$A_{k+2}^{1-\frac{a}{b+c}} - A_k^{1-\frac{a}{b+c}} > \left(2^{\frac{b+c}{a}-1} - 1 \right) A_k^{1-\frac{a}{b+c}} \geq \left(2^{\frac{b+c}{a}-1} - 1 \right) A_1^{1-\frac{a}{b+c}},$$

recall that the last inequality follows from the increasing property of $(A_k)_{k \geq 1}$.

In conclusion, for every $k \geq 0$ we have

$$A_{k+2}^{1-\frac{a}{b+c}} - A_k^{1-\frac{a}{b+c}} \geq C_4 := \min \left\{ \frac{1}{2} \left(1 - \frac{a}{b+c} \right) C_{b+c}^{-\frac{a}{b+c}}, \left(2^{\frac{b+c}{a}-1} - 1 \right) A_1^{1-\frac{a}{b+c}}, A_2^{1-\frac{a}{b+c}} \right\} > 0.$$

Telescoping sum arguments combined with (79) imply for every $k \geq 1$

$$C_4 k \leq A_{2k}^{1-\frac{a}{b+c}} - A_0^{1-\frac{a}{b+c}} \leq A_{2k}^{1-\frac{a}{b+c}} \leq C_2^{\frac{b+c-a}{(b+c)(b+c+1)}} \tau_{2k+2}^{\frac{b+c-a}{b+c+1}}.$$

This gives for every $k \geq 1$

$$\tau_{2k+3} \geq \tau_{2k+2} \geq \widetilde{C}_3 k^{\frac{b+c+1}{b+c-a}},$$

where $\widetilde{C}_3 > 0$. We therefore deduce that there exists $C_3 > 0$ such that

$$\tau_{k+1} \geq C_3 k^{\frac{b+c+1}{b+c-a}} \quad \forall k \geq 1,$$

which gives (77). □

Following a plan identical to the continuous case, we successively consider the case where the control by feedback is formulated in terms of speed, then of gradient.

4.2. Adaptive stepsize rules resulting from the discretization of the velocity based system

In this subsection we specialize the algorithm (PEAS) to the case where $G(y_k, y_{k-1}) = \|y_k - y_{k-1}\|$.

Algorithm 2: Proximal algorithm with adaptive step size defined via velocity

---

Input: $y_0 \neq y_{-1} \in \mathcal{H}$
1 **for** $k = 0, 1, \cdots$ **do**
2   **if** $\nabla f(y_k) = 0$ **then**
3   **stop**
4   **else**
5         $\lambda_k := \|y_k - y_{k-1}\|^{-\frac{p-1}{p}}$
6         $y_{k+1} := \operatorname{prox}_{\lambda_k f}(y_k)$
7   **end**
8 **end**

Theorem 4.2. *Let* $(y_k)_{k \geq 0}$ *be the sequence generated by Algorithm 2. Then, it holds*

$$f(y_k) - \inf_{\mathcal{H}} f = o\left(\frac{1}{k^{2-\frac{1}{p}}}\right) \quad as \ k \to +\infty,$$

and the sequence of iterates $(y_k)_{k \geq 0}$ *converges weakly as* $k \to +\infty$*, and its limit belongs to* $S = \operatorname{argmin}_{\mathcal{H}} f$.

Proof. By the choice of the step size, we have from Theorem 4.1 (iii) that

$$\sum_{k \geq 0} \frac{\tau_k}{\lambda_k} \|y_{k+1} - y_k\|^2 = \sum_{k \geq 0} \tau_k \lambda_k^{-1} \lambda_{k+1}^{-\frac{2p}{p-1}} < +\infty,$$

where $\tau_0 = 0$ and $\tau_k := \sum_{i=0}^{k-1} \lambda_i$ for every $k \geq 1$. We are in position to apply Lemma 4.1 with $(a, b, c) := \left(1, 1, \frac{2p}{p-1}\right)$. We get

$$\tau_{k+1} \geq C_3 k^{2-\frac{1}{p}} \quad \forall k \geq 1. \tag{81}$$

Therefore $\sum_{k \geq 0} \lambda_k = \lim_{k \to +\infty} \tau_k = +\infty$, and we can apply Theorem 4.1 to obtain the conclusion. □

4.3. Adaptive stepsize resulting from the discretization of the gradient based system

Now let us specialize the algorithm (PEAS) to the case where $G(y_k, y_{k-1}) = \|\nabla f(y_k)\|$.

Theorem 4.3. *Let* $(y_k)_{k \geq 0}$ *be the sequence generated by Algorithm 3. Then, it holds*

$$f(y_{k+1}) - \inf_{\mathcal{H}} f = o\left(\frac{1}{k^{2-\frac{1}{p}}}\right) \quad as \ k \to +\infty$$

and the sequence of iterates $(y_k)_{k \geq 0}$ *converges weakly as* $k \to +\infty$*, and its limit belongs to* $S = \operatorname{argmin}_{\mathcal{H}} f$.

Proof. In this case we have from Theorem 4.1 (ii)

$$\sum_{k \geq 0} \tau_k \lambda_k \|\nabla f(y_{k+1})\|^2 = \sum_{k \geq 0} \tau_k \lambda_k \lambda_{k+1}^{-\frac{2p}{p-1}} < +\infty, \tag{82}$$

22

Algorithm 3: Proximal algorithm with adaptive step size defined via gradient

---
Input: $y_0 \in \mathcal{H}$
1 **for** $k = 0, 1, \cdots$ **do**
2   **if** $\nabla f(y_k) = 0$ **then**
3   **stop**
4   **else**
5         $\lambda_k := \|\nabla f(y_k)\|^{-\frac{p-1}{p}}$
6         $y_{k+1} := \text{prox}_{\lambda_k f}(y_k)$
7   **end**
8 **end**

---

where $\tau_0 = 0$ and $\tau_k := \sum_{i=0}^{k-1} \lambda_i$ for every $k \geq 1$.

Let us establish an inequality of the type

$$\|\nabla f(y_k)\| \leq C_k \|\nabla f(y_{k+1})\|,$$

for some sequence $C_k > 0$ which is to be precised. We have for all $k \geq 0$

$$\|\nabla f(y_k)\|^2 = \|\nabla f(y_{k+1})\|^2 - 2\langle \nabla f(y_{k+1}), \nabla f(y_{k+1}) - \nabla f(y_k)\rangle + \|\nabla f(y_{k+1}) - \nabla f(y_k)\|^2$$

$$= \|\nabla f(y_{k+1})\|^2 + \frac{2}{\lambda_k}\langle y_{k+1} - y_k, \nabla f(y_{k+1}) - \nabla f(y_k)\rangle + \|\nabla f(y_{k+1}) - \nabla f(y_k)\|^2$$

$$\leq \|\nabla f(y_{k+1})\|^2 + \left(\frac{2L}{\lambda_k} + L^2\right)\|y_{k+1} - y_k\|^2 = (1 + L\lambda_k)^2 \|\nabla f(y_{k+1})\|^2$$

$$\leq (1 + L\lambda_{k+1})^2 \|\nabla f(y_{k+1})\|^2 \tag{83}$$

where $L > 0$ denotes the Lipschitz constant of ∇f on a bounded set containing the sequence $(y_k)_{k \geq 0}$.

Combining (82) and (83), we get

$$\sum_{k \geq 0} \tau_k \lambda_k \frac{1}{(1 + L\lambda_{k+1})^2}\|\nabla f(y_k)\|^2 = \sum_{k \geq 0} \tau_k \lambda_k \frac{1}{(1 + L\lambda_{k+1})^2}\lambda_k^{-\frac{2p}{p-1}} < +\infty. \tag{84}$$

Let us now show that $\lim_{k \to +\infty} \lambda_k = +\infty$. According to the decreasing property of the sequence $(f(y_k) - \inf_{\mathcal{H}} f)_{k \geq 0}$, by summing inequalities (76) we get

$$\sum_{k \geq 0} \lambda_k \|\nabla f(y_{k+1})\|^2 < +\infty. \tag{85}$$

From the closed-loop rule we deduce that

$$\sum_{k \geq 0} \lambda_k \lambda_{k+1}^{-\frac{2p}{p-1}} < +\infty. \tag{86}$$

Therefore

$$\lim_{k \to +\infty} \lambda_k \lambda_{k+1}^{-\frac{2p}{p-1}} = 0.$$

Since $(\lambda_k)_{k \geq 0}$ is increasing, let us denote by $l > 0$ its limit. If l is finite then, by passing to the limit on the above inequality we get $l^{1 - \frac{2p}{p-1}} = 0$, a clear contradiction with $l > 0$. Therefore

$$\lim_{k \to +\infty} \lambda_k = +\infty.$$

In this case $\frac{1}{(1 + L\lambda_{k+1})^2} \sim (L\lambda_{k+1})^{-2}$, which gives

$$\sum_{k \geq 0} \tau_k \lambda_k^{1 - \frac{2p}{p-1}} \lambda_{k+1}^{-2} < +\infty. \tag{87}$$

We are in position to apply Lemma 4.1 with $(a, b, c) := \left(1, \frac{2p}{p-1} - 1, 2\right)$. We get

$$\tau_{k+1} \geq C_3 k^{2 - \frac{1}{p}} \quad \forall k \geq 1.$$

23

We have $\sum_{k\geq 0} \lambda_k = \lim_{k\to+\infty} \tau_k = +\infty$, and we can apply Theorem 4.1 to obtain, as $k \to +\infty$

$$f(y_{k+1}) - \inf_{\mathcal{H}} f = o\left(\frac{1}{k^{2-\frac{1}{p}}}\right).$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 4.1. Note that the closed-loop control of the velocity and the closed-loop control of the gradient give the same convergence rate of the values. Clearly, we have obtained a faster convergence result compared to the classical Proximal Point Algorithm (PPA), which we also cover since it coincides with $p = 1$ in both cases.

5. Inertial proximal algorithms obtained by closed-loop damping

Let us now consider the convergence properties of the sequences $(x_k)_{k\geq 0}$ which are obtained by applying the averaging process to the sequences generated by Algorithm 2. Indeed, we limit our investigation to the closed loop control of the velocity, the case of the closed loop control of the gradient is very similar. Let us discretize the continuous averaging relation

$$\dot{x}(t) + \frac{\dot{\tau}(t)}{\tau(t)}(x(t) - y(t)) = 0$$

as follows (recall that, because of the choice $q = 1$, we have $\dot{\tau}(t) = \lambda(t)$)

$$x_{k+1} - x_k + \frac{\lambda_k}{\tau_{k+1}}(x_k - y_{k+1}) = 0.$$

Equivalently

$$x_{k+1} = \left(1 - \frac{\lambda_k}{\tau_{k+1}}\right)x_k + \frac{\lambda_k}{\tau_{k+1}}y_{k+1}.$$

This gives the following proximal inertial algorithm:

Algorithm 4: Proximal inertial algorithm with adaptive step size defined via velocity

---
Input: $\tau_0 := 0$ and $x_0, y_0 \neq y_{-1} \in \mathcal{H}$
1 **for** $k = 0, 1, \cdots$ **do**
2 $\quad$ **if** $\nabla f(y_k) = 0$ **then**
3 $\qquad$ **stop**
4 $\quad$ **else**
5 $\qquad$ $\lambda_k := \|y_k - y_{k-1}\|^{-\frac{p-1}{p}}$
6 $\qquad$ $y_{k+1} := \mathrm{prox}_{\lambda_k f}(y_k)$
7 $\qquad$ $\tau_{k+1} := \tau_k + \lambda_k$
8 $\qquad$ $x_{k+1} := \left(1 - \frac{\lambda_k}{\tau_{k+1}}\right)x_k + \frac{\lambda_k}{\tau_{k+1}}y_{k+1}.$
9 $\quad$ **end**
10 **end**

Theorem 5.1. Let $(x_k)_{k\geq 0}$ be the sequence generated by Algorithm 4. Then, it holds

$$f(x_k) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^{2-\frac{1}{p}}}\right) \quad as \ k \to +\infty$$

and the sequence of iterates $(x_k)_{k\geq 0}$ converges weakly as $k \to +\infty$, and its limit belongs to $S = \mathrm{argmin}_{\mathcal{H}} f$.

Proof. Let $k \geq 0$. By definition of x_{k+1} we have

$$\tau_{k+1} x_{k+1} = (\tau_{k+1} - \lambda_k)x_k + \lambda_k y_{k+1}$$

24

which gives (recall that $\tau_{k+1} = \sum_{i=0}^{k} \lambda_i$)

$$\tau_{k+1}x_{k+1} - \tau_k x_k = (\tau_{k+1} - \lambda_k)x_k + \lambda_k y_{k+1} - \tau_k x_k = (\tau_{k+1} - \tau_k - \lambda_k)x_k + \lambda_k y_{k+1} = \lambda_k y_{k+1}.$$

Therefore, by telescoping arguments we obtain

$$x_{k+1} = \frac{\sum_{i=0}^{k} \lambda_i y_{i+1}}{\tau_{k+1}} \quad \forall k \geq 0. \tag{88}$$

By convexity of f we infer

$$f\left(x_{k+1}\right) - \inf_{\mathcal{H}} f = \left(f - \inf_{\mathcal{H}} f\right)\left(x_{k+1}\right) = \left(f - \inf_{\mathcal{H}} f\right)\left(\frac{\sum_{i=0}^{k} \lambda_i y_{i+1}}{\tau_{k+1}}\right)$$

$$\leq \frac{1}{\tau_{k+1}} \sum_{i=0}^{k} \lambda_i \left(f - \inf_{\mathcal{H}} f\right)\left(y_{i+1}\right) = \frac{1}{\tau_{k+1}} \sum_{i=0}^{k} \lambda_i \left(f\left(y_{i+1}\right) - \inf_{\mathcal{H}} f\right),$$

By Theorem 4.1 (i), we have $\sum_{k\geq 0} \lambda_k \left(f\left(y_{k+1}\right) - \inf_{\mathcal{H}} f\right) < +\infty$, and by (81) we have $\tau_{k+1} \geq k^{2-\frac{1}{p}}$, which gives the claim. The weak convergence of $(x_k)_{k\geq 0}$ to an element in $S = \operatorname{argmin}_{\mathcal{H}} f$ follows from the weak convergence of $(y_k)_{k\geq 0}$ and the Stolz-Cesáro Theorem. $\qquad\square$

5.1. **Geometric interpretation of Algorithm 4**

First note that Algorithm 4 can be equivalently written as follows

$$x_{k+1} = \left(1 - \frac{\lambda_k}{\tau_{k+1}}\right)x_k + \frac{\lambda_k}{\tau_{k+1}} \operatorname{prox}_{\lambda_k f}\left(x_{k-1} + \frac{\tau_k}{\lambda_{k-1}}(x_k - x_{k-1})\right). \tag{89}$$

Since $\frac{\tau_k}{\lambda_{k-1}} > 1$, the algorithm first involves an extrapolation step (this is the inertial aspect), then a proximal step, and finally a relaxation step which balances the inertia effect and dampens the oscillations. This is shown in the figure below. We set $\theta_k = \frac{\lambda_k}{\tau_{k+1}} \in]0,1[$.
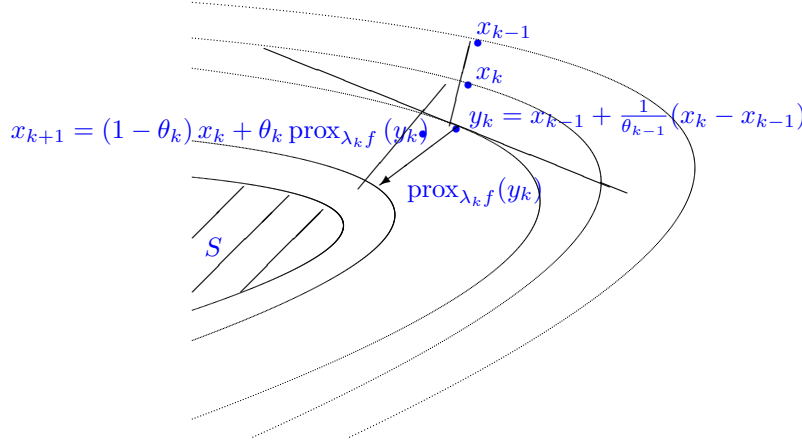
FIGURE 1. A geometrical illustration of Algorithm 4.

Despite some analogies, Algorithm 4 is different from the relaxed inertial proximal algorithm (RIPA) considered by Attouch and Cabot in [7], and which writes

$$\begin{cases} y_k & = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} & = (1-\rho_k)y_k + \rho_k \operatorname{prox}_{\lambda_k f}(y_k) \end{cases} \tag{90}$$

As main difference, in Algorithm 4 the relaxation is taken between x_k and $\operatorname{prox}_{\lambda_k f}(y_k)$, while in (RIPA) it is taken between y_k and $\operatorname{prox}_{\lambda_k f}(y_k)$. Consequently, Algorithm 4 involves a Hessian damping effect which is not present in (RIPA). Note in Algorithm 4 the balance between the extrapolation (inertial, acceleration) effect and the relaxation effect. Moreover, our construction provides coefficients which are generated automatically

in closed loop way, whereas in (RIPA) they require subtle adjustment. The importance of the relaxation technique when combined with inertia has been put to the fore in [22]. According to (88), x_{k+1} can be interpreted as an average of the $\{y_n : 0 \leq n \leq k+1\}$, which makes our approach somewhat analogous to the nonlinear averaging technique developed in [33], where it is assumed that there is a unique minimizer. Averaging techniques have also been used in [32] in the context of hybrid systems. Indeed, adjusting the damping in a closed-loop ad hoc manner bears some analogy to restarting methods.

5.2. **Extension to the nonsmooth setting**

Since our proposed numerical algorithms are proximal methods, they can be used also to minimize nonsmooth and convex functions. Given an optimization problem

$$\min \{f(x) : x \in \mathcal{H}\}, \tag{91}$$

where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, lower semicontinuous, and convex function with $S = \mathrm{argmin}_{\mathcal{H}} f \neq \emptyset$, one can equivalently consider the problem

$$\min \{f_\gamma(x) : x \in \mathcal{H}\}, \tag{92}$$

where $f_\gamma : \mathcal{H} \to \mathbb{R}$ stands for the Moreau envelope of f with parameter $\gamma > 0$, defined as

$$f_\gamma(x) = \inf_{y \in \mathcal{H}} \left\{ f(x) + \frac{1}{2\gamma} \|x - y\|^2 \right\}.$$

The two problems share the same optimal value and solution set $\mathrm{argmin}_{\mathcal{H}} f = \mathrm{argmin}_{\mathcal{H}} f_\gamma$, while the Moreau envelope is convex and differentiable and it has a γ^{-1}-Lipschitz continuous gradient ([14,15]). The formulas for the proximal operators of f and f_γ are closely related by a simple convex combination, which makes no difference when solving (91) and (92) conceptually. We can therefore exploit these premises to apply our methods to a broader class of functions that are only assumed to be proper, lower semicontinuous, and convex.

6. Numerical experiments

In this section, we carry out numerical experiments in order to demonstrate the effectiveness of the methods we have proposed.

Let $q \geq 1$ and $X \in \mathbb{R}^{m \times n}$. We denote by $\|\cdot\|_{\mathcal{S}_q}$ the Schatten q-norm of X, which is defined as

$$\|X\|_{\mathcal{S}_q}^q = \sum_i \sigma_i^q(X),$$

where $\sigma_i(X)$ denote the ith-singular value of X. In case $q := 1$ it gives the nuclear norm, which we denote by $\|\cdot\|_*$, and in case $q := 2$ it gives the Frobenius norm, which we denote by $\|\cdot\|_\mathsf{F}$. Further, we denote by $E_{\lambda_-,\lambda_+}(\mathbb{R}^{n \times n})$ the set of positive semidefinite $n \times n$ matrices with eigenvalues belonging to $[\lambda_-, \lambda_+]$. The logarithm of a matrix $X \in \mathbb{R}^{n \times n}$ is the matrix $Y := \log(X) \in \mathbb{R}^{n \times n}$ such that

$$X = e^Y = \sum_{i \geq 0} \frac{1}{i!} Y^i.$$

The functions used in the numerical experiments were:
a) "log det + *Nuclear norm*"

$$f : \mathbb{R}^{n \times n} \to \mathbb{R} \cup \{+\infty\}, \quad f(X) := \begin{cases} -\log(\det(X)) + \mu \|X\|_* & \text{if } X \succ 0, \\ +\infty & \text{otherwise}, \end{cases}$$

b) "log det + *Squared Frobenius norm*"

$$f : \mathbb{R}^{n \times n} \to \mathbb{R} \cup \{+\infty\}, \quad f(X) := \begin{cases} -\log(\det(X)) + \mu \|X\|_\mathsf{F}^2 & \text{if } X \succ 0, \\ +\infty & \text{otherwise}, \end{cases}$$

26

c) "log det + *Bounds on eigenvalues*"

$$f : \mathbb{R}^{n \times n} \to \mathbb{R} \cup \{+\infty\}, \quad f(X) := \begin{cases} -\log\left(\det\left(X\right)\right) & \text{if } X \in E_{\lambda_-,\lambda_+}\left(\mathbb{R}^{n \times n}\right), \\ +\infty & \text{otherwise}, \end{cases}$$

d) "*von Neumann entropy + Nuclear norm*"

$$f : \mathbb{R}^{n \times n} \to \mathbb{R} \cup \{+\infty\}, \quad f(X) := \begin{cases} \text{trace}\left(X \log\left(X\right)\right) + \mu \left\|X\right\|_* & \text{if } X \succeq 0, \\ +\infty & \text{otherwise}, \end{cases}$$

e) "*von Neumann entropy + Squared Frobenius norm*"

$$f : \mathbb{R}^{n \times n} \to \mathbb{R} \cup \{+\infty\}, \quad f(X) := \begin{cases} \text{trace}\left(X \log\left(X\right)\right) + \mu \left\|X\right\|_{\mathsf{F}}^2 & \text{if } X \succeq 0, \\ +\infty & \text{otherwise}, \end{cases}$$

f) "*von Neumann entropy + Bounds on eigenvalues*"

$$f : \mathbb{R}^{n \times n} \to \mathbb{R} \cup \{+\infty\}, \quad f(X) := \begin{cases} \text{trace}\left(X \log\left(X\right)\right) & \text{if } X \in E_{\lambda_-,\lambda_+}\left(\mathbb{R}^{n \times n}\right), \\ +\infty & \text{otherwise}, \end{cases}$$

g) "*Ridge + Schatten q-penalty*"

$$f : \mathbb{R}^{m \times n} \to \mathbb{R}, \quad f(X) := \frac{1}{2}\left\|X\right\|_{\mathsf{F}}^2 + \mu \left\|X\right\|_{\mathcal{S}_q}^q.$$

For all cases we considered $\mu := 10^{-1}$ and solved the optimization problem (92) for $\gamma := 1$. For this purpose we used Algorithm 2 and Algorithm 3 when $p := 2$, the classical proximal point algorithm (PPA) ([14, 15]) and the special case of FISTA obtained in the absence of the smooth term ([16]), which we also compared with each other. For the formulas of the proximal operators of the seven objective functions, see [15, 17, 20].

We have set $n := 10^2$, and for the Ridge + Schatten q-penalty we have chosen $m \in \left\{1, 10^2, 10^4\right\}$ to see to what extent the dimension of the matrix affects the numerical performance, also for different values of q. The initial point Y_0 has been generated randomly, whereas for Algorithm 2 we have simply added 1 to all its entries to obtain Y_{-1}.

We have stopped the algorithms either when

$$\frac{\left\|Y_{k+1} - Y_k\right\|}{\left\|Y_{k+1}\right\|} < \texttt{Tol}$$

or if they a maximum allowed number of iterations $\texttt{Ite\_max}$ has been exceeded. We have set $\texttt{Tol} := 10^{-16}$ and $\texttt{Ite\_max} := 10^3$. The performance of the four algorithms has been compared in terms of $f\left(\text{prox}_f\left(Y_k\right)\right) - f_*$ in logarithmic scale, where f_* is the minimum value the objective function takes over all generated sequences.

According to the main convergence theorems, for $p = 2$, Algorithm 2 and Algorithm 3 have rates of convergence of $o\left(k^{-3/2}\right)$ as $k \to +\infty$, which is faster than that of PPA, but worse than that of FISTA. As Figure 2 shows for the objective functions (a) - (f), Algorithm 2 and Algorithm 3 demonstrate competitive performance against PPA and FISTA, outperforming them in many instances. On the other hand, they do not exhibit oscillations, a phenomenon known to occur with momentum algorithms. The better convergence properties of the algorithms presented in this paper are confirmed for the different instances considered in the Figures 3 - 5 when minimizing the *Ridge + Schatten q-penalty function*, especially as the values for q become larger. We believe that the consistent improvements in the convergence of the function values observed for Algorithm 2 and Algorithm 3 are due to the fact that these algorithms make much better use of local information and adjust the step size accordingly.

7. Conclusion and perspective

Our study proposes new fast adaptive optimization methods for convex optimization. We have shown that the time scaling and averaging technique, previously developed by the authors in the context of non-autonomous systems, can be developed by taking closed-loop time parameterization, giving rise to autonomous dynamics. The method turns out to be flexible, because it is based on elementary mathematical
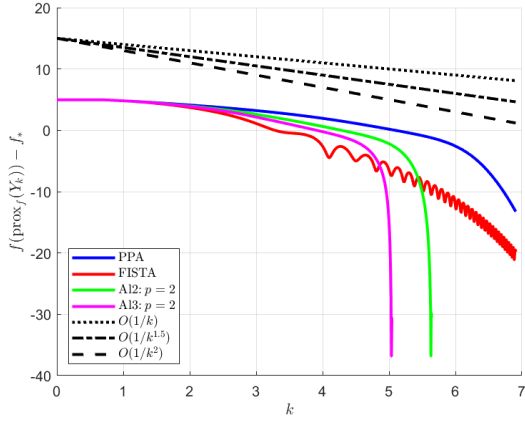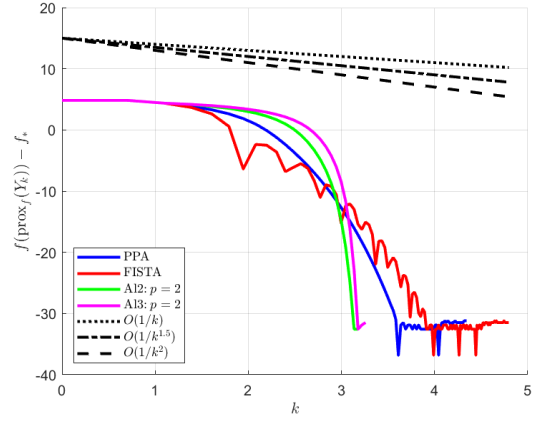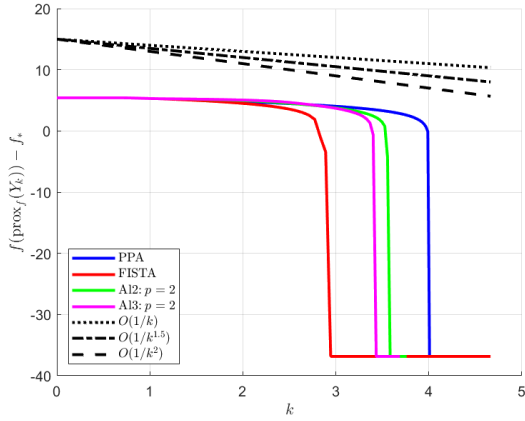
(A) "log det + *Nuclear norm*"
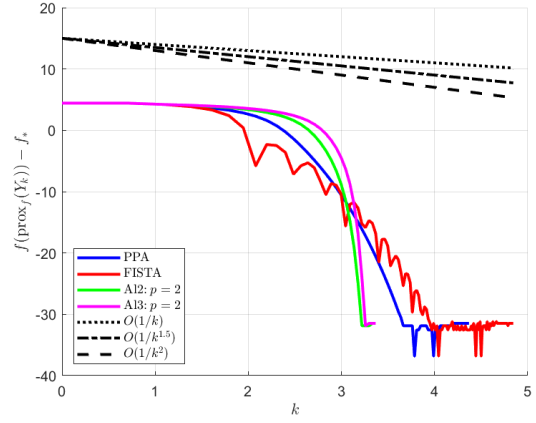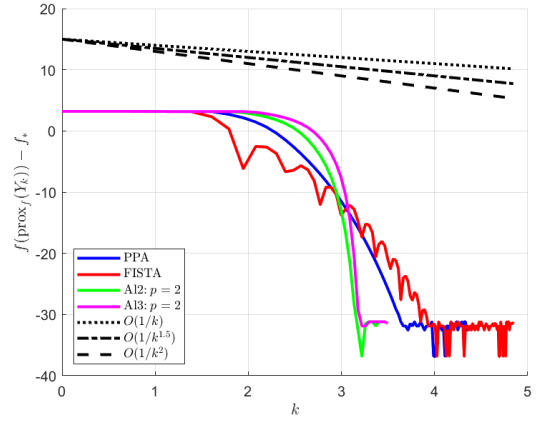

(D) "*von Neumann entropy + Nuclear norm*"
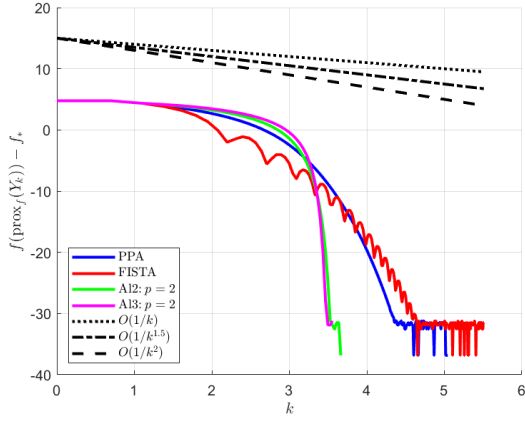
(B) "log det + *Squared Frobenius norm*"

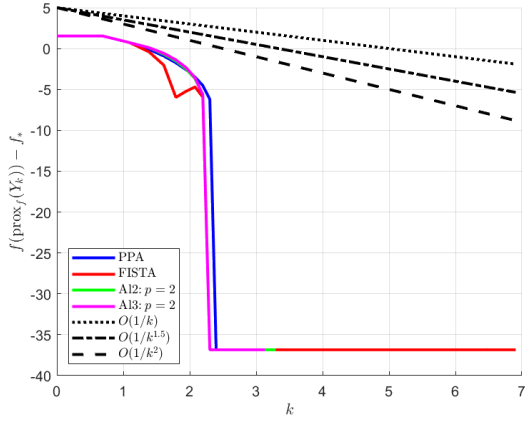

(E) "*von Neumann entropy + Squared Frobenius norm*"
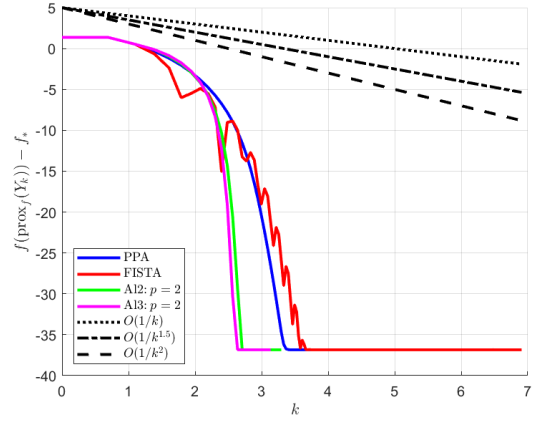

(C) "log det + *Bounds on eigenvalues*"

(F) "*von Neumann entropy + Bounds on eigenvalues*"

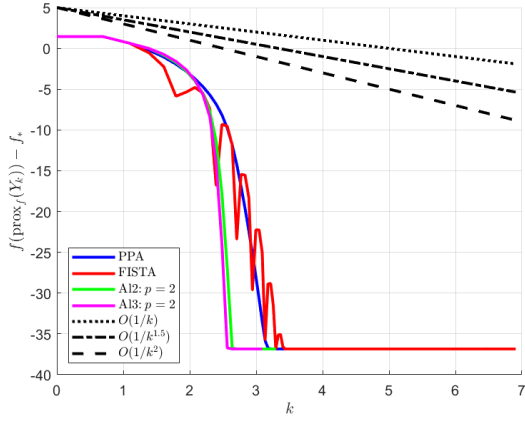FIGURE 2. Numerical comparisons between PPA, FISTA, and Algorithm 2 and Algorithm 3 for $n = 10^2$.

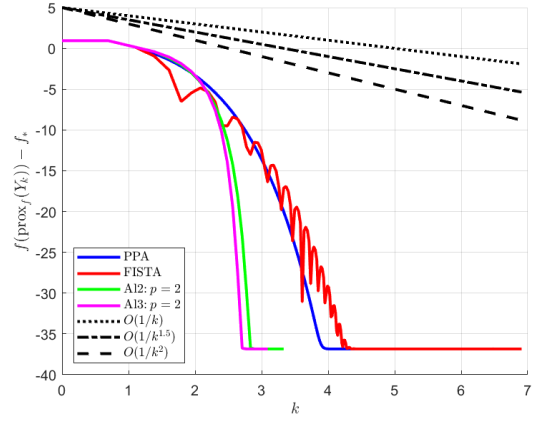tools, namely the dynamics of the steepest descent, and the operations of temporal parameterization and

(A) "*Ridge + Nuclear norm*"

(C) "*Ridge + Schatten 3/2-penalty*"

(B) "*Ridge + Schatten 4/3-penalty*"

(D) "*Ridge + Schatten 4-penalty*"

FIGURE 3. Numerical comparisons between PPA, FISTA, and Algorithm 2 and Algorithm 3 for $(m, n) := \left(1, 10^2\right)$.

averaging. It is therefore not necessary to redo a Lyapunov analysis, one relies on the classic results for the steepest descent. The results obtained for the continuous dynamics pass quite naturally to the corresponding proximal algorithms, where the iterates are expressed in a direct way according to the proximal terms. This study is one of the very first to develop an algorithmic framework based on autonomous dynamics and which, when specialized, provides the convergence rates of the dynamical surrogate of the Nesterov acceleration gradient method. Another important aspect of our analysis is that it exhibits Hessian-driven damping, which plays a key role in damping oscillations. Our work opens up many perspectives, our method naturally extending to gradient algorithms, proximal-gradient algorithms for composite optimization, cocercive monotone operators, and the study of the stochastic version, to name only a few.
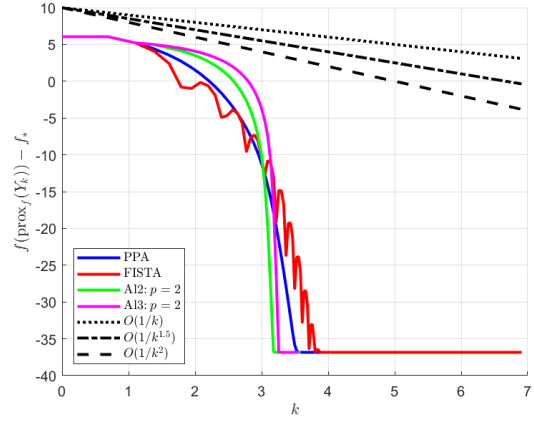
8. APPENDIX

8.1. **Classical facts concerning the continuous steepest descent**
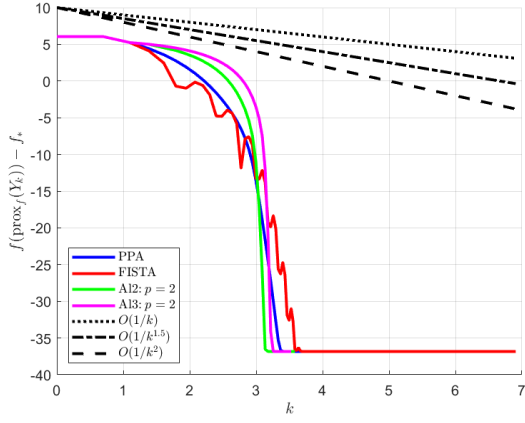
Consider the classical continuous steepest descent
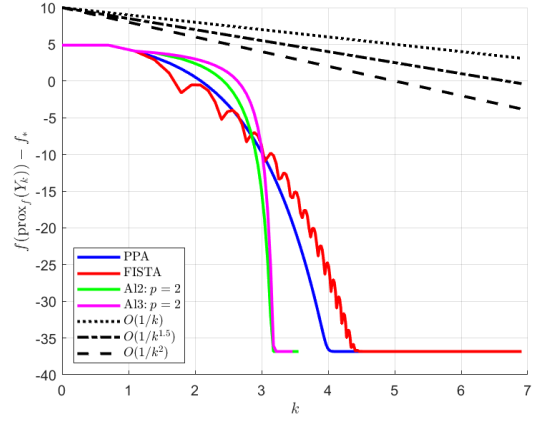
$$\text{(SD)} \quad \dot{z}(t) + \nabla f(z(t)) = 0. \tag{93}$$
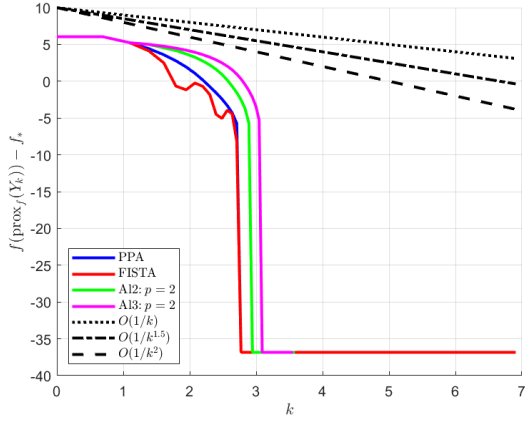
(A) "*Ridge + Nuclear norm*"

(C) "*Ridge + Schatten 3/2-penalty*"

(B) "*Ridge + Schatten 4/3-penalty*"

(D) "*Ridge + Schatten 4-penalty*"

FIGURE 4. Numerical comparisons between PPA, FISTA, and Algorithm 2 and Algorithm 3 for $(m, n) := \left(10^2, 10^2\right)$.

Under the standing assumption $(\mathcal{A})$ on f, we know that, for any $z_0 \in \mathcal{H}$ there exists a unique classical global solution $z \in \mathcal{C}^1([t_0, +\infty[\colon \mathcal{H})$ of (SD) satisfying $z(t_0) = z_0$, see [5, Theorem 17.1.1]. We fix t_0 as the origin of time. Recall classical facts concerning the continuous steepest descent.

Theorem 8.1. *Suppose that $f\colon \mathcal{H} \to \mathbb{R}$ satisfies $(\mathcal{A})$. Let $z\colon [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of*

$$\dot{z}(t) + \nabla f(z(t)) = g(t) \tag{94}$$

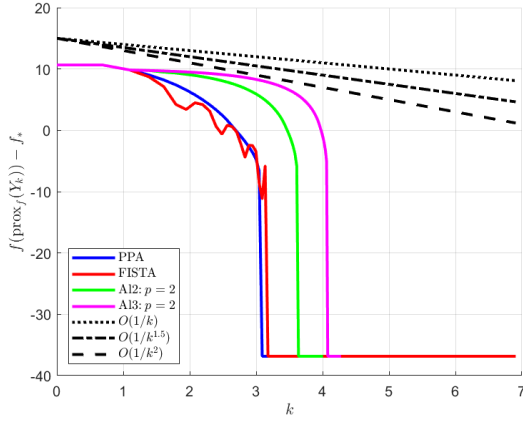*where $g\colon [t_0, +\infty[\to \mathcal{H}$ is such that*

$$\int_{t_0}^{+\infty} \|g(t)\| \, dt < +\infty \ \text{and} \ \int_{t_0}^{+\infty} t \, \|g(t)\|^2 \, dt < +\infty. \tag{95}$$

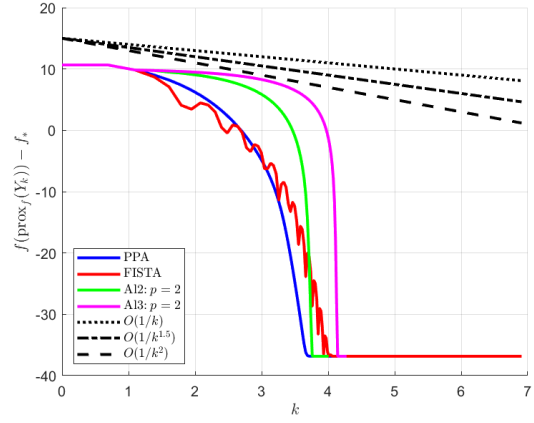*Then, the following statements are satisfied:*

(i) (convergence of gradients towards zero) $\quad \|\nabla f(z(t))\| = o\left(\dfrac{1}{\sqrt{t}}\right)$ *as $t \to +\infty$.*

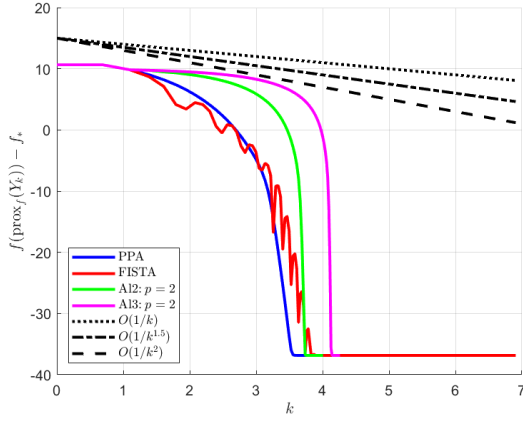*(ii) (integral estimate of the velocities)* $\displaystyle\int_{t_0}^{+\infty} t \, \|\dot{z}(t)\|^2 \, dt < +\infty.$

(iii) (integral estimate of the gradients) $\displaystyle\int_{t_0}^{+\infty} t \, \|\nabla f(z(t))\|^2 \, dt < +\infty.$
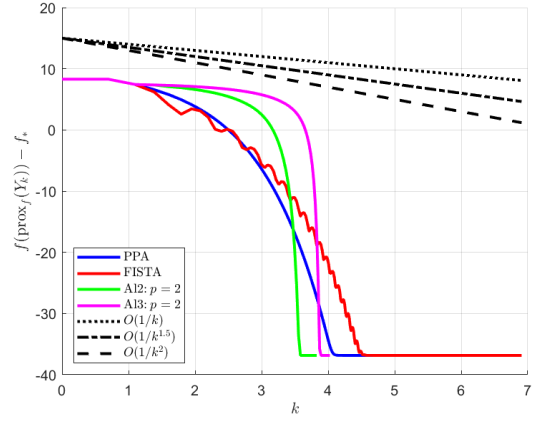
30

(A) "*Ridge + Nuclear norm*"

(C) "*Ridge + Schatten 3/2-penalty*"

(B) "*Ridge + Schatten 4/3-penalty*"

(D) "*Ridge + Schatten 4-penalty*"

FIGURE 5. Numerical comparisons between PPA, FISTA, and Algorithm 2 and Algorithm 3 for $(m,n) := \left(10^4, 10^2\right)$.

(iv) *(convergence of values)* $f\left(z\left(t\right)\right) - \inf_{\mathcal{H}} f = o\left(\dfrac{1}{t}\right)$ *as* $t \to +\infty$.

(v) *(improved convergence rates of gradients) if* $g(t) \equiv 0$, *then* $\|\nabla f\left(z\left(t\right)\right)\| = o\left(\dfrac{1}{t}\right)$ *as* $t \to +\infty$.

(vi) *The solution trajectory* $z(t)$ *converges weakly as* $t \to +\infty$, *and its limit belongs to* $\mathcal{S} = \operatorname{argmin} f$.

If $g(t) \equiv 0$, we have that $t \mapsto \|\nabla f\left(z\left(t\right)\right)\|$ is nonincreasing, since in this case

$$\frac{d}{dt}\|\nabla f\left(z\left(t\right)\right)\|^2 = 2\left\langle \nabla f\left(z\left(t\right)\right), \frac{d}{dt}\nabla f\left(z\left(t\right)\right)\right\rangle = -2\left\langle \dot{z}\left(t\right), \frac{d}{dt}\nabla f\left(z\left(t\right)\right)\right\rangle \leq 0 \quad \forall t \geq t_0.$$

Therefore, from the integral estimate of the gradients we deduce that $\|\nabla f\left(z\left(t\right)\right)\| = o\left(\frac{1}{t}\right)$.

8.2. **Auxiliary result**

Opial's Lemma is a basic ingredient of the convergence analysis.

Lemma 8.1. *(Opial) Let* S *be a nonempty subset of* $\mathcal{H}$ *and let* $(x_k)_{k \geq 0}$ *be a sequence in* $\mathcal{H}$. *Assume that*

(i) *for every* $z \in S$, $\lim_{k \to +\infty} \|x_k - z\|$ *exists;*

(ii) *every weak sequential limit point of* $(x_k)_{k \geq 0}$, *as* $k \to +\infty$, *belongs to* S.

Then, $(x_k)_{k \geq 0}$ *converges weakly as* $k \to +\infty$*, and its limit belongs to S.*

Acknowledgements

References

[1] C.D. Alecsa, S. László, T. Pinta, *An extension of the second order dynamical system that models Nesterov's convex gradient method*, Appl. Math. Opt. **84** (2021), 1687–1716.

[2] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, *Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling*, Evol. Eq. Control. Theor. **11**(2) (2022), 487–514.

[3] H. Attouch, R.I. Boţ, E.R. Csetnek, *Fast optimization via inertial dynamics with closed-loop damping*. J. Eur. Math. Soc. **25**(5) (2023), 1985–2056.

[4] H. Attouch, R.I. Boţ, D.-K. Nguyen, *Fast convex optimization via time scale and averaging of the steepest descent*, Math. Oper. Res. (2024), DOI: 10.1287/moor.2023.0186.

[5] H. Attouch, G. Buttazzo, G. Michaille, *Variational Analysis in Sobolev and BV Spaces. Applications to PDE's and Optimization*, MOS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, PA, (2014).

[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, *Convergence of a relaxed inertial proximal algorithm for maximally monotone operators*, Math. Program. **184** (2020), 243–287.

[8] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order algorithms via inertial systems with Hessian driven damping*, Math. Program. **193** (2022), 113–155.

[9] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping*, Optimization, **72**(5) (2023), 1199–1238.

[10] H. Attouch, J. Fadili, V. Kungurtsev, *On the effect of perturbations, errors in first-order optimization methods with inertia and Hessian driven damping*, Evol. Eq. Control. Theor., **12**(1) (2023), 71–117.

[11] H. Attouch, J. Peypouquet, P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Diff. Eq. **261** (2016), 5734–5783.

[12] H. Attouch, P. Redont, B. F. Svaiter, *Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces*, J. Optim. Theory Appl. **157**(3) (2013), 624–650.

[13] H. Attouch, B. F. Svaiter, *A continuous dynamical Newton-Like approach to solving monotone inclusions*, SIAM J. Control Optim. **49**(2) (2011), 574–598.

[14] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CSM Books in Mathematics, Springer, Cham, (2011).

[15] A. Beck, *First-Order Methods in Optimization*, MOS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, PA, (2017).

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

[17] A. Benfenati, E. Chouzenoux, J.–C. Pesquet, *Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation*, Signal Process. **169** (2020), Article 107417.

[18] R. I. Boţ, E. R. Csetnek, S.C. László, *Tikhonov regularization of a second order dynamical system with Hessian damping*, Math. Program. **189** (2021), 151–186.

[19] C. Castera, J. Bolte, C. Févotte, E. Pauwels, *An inertial Newton algorithm for deep learning*, J. Mach. Learn. Res. **22** (2021), 1–31.

[20] G. Chierchia, E. Chouzenoux, P. L. Combettes, and J.-C. Pesquet, *The Proximity Operator Repository* https://proximity-operator.net/

[21] A. Haraux, M. A. Jendoubi, *The Convergence Problem for Dissipative Autonomous Systems*, Springer Briefs in Mathematics, (2015).

[22] F. Iutzeler, J.M. Hendrickx, *A generic online acceleration scheme for optimization algorithms via relaxation and inertia*, Opt. Methods Soft. **34**(2) (2019), 383–405.

[23] T. Lin, M. I. Jordan, *A control-theoretic perspective on optimal high-order optimization*, Math. Program. **195**(1) (2022), 929–975.

[24] T. Lin, M. I. Jordan, *Monotone inclusions, acceleration and closed-loop control*, Math. Oper. Res. **48**(4) (2023),2353–2382.

[25] T. Lin, M. I. Jordan, *A continuous-time perspective on global acceleration for monotone equation problems*, Comm. Opt. Theory **2024** (2024), Article 40, 1–25.

[26] S. Maier, C. Castera, P. Ochs, *Near-optimal closed-loop method via Lyapunov damping for convex optimization*, ArXiv: 2311.10053.

[27] P.L. Maingé, F. Labarre, *First-order frameworks for continuous Newton-like dynamics governed by maximally monotone operators*, Set-Valued Var. Anal. **30**(2022), 425–451.

[28] M. Muehlebach, M. I. Jordan, *A dynamical systems perspective on Nesterov acceleration*, Proc. of ICML, PMLR 27, (2019).

[29] R. D. C. Monteiro, B. F. Svaiter, *An Accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods*, SIAM J. Optim. **23**(2) (2013), 1092–1125.

[30] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$*, Soviet Math. Dokl. **27** (1983), 372–376.

[31] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, Boston, (2004).

[32] J. I. Poveda, N. Li, *Robust hybrid zero-order optimization algorithms with acceleration via averaging in time*, Automatica **123** (2021), Article 109361.

[33] D. Scieur, A. D'Aspremont, F. Bach, *Regularized nonlinear acceleration*, Adv. in Neural Inform. Proc. Systems **29** (2016).

[34] B. Shi, S. S. Du, M. I. Jordan, W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, Math. Program. **195** (2022), 79–148.

[35] C. Song, Y. Jiang, Y. Ma, *Unified acceleration of high-order algorithms under general Hölder continuity*, SIAM J. Optim. **31**(3) (2021), 1797–1826.

[36] W. J. Su, S. Boyd, E. J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, Adv. in Neural Inform. Proc. Systems **27** (2014).