



**HAL**  
open science

# A doubly nonlinear evolution system with threshold effects associated with dry friction

Samir Adly, Hedy Attouch, Manh Hung Le

► **To cite this version:**

Samir Adly, Hedy Attouch, Manh Hung Le. A doubly nonlinear evolution system with threshold effects associated with dry friction. 2023. hal-04164463

**HAL Id: hal-04164463**

**<https://cnrs.hal.science/hal-04164463>**

Preprint submitted on 18 Jul 2023

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

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

# A doubly nonlinear evolution system with threshold effects associated with dry friction

Samir Adly\*      Hedy Attouch†      Manh Hung Le‡

**ABSTRACT.** In this paper, we investigate the asymptotic behavior of inertial dynamics with dry friction within the context of a Hilbert framework for convex differentiable optimization. Our study focuses on a doubly nonlinear first-order evolution inclusion that encompasses two potentials. In our analysis, we specifically focus on two main components: the differentiable function  $f$  that needs to be minimized, which influences the system's state through its gradient, and the nonsmooth dry friction potential denoted as  $\varphi = r\|\cdot\|$ . It's important to note that the dry friction term acts on a linear combination of the velocity vector and the gradient of  $f$ . Consequently, any stationary point in our system corresponds to a critical point of  $f$ , unlike the case where only the velocity vector is involved in the dry friction term, resulting in an approximate critical point of  $f$ . To emphasize the crucial role of  $\nabla f(x)$ , we also explore the dual formulation of this dynamic, which possesses a Riemannian gradient structure. To address these dynamics, we employ the recently developed generic acceleration approach by Attouch, Bot, and Nguyen. This approach involves the time scaling of a continuous first-order differential equation, followed by the application of the method of averaging. By applying this methodology, we derive fast convergence results for second-order time-evolution systems with dry friction, asymptotically vanishing viscous damping, and implicit Hessian-driven damping.

Mathematics Subject Classifications: 37N40, 46N10, 49M30, 65B99, 65K05, 65K10, 90B50, 90C25.

Key words and phrases: fast convex optimization; damped inertial dynamic; dry friction; time scaling; averaging; Nesterov algorithms; Hessian driven damping.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Some historical facts . . . . .	4
1.1.1	Viscous friction . . . . .	4
1.1.2	Dry friction . . . . .	5
1.1.3	Hessian-driven damping . . . . .	5
1.2	Contents . . . . .	6

---

\*Laboratoire XLIM, Université de Limoges, 123, avenue Albert Thomas, 87060 Limoges, France. E-mail: samir.adly@unilim.fr

†IMAG, Université Montpellier, CNRS, Place Eugène Bataillon, 34095 Montpellier CEDEX 5, France. E-mail: hedy.attouch@umontpellier.fr, Supported by COST Action: CA16228

‡Laboratoire XLIM, Université de Limoges, 123, avenue Albert Thomas, 87060 Limoges, France. E-mail: manhhung.le@etu.unilim.fr

<b>2</b>	<b>Study of the first order system (DRYAD)</b>	<b>6</b>
2.1	Wellposedness, and energy estimates: $f$ not necessarily convex . . . . .	6
2.2	(DRYAD) seen as the perturbed gradient flow: $f$ convex . . . . .	8
<b>3</b>	<b>A dual approach to (DRYAD)</b>	<b>9</b>
<b>4</b>	<b>Applying the time scaling and averaging techniques to (DRYAD)</b>	<b>13</b>
4.1	Time scaling . . . . .	13
4.2	Averaging . . . . .	13
<b>5</b>	<b>Applying the time scaling and averaging techniques to the dual system (DDRYAD)</b>	<b>16</b>
<b>6</b>	<b>Numerical results</b>	<b>17</b>
<b>7</b>	<b>Conclusion</b>	<b>21</b>
<b>A</b>	<b>Asymptotic convergence rates for the perturbed gradient flow</b>	<b>21</b>

## 1 Introduction

In recent years, the interplay between continuous optimization and the theory of dynamical systems has resulted in significant advancements in the field of applied mathematics. The investigation of the long-term behavior of inertial dynamics, particularly within the context of a Hilbert space for convex differentiable optimization, has become a focal point. In this paper, we delve into a new layer of complexity by considering threshold effects associated with dry friction in the framework of inertial dynamics. We lay our foundation on a doubly nonlinear first-order evolution equation that involves two potentials. The differentiable function  $f$  to be minimized interacts with the system's state via its gradient and the nonsmooth dry friction potential  $\varphi = r\|\cdot\|$ ,  $r > 0$ , that operates on a linear combination of the velocity vector and the gradient of  $f$  through its convex subdifferential. These two potential components interplay to shape the dynamics of the system. In order to shed light on the centrality of  $\nabla f(x)$ , we adopt a dual formulation approach, featuring a Riemannian gradient structure, thus providing a deeper insight into the dynamics of the system. Building on the general acceleration method proposed by Attouch, Bot, and Nguyen [14], and recently extended by Adly and Attouch [4] to dry friction, our methodology incorporates time scaling and averaging of a first-order continuous differential equation. These techniques pave the way for obtaining fast convergence results for second-order time-evolution systems that include dry friction, asymptotic vanishing damping, and Hessian-driven damping in an implicit form. In this paper, we develop these concepts, provide mathematical proofs in support of our results and illustrate this through numerical simulations. We believe that these new results can contribute to the understanding and development of accelerated gradient methods from the continuous time perspective, potentially providing valuable insight into intricate optimization problems. Let us just briefly recall some facts about previous related works and at the same time highlight differences between those and our paper. First, acting as the basis upon which our current paper is built is the work by Attouch, Bot, and Nguyen [14] on the acceleration of first order dynamics via the time scaling and averaging techniques. The authors in that paper provide a generic approach by which second order dynamics with improved convergence properties can be deduced from first order ones; what is notable is while one needs to develop a Lyapunov analysis for the convergence of the first order dynamic, the improved convergence properties of the resulting second order dynamic obtained by the the time scaling and averaging techniques can be yielded solely by the differential and integral calculus. Making use of this acceleration approach, we, in this paper, manage to speed up the convergence of a doubly nonlinear evolution equation that involves the presence of dry friction. However, the addition we make, which is

not performed in the original paper [14], is that we further propose a dual approach to the initial evolution system by introducing a dual dynamic with the function variable being the gradient of the function to be minimized  $f$ . The study of the dual dynamic makes it possible to have a better understanding of the properties of the gradient of  $f$ . This dual dynamic, which has a Riemannian gradient structure, further yields a second order dynamic with accelerated convergence rates via the time scaling and averaging techniques. Considering the dual approach is, in fact, initiated by Adly and Attouch in [4] where they also study a doubly nonlinear evolution system involving dry friction which turns out to be a special case of our first order dynamic. What distinguishes our work with [4] is largely in the first order dynamic itself. Equipped with a slightly different dry friction term, our first order dynamic improves that of [4] in the sense that the limit point of the solution trajectory is now the exact critical point, not just an approximate one. This difference in the dynamic will be precisely indicated shortly.

Throughout this paper,  $\mathcal{H}$  is a real Hilbert space equipped with the scalar product  $\langle \cdot, \cdot \rangle$  and the associated norm  $\| \cdot \|$ . We first look at the first-order evolution equation.

$$\text{(DRYAD)} \quad \gamma(\dot{x}(t) + \beta \nabla f(x(t))) + \partial \varphi(\dot{x}(t) + \beta \nabla f(x(t))) + \nabla f(x(t)) \ni 0, \quad t \in [t_0, \infty)$$

that is a doubly nonlinear dynamic that involves two potentials.

We make the following standing assumptions on the two potentials  $f$  and  $\varphi$ .

$$\begin{cases} f : \mathcal{H} \rightarrow \mathbb{R} \text{ is a continuously differentiable function which is bounded from below.} \\ \nabla f \text{ is Lipschitz continuous on the bounded sets of } \mathcal{H}. \\ \varphi : \mathcal{H} \rightarrow \mathbb{R} \text{ satisfies } \varphi(x) = r\|x\| \text{ for some } r > 0 \text{ and } \gamma > 0, \beta \geq 0. \end{cases}$$

This doubly nonlinear differential inclusion contains the term  $\partial \varphi(\dot{x}(t) + \beta \nabla f(x(t)))$  attached to dry friction, hence the abbreviation (DRYAD) for Dry friction Acting Doubly. The case  $\beta = 0$  and  $\gamma = 1$ , was studied in [4]. It's worth noticing that the basic starting dynamic for the majority of gradient methods in optimization is the steepest descent method. The first potential, designated as  $f$ , affects the system's state via its gradient and is a differentiable function to be minimized. The velocity vector is affected by the second potential  $\varphi = r\| \cdot \|$ . The study of the associated dynamics' asymptotic behavior is significantly altered by the presence of this nonsmooth dry friction potential.

One distinctive characteristic of (DRYAD) is the inclusion of the dry friction term  $\partial \varphi(\dot{x}(t) + \beta \nabla f(x(t)))$ , which incorporates both the velocity vector and the gradient of  $f$ . This differentiation sets it apart from previously studied dynamics, where the dry friction term exclusively involves the velocity vector. Although seemingly straightforward, this modification significantly alters the dynamics in comparison to those investigated in [2–4, 6]. An advantageous aspect of representing the dry friction term in this new form is that each trajectory generated by (DRYAD) converges towards a critical point of  $f$ , specifically a minimizer in the case of convex  $f$ . In fact, any stationary point  $x_\infty$  of the dynamic (DRYAD) satisfies  $\partial \varphi(\beta \nabla f(x_\infty)) + (1 + \gamma \beta) \nabla f(x_\infty) \ni 0$ . This condition is equivalent to  $\beta \nabla f(x_\infty) = \text{prox}_{\frac{\beta}{1+\gamma\beta} \varphi}(0)$ , which, in combination with the dry friction property (DF)<sub>r</sub>, implies that  $\nabla f(x_\infty) = 0$  if  $\beta > 0$  (see Lemma 1.1). Thus,  $x_\infty$  corresponds to a critical point of  $f$ . In contrast, in the case  $\beta = 0$ , each trajectory generated by the dynamic converges towards an ‘‘approximate’’ critical point  $x_\infty$  of  $f$ , characterized by  $-\nabla f(x_\infty) \in \partial \varphi(0)$ .

To emphasize the role played by the gradient, we also examine the dual approach that involves the dual variable  $g(x) = \nabla f(x)$ , and the corresponding evolution reads

$$\nabla^2 f^*(g(t)) \dot{g}(t) + \frac{\gamma \beta + 1}{\gamma} g(t) - \frac{1}{\gamma} \text{proj}_{B(0,r)}(g(t)) = 0,$$

thus making appear the Riemannian structure associated with the Hessian of the convex Fenchel conjugate function  $f^*$  (when this function is assumed of class  $C^2$ ) associated with  $f$ . Here,  $\text{proj}_{B(0,r)}$  denotes

the projection operator onto the closed ball  $B(0, r)$ . Our first investigation focuses on the convergence properties of the trajectories produced by the primal evolution system (DRYAD) and its dual.

Next, we leverage the universal acceleration approach developed by Attouch, Bot, and Nguyen [14], wherein they employ a time scaling technique on a first-order continuous differential equation and subsequently apply the method of averaging. These techniques give a second-order evolution system when applied to (DRYAD)

$$\ddot{z}(s) + \frac{\alpha}{s}\dot{z}(s) + \frac{\gamma\beta + 1}{\gamma}\nabla f\left(z(s) + \frac{s}{\alpha - 1}\dot{z}(s)\right) + \frac{1}{\gamma}\nabla\varphi_{\frac{s}{\gamma(\alpha-1)}}\left(-\frac{s}{\gamma(\alpha-1)}\nabla f\left(z(s) + \frac{s}{\alpha-1}\dot{z}(s)\right)\right) = 0,$$

that involves dry friction aspects (smoothly via the gradient of the Moreau envelope  $\nabla\varphi_{\frac{s}{\gamma(\alpha-1)}}$  of  $\varphi$ ), asymptotically vanishing viscous damping (which is closely related to Nesterov's accelerated gradient method), and a damping term that is driven by the Hessian of  $f$  in an implicit form. Doing the same for the dual dynamic, we obtain

$$\nabla^2 f^*\left(w(s) + \frac{s}{\alpha - 1}\dot{w}(s)\right)\left(\ddot{w}(s) + \frac{\alpha}{s}\dot{w}(s)\right) + \nabla\Psi_\beta^*\left(w(s) + \frac{s}{\alpha - 1}\dot{w}(s)\right) = 0.$$

In the case of these inertial systems, there is no necessity to conduct a Lyapunov analysis due to the utilization of the scaling and averaging method. Instead, we exploit the convergence results of the first-order system (DRYAD) by employing techniques from differential and integral calculus. Consequently, we achieve fast convergence results for second-order time-evolution systems that incorporate dry friction, asymptotically vanishing viscous damping, and Hessian-driven damping in the implicit form.

## 1.1 Some historical facts

Let's discuss the function and significance of each damping term involved in our inertial dynamics.

### 1.1.1 Viscous friction

The term  $\gamma\dot{x}(t)$  in (DRYAD) models the viscous damping with a positive coefficient  $\gamma > 0$ . This is linked to the heavy ball with friction method of Polyak [33, 34]. Precisely, in [34] Polyak introduced the Heavy Ball with Friction method, which is based on the following inertial system with a fixed viscous damping coefficient

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

The Heavy-Ball Method (HBF) ensures exponential convergence of  $f(x(t))$  to  $\min_{\mathcal{H}} f$  for a smooth strongly convex function  $f$ . The convergence rate of (HBF) for general convex functions is  $\mathcal{O}(1/t)$ , which isn't faster than the steepest descent approach. Su-Boyd-Candès' approach of introducing a vanishing viscous damping coefficient in [38], denoted by  $\gamma(t) = \alpha/t$ , where  $\alpha$  is a positive parameter, made a substantial addition to the field. The corresponding ordinary differential equation (ODE) known as the Su-Boyd-Candès dynamic represents a continuous surrogate of the Nesterov accelerated gradient (NAG) method and is given by

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

We have the inversely quadratic convergence rate of the values  $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}(1/t^2)$  for any trajectory  $x(t)$  of  $(\text{AVD})_\alpha$  with  $\alpha \geq 3$ . The viscous damping coefficient  $\frac{\alpha}{t}$  vanishes (tends to zero) as time  $t$  approaches infinity, hence the terminology Asymptotic Vanishing Damping. The convergence properties of the dynamic  $(\text{AVD})_\alpha$  have been the subject of many recent studies, see [8, 9, 15–18, 20, 23, 24, 32, 38]. The case where the parameter  $\alpha = 3$  is crucial since it matches Nesterov's historical

algorithm. With the exception of the one dimensional case, where convergence of the trajectories has been demonstrated [18], the question of whether the trajectories converge in this case is still unanswered. According to Attouch-Chbani-Peypouquet-Redont [17], each trajectory weakly converges to a minimizer of  $f$  for values  $\alpha > 3$ . The corresponding algorithmic result was obtained by Chambolle-Dossal [27]. Furthermore, it has been proved in [20] and [32] that for  $\alpha > 3$ , the asymptotic convergence rate of the values is actually  $o(1/t^2)$ . Apidopoulos-Aujol-Dossal [11] and Attouch-Chbani-Riahi [18] investigated the subcritical situation where  $\alpha < 3$  and showed that the convergence rate of the objective values is  $\mathcal{O}(t^{-\frac{2\alpha}{3}})$ . These rates are optimal, which means they can be reached or approached arbitrarily closely.

### 1.1.2 Dry friction

Following [2, 3, 5], we say that the potential function  $\varphi$  satisfies the dry friction property  $(DF)_r$ ,  $r > 0$ , if the following properties are satisfied:

$$(DF)_r \quad \begin{cases} \varphi : \mathcal{H} \rightarrow \mathbb{R}_+ \text{ is convex continuous,} \\ \min_{\xi \in \mathcal{H}} \varphi(\xi) = \varphi(0) = 0, \\ \varphi(\xi) \geq r\|\xi\| \quad \forall \xi \in \mathcal{H}. \end{cases}$$

The function  $\varphi(x) = r\|x\|$ ,  $r > 0$  is a model example of potential which satisfies the dry friction property, which will be used throughout this paper. An important property associated with dry friction is stated in the lemma below (see [2, 3, 5] for further details).

**Lemma 1.1** *Suppose that  $\varphi : \mathcal{H} \rightarrow \mathbb{R}_+$  satisfies  $(DF)_r$ . Then one has  $\overline{\mathbb{B}}(0, r) \subset \partial\varphi(0)$ , and therefore*

$$\|x\| \leq \lambda r \implies \text{prox}_{\lambda\varphi}(x) = 0.$$

In the above formula,  $\text{prox}_{\varphi} : \mathcal{H} \rightarrow \mathcal{H}$  denotes the proximal mapping associated with the convex function  $\varphi$ . Recall that, for any  $x \in \mathcal{H}$ , for any  $\lambda > 0$

$$\text{prox}_{\lambda\varphi}(x) = \text{argmin}_{\xi \in \mathcal{H}} \left\{ \lambda\varphi(\xi) + \frac{1}{2}\|x - \xi\|^2 \right\}.$$

For a thorough background on convex analysis in Hilbert spaces, we refer to [25].

Lemma 1.1 establishes a thresholding property for the proximal operator associated with a dry friction potential.

Dry friction holds significant importance in the realm of mechanics as it induces stabilization of mechanical systems within finite time. This stands in contrast to viscous damping, which tends to produce numerous small oscillations asymptotically. Consequently, dry friction serves as an appealing tool for optimization purposes. Although the use of dry friction in optimization is a relatively recent topic, initial findings regarding the property of finite convergence under the influence of dry friction were obtained by Adly, Attouch, and Cabot [6]. Corresponding results for Partial Differential Equations have been established in [10, 28, 29, 36].

### 1.1.3 Hessian-driven damping

The combination of viscous friction with dry friction and Hessian driven damping has been considered by Adly and Attouch in [2, 3, 5]. The Hessian driven damping has a natural connection with the strong damping property in mechanics and physics, see [30]. It helps to control and attenuate the oscillation effects that occur naturally with inertial systems. Recent research has concentrated on the inertial dynamic

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

which combines asymptotic vanishing damping with Hessian-driven damping. The corresponding algorithms involve a correcting term in the Nesterov accelerated gradient method which reduces the oscillatory aspects, see Attouch-Peypouquet-Redont [21], Attouch-Chbani-Fadili-Riahi [22], Shi-Du-Jordan-Su [37]. Related to this is the Inertial System with Implicit Hessian Damping

$$\text{(SIHD)} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t) + \beta(t)\dot{x}(t)) = 0,$$

considered by Alecsa-László-Pinta in [7], see also Attouch-Fadili-Kungurtsev [19] in the perturbed case. The justification for using the term “implicit” stems from the observation that through Taylor expansion (as  $t \rightarrow \infty$  we obtain  $\dot{x}(t) \rightarrow 0$ ) one has

$$\nabla f(x(t) + \beta(t)\dot{x}(t)) \approx \nabla f(x(t)) + \beta(t)\nabla^2 f(x(t))\dot{x}(t),$$

hence making the Hessian damping appear indirectly.

## 1.2 Contents

The structure of this paper is as follows. In Section 2, we study the first order system (DRYAD), where we show the wellposedness of the system, energy estimates and some convergence properties. A dual approach to (DRYAD) is studied in Section 3. Then, in Section 4, we use the time scaling and averaging technique to (DRYAD) to obtain an inertial dynamic with accelerated convergence results. We also employ these techniques for the dual dynamic of (DRYAD) in Section 5. We illustrate our theoretical results with some numerical examples in Section 6.

## 2 Study of the first order system (DRYAD)

### 2.1 Wellposedness, and energy estimates: $f$ not necessarily convex

Recall that our approach is based on the dynamical system (DRYAD)

$$\text{(DRYAD)} \quad \gamma(\dot{x}(t) + \beta\nabla f(x(t))) + \partial\varphi\left(\dot{x}(t) + \beta\nabla f(x(t))\right) + \nabla f(x(t)) \ni 0,$$

which is a doubly nonlinear evolution equation.

Let us first observe that the Cauchy problem associated with (DRYAD) is well posed. In fact, we can rewrite (DRYAD) as follows

$$\left(I + \frac{1}{\gamma}\partial\varphi\right)(\dot{x}(t) + \beta\nabla f(x(t))) \ni -\frac{1}{\gamma}\nabla f(x(t)).$$

This is equivalent to

$$\dot{x}(t) + \beta\nabla f(x(t)) = \text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma}\nabla f(x(t))\right), \quad (2.1)$$

where  $\text{prox}_{\frac{1}{\gamma}\varphi}$  denotes the proximal operator which is single-valued since  $\varphi : \mathcal{H} \rightarrow \mathbb{R}$  is a convex function (hence  $\partial\varphi$  is a maximally monotone operator). Set  $T : \mathcal{H} \rightarrow \mathcal{H}$  defined by  $T(y) = -\beta y + \text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma}y\right)$ , Equation (2.1) can be cast under the form

$$\dot{x}(t) = T(\nabla f(x(t))) = F(x(t)),$$

where  $F = T \circ \nabla f$ . Since  $T$  is globally Lipschitz continuous and  $\nabla f$  is Lipschitz continuous on the bounded sets,  $F$  is Lipschitz continuous on the bounded sets. This property guarantees the existence and

uniqueness of a local solution to (DRYAD) according to the classical Cauchy-Lipschitz theorem. To pass from a local to a global solution, we need the following energy estimates

$$\gamma \int_{t_0}^t \|y(s)\|^2 ds + r \int_{t_0}^t \|y(s)\| ds + \beta \int_{t_0}^t \|\nabla f(x(s))\|^2 ds + f(x(t)) \leq f(x(t_0)),$$

where  $y(t) = \dot{x}(t) + \beta \nabla f(x(t))$ .

This, combined with  $f$  being bounded below, classically implies the global existence property. In order to achieve this energy estimate, we proceed as follows. Taking the scalar product of (DRYAD) with  $y(t)$ , we obtain

$$\gamma \|y(t)\|^2 + \langle \partial\varphi(y(t)), y(t) \rangle + \langle \nabla f(x(t)), \dot{x}(t) + \beta \nabla f(x(t)) \rangle = 0$$

As a property of dry friction, we have

$$\langle \partial\varphi(y(t)), y(t) \rangle \geq r \|y(t)\|.$$

Therefore,

$$\gamma \|y(t)\|^2 + r \|y(t)\| + \beta \|\nabla f(x(t))\|^2 + \frac{d}{dt} f(x(t)) \leq 0.$$

Taking the integration from  $t_0$  to  $t$  yields the energy estimate. We summarize what has just been shown in the following theorem.

**Theorem 2.1** *Given an arbitrary  $x_0 \in \mathcal{H}$ , there exists a unique global solution trajectory  $x : [t_0, \infty) \rightarrow \mathcal{H}$  such that  $x(t_0) = x_0$  to the system (DRYAD). Furthermore, we have the following properties*

- $t \mapsto f(x(t))$  is decreasing
- $\int_{t_0}^{\infty} \|\dot{x}(t) + \beta \nabla f(x(t))\|^2 dt < \infty$
- $\int_{t_0}^{\infty} \|\dot{x}(t) + \beta \nabla f(x(t))\| dt < \infty$
- $\int_{t_0}^{\infty} \|\nabla f(x(t))\|^2 dt < \infty$ .

Taking advantage of the following form of (DRYAD)

$$\dot{x}(t) + \beta \nabla f(x(t)) = \text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma} \nabla f(x(t))\right),$$

it is easy to show, using Lemma 1.1, that if  $\|\nabla f(x(t))\| \leq r$  then the system (DRYAD) becomes the gradient flow (continuous steepest descent method), that is,  $\dot{x}(t) + \beta \nabla f(x(t)) = 0$ . We will see later that  $\nabla f(x(t))$  tends to zero as  $t$  tends to  $+\infty$ . There is therefore a change in the nature of the dynamic after a certain time, going from a doubly nonlinear evolution equation to the gradient flow (without perturbation). Since the acceleration of gradient flow is well understood, it is interesting to examine the new dynamics and their convergence properties attached to our approach.

We initially examine the primal problem and subsequently explore its dual approach.



## 2.2 (DRYAD) seen as the perturbed gradient flow: $f$ convex

Let us start from the equivalent formulation of (DRYAD) given by

$$\dot{x}(t) + \beta \nabla f(x(t)) = \text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma} \nabla f(x(t))\right).$$

Let us show that the right hand side of the above equality, defined by  $g(t) := \text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma} \nabla f(x(t))\right)$ , satisfies

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

Indeed we have

$$\int_{t_0}^{\infty} \|g(t)\| dt = \int_{\|\nabla f(x(t))\| \leq r} \|g(t)\| dt + \int_{\|\nabla f(x(t))\| > r} \|g(t)\| dt.$$

On the set  $\{\|\nabla f(x(t))\| \leq r\}$ , we have according to Lemma 1.1

$$g(t) = \text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma} \nabla f(x(t))\right) = 0.$$

Hence,

$$\int_{t_0}^{\infty} \|g(t)\| dt = \int_{\|\nabla f(x(t))\| > r} \|g(t)\| dt.$$

Then note that the proximal mapping of  $\varphi$  is nonexpansive, and is equal to zero at zero. So we have

$$\|g(t)\| = \|\text{prox}_{\frac{1}{\gamma}\varphi}\left(-\frac{1}{\gamma} \nabla f(x(t))\right)\| \leq \frac{1}{\gamma} \|\nabla f(x(t))\|.$$

On the set  $\{\|\nabla f(x(t))\| > r\}$ , we deduce that

$$\|g(t)\| \leq \frac{1}{\gamma} \|\nabla f(x(t))\| \leq \frac{1}{\gamma r} \|\nabla f(x(t))\|^2.$$

Therefore,

$$\int_{t_0}^{\infty} \|g(t)\| dt = \int_{\|\nabla f(x(t))\| > r} \|g(t)\| dt \leq \frac{1}{\gamma r} \int_{\|\nabla f(x(t))\| > r} \|\nabla f(x(t))\|^2 dt.$$

Finally we get,

$$\int_{t_0}^{\infty} \|g(t)\| dt \leq \frac{1}{\gamma r} \int_{t_0}^{\infty} \|\nabla f(x(t))\|^2 dt.$$

According to Theorem 2.1 we have ,

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

Thus, (DRYAD) is the gradient flow with a right-hand side in  $\mathbb{L}^1(t_0, +\infty)$ , which classically preserves the convergence properties of the gradient flow. We refer to [26] for further details on this topic.

We have established that (DRYAD) can be regarded as a perturbed gradient flow with the perturbation belonging to  $\mathbb{L}^1(t_0, +\infty)$ . As a result, it exhibits the classical convergence properties observed in the gradient flow literature [26]. Additionally, as we will prove in the subsequent section, (DRYAD) eventually transforms into the unperturbed gradient flow after a finite time. Consequently, it inherits all the convergence rates detailed in Theorem A.1.

**Remark 2.1** It is an open question to obtain convergence rates for the perturbed gradient flow equation (A.1) under the sole assumption  $\int_{t_0}^{+\infty} \|g(t)\| dt < +\infty$ . We know that there is convergence of the trajectories, but in the above result, to get convergence rates, we need also to use the energy assumption  $\int_{t_0}^{+\infty} t \|g(t)\|^2 dt < +\infty$ .

### 3 A dual approach to (DRYAD)

Examining the dual dynamic of (DRYAD) can help us better comprehend the convergence properties of gradients, which is of fundamental importance in (DRYAD).

To begin with, let us recall our dynamical system

$$\gamma(\dot{x}(t) + \beta\nabla f(x(t))) + \partial\varphi(\dot{x}(t) + \beta\nabla f(x(t))) + \nabla f(x(t)) \ni 0.$$

Set  $\Psi(x) = \frac{\gamma}{2}\|x\|^2 + \varphi(x)$ . We have  $\partial\Psi(x) = (\gamma I + \partial\varphi)(x)$ . We transform the original system as follows

$$\begin{aligned} A &:= \gamma(\dot{x}(t) + \beta\nabla f(x(t))) + \partial\varphi(\dot{x}(t) + \beta\nabla f(x(t))) + \nabla f(x(t)) \ni 0, \\ &\iff -\nabla f(x(t)) \in \partial\Psi(\dot{x}(t) + \beta\nabla f(x(t))) \\ &\iff \nabla f(x(t)) \in \partial\Psi(-\dot{x}(t) - \beta\nabla f(x(t))), \\ &\iff -\dot{x}(t) - \beta\nabla f(x(t)) \in \partial\Psi^*(\nabla f(x(t))) \\ &\iff \dot{x}(t) + \partial G(\nabla f(x(t))) \ni 0, \text{ where } G(x) = \frac{\beta}{2}\|x\|^2 + \Psi^*(x) \\ &\iff -\dot{x}(t) \in \partial G(\nabla f(x(t))) \\ &\iff \dot{x}(t) \in \partial G(-\nabla f(x(t))) \\ &\iff -\nabla f(x(t)) \in \partial G^*(\dot{x}(t)) \\ &\iff \partial G^*(\dot{x}(t)) + \nabla f(x(t)) \ni 0 \end{aligned}$$

We have  $G^* = (\frac{\beta}{2}\|\cdot\|^2 + \Psi^*)^*$  which is exactly the Moreau envelop of  $\Psi$ , denoted by  $\Psi_\beta$ . Hence, we have the dual dynamical system of (DRYAD) having the following form

$$\nabla\Psi_\beta(\dot{x}(t)) + \nabla f(x(t)) = 0.$$

Set  $g(t) = \nabla f(x(t))$ , and the idea is to transform this dynamical system such that the left hand side is a function of  $g(t)$ . To this end, we have

$$\begin{aligned} \nabla\Psi_\beta(\dot{x}(t)) + g(t) = 0 &\iff -g(t) = \nabla\Psi_\beta(\dot{x}(t)) \\ &\iff \dot{x}(t) = \nabla\Psi_\beta^*(-g(t)). \end{aligned}$$

Assume that  $f$  is a convex function, it follows that

$$g(t) = \nabla f(x(t)) \iff x(t) \in \partial f^*(g(t)).$$

Therefore,

$$\frac{d}{dt}(\partial f^*(g(t))) - \nabla\Psi_\beta^*(-g(t)) \ni 0.$$

Since  $\Psi_\beta = (\frac{\beta}{2}\|\cdot\|^2 + \Psi^*)^*$ , we have  $\Psi_\beta^* = \frac{\beta}{2}\|\cdot\|^2 + \Psi^*$  and hence  $\nabla\Psi_\beta^*(x) = \beta x + \nabla\Psi^*(x)$ . Let us recall that  $\Psi(x) = \frac{\gamma}{2}\|x\|^2 + \varphi(x)$ . Based on the specific form of the dry friction, we can compute the Fenchel conjugate of  $\Psi$  as follows

$$\Psi^*(x) = \frac{1}{2\gamma}\text{dist}^2(x, B(0, r)).$$

Hence, we can compute its gradient  $\nabla\Psi^*(x) = \frac{1}{\gamma}(x - \text{proj}_{B(0,r)}(x))$ . Therefore

$$\nabla\Psi_\beta^*(x) = \beta x + \nabla\Psi^*(x) = \beta x + \frac{1}{\gamma}(x - \text{proj}_{B(0,r)}(x)) = \frac{\gamma\beta + 1}{\gamma}x - \frac{1}{\gamma}\text{proj}_{B(0,r)}(x).$$

Plugging this into  $\frac{d}{dt}(\partial f^*(g(t))) - \nabla\Psi_\beta^*(-g(t)) \ni 0$ , we have

$$\frac{d}{dt}(\partial f^*(g(t))) + \frac{\gamma\beta + 1}{\gamma}g(t) - \frac{1}{\gamma}\text{proj}_{B(0,r)}(g(t)) \ni 0. \quad (3.1)$$

Assume that  $f^*$  is of class  $C^2$ , this system can be equivalently written as

$$\text{(DDRYAD)} \quad \nabla^2 f^*(g(t))\dot{g}(t) + \frac{\gamma\beta + 1}{\gamma}g(t) - \frac{1}{\gamma}\text{proj}_{B(0,r)}(g(t)) = 0,$$

thus making appear the Riemannian structure associated with the Hessian of the convex function  $f^*$ . Let us summarize the above results in the following statement, and establish the convergence rates of the gradients.

**Theorem 3.1** *Let  $x : [t_0, \infty) \rightarrow \mathcal{H}$  be a global solution trajectory of (DRYAD). Suppose that  $f$  is convex. Then  $g(t) := \nabla f(x(t))$  is a solution trajectory of the generalized Riemannian flow.*

$$\frac{d}{dt}(\partial f^*(g(t))) + \frac{\gamma\beta + 1}{\gamma}g(t) - \frac{1}{\gamma}\text{proj}_{B(0,r)}(g(t)) \ni 0.$$

Furthermore, the following convergence properties hold as  $t \rightarrow \infty$

- The function  $t \mapsto D(g(t), 0)$  is decreasing where

$$D(g(t), 0) = f^*(0) - f^*(g(t)) + \langle \nabla f^*(g(t)), g(t) \rangle$$

- $\|g(t)\| = \|\nabla f(x(t))\| = o\left(\frac{1}{\sqrt{t}}\right)$

**Proof.** First, we define the Bregman distance function

$$D(g(t), 0) = f^*(0) - f^*(g(t)) + \langle \nabla f^*(g(t)), g(t) \rangle.$$

We have  $t \mapsto D(g(t), 0)$  is nonnegative because  $f^*$  is convex.

Let us derivate  $t \mapsto D(g(t), 0)$  to get

$$\begin{aligned} \frac{d}{dt}D(g(t), 0) &= -\frac{d}{dt}f^*(g(t)) + \left\langle \frac{d}{dt}\nabla f^*(g(t)), g(t) \right\rangle + \langle \nabla f^*(g(t)), \dot{g}(t) \rangle \\ &= \left\langle \frac{d}{dt}\nabla f^*(g(t)), g(t) \right\rangle \\ &= -\frac{\gamma\beta + 1}{\gamma}\|g(t)\|^2 + \frac{1}{\gamma}\langle \text{proj}_{B(0,r)}(g(t)), g(t) \rangle, \end{aligned}$$

where the last inequality comes from the dual inclusion (3.1).

Define  $h(x) = \gamma\Psi^*(x) = \frac{1}{2}\text{dist}^2(x, B(0, r))$ . We have  $h$  is a smooth and convex function. Therefore, according to the first order characteristic of convex functions we have

$$0 \geq h(g(t)) - \langle g(t) - \text{proj}_{B(0,r)}(g(t)), g(t) \rangle.$$

Therefore,

$$\langle \text{proj}_{B(0,r)}(g(t)), g(t) \rangle = \|g(t)\|^2 - \langle g(t) - \text{proj}_{B(0,r)}(g(t)), g(t) \rangle \leq \|g(t)\|^2 - h(g(t))$$

Using this inequality gives us the following estimate of  $\frac{d}{dt}D(g(t), 0)$

$$\frac{d}{dt}D(g(t), 0) \leq -\beta\|g(t)\|^2 - \frac{1}{\gamma}h(g(t))$$

Considering the integral while acknowledging the nonnegativity of  $D(g(t), 0)$  yields

$$\int_{t_0}^t \beta\|g(s)\|^2 + \frac{1}{\gamma}h(g(s))ds \leq D(g(t_0), 0)$$

Due to the nonnegativity of  $\beta\|g(s)\|^2 + \frac{1}{\gamma}h(g(s))$ , we can take the limit as  $t$  tends to infinity and obtain

$$\int_{t_0}^{\infty} \beta\|g(s)\|^2 + \frac{1}{\gamma}h(g(s))ds \leq D(g(t_0), 0) < \infty$$

Let us now pass from this integral property to asymptotic properties. Recall that we have

$$\frac{d}{dt}(\partial f^*(g(t))) + \frac{\gamma\beta + 1}{\gamma}g(t) - \frac{1}{\gamma}\text{proj}_{B(0,r)}(g(t)) \ni 0$$

Taking the scalar product of both sides of this inclusion with  $\dot{g}(t)$ , we obtain

$$\begin{aligned} & \left\langle \frac{d}{dt}(\partial f^*(g(t))), \dot{g}(t) \right\rangle + \frac{\beta}{2} \frac{d}{dt} \|g(t)\|^2 + \frac{1}{\gamma} \langle g(t) - \text{proj}_{B(0,r)}(g(t)), \dot{g}(t) \rangle = 0 \\ & \iff \left\langle \frac{d}{dt}(\partial f^*(g(t))), \dot{g}(t) \right\rangle + \frac{\beta}{2} \frac{d}{dt} \|g(t)\|^2 + \frac{1}{\gamma} \langle \dot{h}(g(t)), \dot{g}(t) \rangle = 0 \\ & \iff \left\langle \frac{d}{dt}(\partial f^*(g(t))), \dot{g}(t) \right\rangle + \frac{\beta}{2} \frac{d}{dt} \|g(t)\|^2 + \frac{1}{\gamma} \frac{d}{dt} h(g(t)) = 0 \end{aligned}$$

Due to the convexity of  $f^*$ , we have  $\langle \frac{d}{dt}(\partial f^*(g(t))), \dot{g}(t) \rangle = \langle \nabla^2 f^*(g(t))\dot{g}(t), \dot{g}(t) \rangle \geq 0$ . Combining this with the last equality, we have

$$\frac{d}{dt} \left( \frac{\beta}{2} \|g(t)\|^2 + \frac{1}{\gamma} h(g(t)) \right) \leq 0,$$

which means that the function  $t \mapsto \frac{\beta}{2} \|g(t)\|^2 + \frac{1}{\gamma} h(g(t))$  is decreasing. On the other hand, we have

$$\int_{t_0}^{\infty} \frac{\beta}{2} \|g(s)\|^2 + \frac{1}{\gamma} h(g(s)) ds \leq \int_{t_0}^{\infty} \beta \|g(s)\|^2 + \frac{1}{\gamma} h(g(s)) ds < \infty.$$

Therefore we obtain that

$$\lim_{t \rightarrow \infty} t \left( \frac{\beta}{2} \|g(t)\|^2 + \frac{1}{\gamma} h(g(t)) \right) = 0.$$

Or equivalently (due to the nonnegativity of the underlying functions and the fact that  $h(g(t)) \leq \frac{1}{2} \|g(t)\|^2$ )

$$\lim_{t \rightarrow \infty} t \|g(t)\|^2 = \lim_{t \rightarrow \infty} t \|\nabla f(x(t))\|^2 = 0, \text{ or } \|g(t)\| = \|\nabla f(x(t))\| = o\left(\frac{1}{\sqrt{t}}\right).$$

The last conclusion is obtained due to the following lemma (see [1, Lemma 5.2]).

**Lemma 3.1** *Let  $h : [t_0, \infty] \rightarrow \mathbb{R}^+$  be a nonincreasing function belonging to  $\mathbb{L}^1([t_0, \infty])$ . Then it holds that  $\lim_{t \rightarrow +\infty} th(t) = 0$ .*

The proof of Theorem 3.1 is thereby completed. ■

As a consequence we have the following corollary.

**Corollary 3.1** *Suppose that  $f : \mathcal{H} \rightarrow \mathbb{R}$  is a convex differentiable function that satisfies  $S = \operatorname{argmin} f \neq \emptyset$ . Let  $x : [t_0, +\infty[ \rightarrow \mathcal{H}$  be a solution trajectory of (DRYAD). Then the following statements are satisfied:*

(i) *(convergence of gradients towards zero)  $\|\nabla f(x(t))\| = o\left(\frac{1}{\sqrt{t}}\right)$  as  $t \rightarrow +\infty$ .*

(ii) *(integral estimate of the gradients)  $\int_{t_0}^{+\infty} t \|\nabla f(x(t))\|^2 dt < +\infty$ .*

(iii) *(convergence of values)  $f(x(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t}\right)$  as  $t \rightarrow +\infty$ .*

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

**Proof.** It follows from Theorem 3.1 that  $\nabla f(x(t))$  approaches zero as  $t$  approaches infinity. As previously stated, when  $\|\nabla f(x(t))\| \leq r$ , the system becomes the gradient flow. Consequently, there exists a  $T > 0$  such that for all  $t \geq T$ , the system is the gradient flow. Therefore, any solution trajectory generated by (DRYAD) inherits all the convergence rates of the continuous steepest descent method listed in Theorem A.1 (in the appendix). ■

When  $\mathcal{H}$  is a finite dimensional Euclidian space, we have the following corollary concerning the convergence property (DRYAD) when the objective function  $f$  is not necessarily convex but satisfies the Kurdyka–Lojasiewicz property. Let us recall some basic facts concerning the Kurdyka–Lojasiewicz property, which we briefly designate by (KL). No convexity assumption is made on the function  $f$  to be minimized. A function  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  satisfies the (KL) property if its values can be reparametrized in the neighborhood of each of its critical points, so that the resulting function becomes sharp. This means that there exists a continuous, concave, increasing function  $\theta$  such that for all  $u$  in a slice of  $f$ , we have

$$\|\nabla(\theta \circ f)(u)\| \geq 1.$$

The function  $\theta$  captures the geometry of  $f$  around its critical points, and is called a desingularizing function; see [12, 13], for further details.

**Corollary 3.2** *Suppose that  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  is a differentiable function that satisfies the (KL) property. Then, any bounded solution trajectory of (DRYAD) has a finite length and hence converges to a critical point of  $f$ , as  $t \rightarrow \infty$*

**Proof.** Since (DRYAD) arrives at the regime of the gradient flow from a sufficiently large time, the statement of the corollary follows from the result of Lojasiewicz [31] ■

**Remark 3.1** Let us present a dual viewpoint on the finite time stabilization property. Since  $\nabla f(x(t))$  converges to zero as  $t \rightarrow \infty$ , from a sufficiently large time  $T$ , we have  $\nabla f(x(t)) \in B(0, r)$  and hence  $g(t) - \operatorname{proj}_{B(0, r)}(g(t)) = 0$ . As a result, the dual dynamical system becomes

$$\frac{d}{dt}(\partial f^*(g(t))) + \beta g(t) = 0,$$

or equivalently

$$\nabla^2 f^*(g(t))\dot{g}(t) + \beta g(t) = 0,$$

when  $f^*$  is assumed to be twice continuously differentiable.

## 4 Applying the time scaling and averaging techniques to (DRYAD)

Before going further, let us mention that these techniques of scaling and averaging were initiated by At-touch, Bot and Nguyen [14] for the gradient flow method. In our specific case, we have adapted these techniques to address our problem, and we now recapitulate only the essential elements required for our analysis.

### 4.1 Time scaling

The time scaling technique is in fact a change of variable  $t = \tau(s)$ , where  $\tau$  is an increasing function from  $\mathbb{R}^+$  to  $\mathbb{R}^+$  which is continuously differentiable and which tends to  $\infty$  when  $s \rightarrow \infty$ . Set

$$y(s) = x(\tau(s))$$

We have

$$\dot{y}(s) = \dot{\tau}(s)\dot{x}(\tau(s))$$

As a result, to obtain the corresponding dynamical system associated with the new trajectory  $y(s)$ , we respectively replace  $x(t)$  and  $\dot{x}(t)$  in the original system with  $y(s)$  and  $\frac{\dot{y}(s)}{\dot{\tau}(s)}$ . To this end, we obtain

$$\gamma \left( \frac{\dot{y}(s)}{\dot{\tau}(s)} + \beta \nabla f(y(s)) \right) + \partial \varphi \left( \frac{\dot{y}(s)}{\dot{\tau}(s)} + \beta \nabla f(y(s)) \right) + \nabla f(y(s)) \ni 0.$$

Using the positive homogeneity of degree zero of  $\partial \varphi$ , this system can be simplified as

$$\dot{y}(s) + \frac{\dot{\tau}(s)}{\gamma} \partial \varphi \left( \dot{y}(s) + \beta \dot{\tau}(s) \nabla f(y(s)) \right) + \frac{\gamma \beta + 1}{\gamma} \dot{\tau}(s) \nabla f(y(s)) \ni 0 \quad (4.1)$$

### 4.2 Averaging

Let us attach to  $y(\cdot)$  the new function  $z : [s_0, \infty] \rightarrow \mathcal{H}$  define by

$$\dot{z}(s) + \frac{1}{\dot{\tau}(s)} (z(s) - y(s)) = 0,$$

with  $z(s_0) = y(s_0) = x_0$  given in  $\mathcal{H}$ . Equivalently,

$$y(s) = z(s) + \dot{\tau}(s)\dot{z}(s).$$

By temporal derivation we have

$$\dot{y}(s) = (1 + \ddot{\tau}(s))\dot{z}(s) + \dot{\tau}(s)\ddot{z}(s)$$

After plugging  $y(s)$  and  $\dot{y}(s)$  by their expressions in terms of  $z(t)$  into the dynamical system obtained from time scaling (4.1) and dividing both sides by  $\dot{\tau}(s)$  we have

$$\ddot{z}(s) + \frac{\ddot{\tau}(s) + 1}{\dot{\tau}(s)} \dot{z}(s) + \frac{\gamma \beta + 1}{\gamma} \nabla f(z(s) + \dot{\tau}(s)\dot{z}(s)) + \frac{1}{\gamma} \partial \varphi(a(s)) \ni 0 \quad (4.2)$$

where  $a(s) = \dot{y}(s) + \beta \dot{\tau}(s) \nabla f(y(s))$ .

Let us return to the dynamical system (4.1) and express it using the notation  $a(s)$ . So we have

$$a(s) + \frac{\dot{\tau}(s)}{\gamma} \partial \varphi(a(s)) + \frac{\dot{\tau}(s)}{\gamma} \nabla f(y(s)) = 0$$

Therefore

$$\begin{aligned}
\partial\varphi(a(s)) &= -\frac{\gamma}{\dot{\tau}(s)}a(s) - \nabla f(y(s)) \\
&= -\frac{\gamma}{\dot{\tau}(s)}\left(I + \frac{\dot{\tau}(s)}{\gamma}\partial\varphi\right)^{-1}\left(-\frac{\dot{\tau}(s)}{\gamma}\nabla f(y(s))\right) - \nabla f(y(s)) \\
&= \frac{\gamma}{\dot{\tau}(s)}\left[-\left(I + \frac{\dot{\tau}(s)}{\gamma}\partial\varphi\right)^{-1}\left(-\frac{\dot{\tau}(s)}{\gamma}\nabla f(y(s))\right) - \frac{\dot{\tau}(s)}{\gamma}\nabla f(y(s))\right] \\
&= \nabla\varphi_{\frac{\dot{\tau}(s)}{\gamma}}\left(-\frac{\dot{\tau}(s)}{\gamma}\nabla f(y(s))\right) \\
&= \nabla\varphi_{\frac{\dot{\tau}(s)}{\gamma}}\left(-\frac{\dot{\tau}(s)}{\gamma}\nabla f(z(s) + \dot{\tau}(s)\dot{z}(s))\right)
\end{aligned}$$

Plugging this into the dynamical system (4.2) gives

$$\ddot{z}(s) + \frac{\ddot{\tau}(s) + 1}{\dot{\tau}(s)}\dot{z}(s) + \frac{\gamma\beta + 1}{\gamma}\nabla f(z(s) + \dot{\tau}(s)\dot{z}(s)) + \frac{1}{\gamma}\nabla\varphi_{\frac{\dot{\tau}(s)}{\gamma}}\left(-\frac{\dot{\tau}(s)}{\gamma}\nabla f(z(s) + \dot{\tau}(s)\dot{z}(s))\right) = 0 \quad (4.3)$$

Finally we have obtained a second order dynamical system by doing time scaling and averaging from the original first order dynamic. In order to have fast convergence properties, we choose  $\tau$  such that the viscous damping coefficient in this dynamical system asymptotically vanishes as follows

$$\frac{\ddot{\tau}(s) + 1}{\dot{\tau}(s)} = \frac{\alpha}{s},$$

for some  $\alpha > 1$ . We can easily show that this is achieved by setting  $\tau(s) = \frac{s^2}{2(\alpha-1)}$ .

Hence, the dynamical system (4.3) becomes

$$(\text{iDRYAD}) \quad \ddot{z}(s) + \frac{\alpha}{s}\dot{z}(s) + \frac{\gamma\beta + 1}{\gamma}\nabla f\left(z(s) + \frac{s}{\alpha-1}\dot{z}(s)\right) + \frac{1}{\gamma}\nabla\varphi_{\frac{s}{\gamma(\alpha-1)}}\left(-\frac{s}{\gamma(\alpha-1)}\nabla f\left(z(s) + \frac{s}{\alpha-1}\dot{z}(s)\right)\right) = 0$$

The corresponding convergence properties for  $z(s)$  with this specific choice of  $\tau$  are captured in the following theorem

**Theorem 4.1** *Let  $f$  be a convex smooth function whose gradient is Lipschitz continuous on the bounded sets and such that  $\text{argmin} f$  is non empty. Assume  $\alpha > 3$ , let  $z : [s_0, \infty] \rightarrow \mathcal{H}$  be a solution trajectory of*

$$\ddot{z}(s) + \frac{\alpha}{s}\dot{z}(s) + \frac{\gamma\beta + 1}{\gamma}\nabla f\left(z(s) + \frac{s}{\alpha-1}\dot{z}(s)\right) + \frac{1}{\gamma}\nabla\varphi_{\frac{s}{\gamma(\alpha-1)}}\left(-\frac{s}{\gamma(\alpha-1)}\nabla f\left(z(s) + \frac{s}{\alpha-1}\dot{z}(s)\right)\right) = 0$$

*Then we have the following properties*

- $f(z(s)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/s^2)$
- $\|\nabla f(z(s))\| = \mathcal{O}(1/s)$
- $\int_{s_0}^{\infty} s^3 \|\nabla f(z(s) + \frac{s}{\alpha-1}\dot{z}(s))\|^2 ds < \infty$
- *only assume that  $\alpha > 1$ , then  $z(s)$  converges weakly and its limit belongs to  $\text{argmin} f$*

**Proof.** Since our original system follows the steepest descent method after a finite time, the results are achieved according to [14]. Let us present the main lines. First, we interpret the transition from  $y$  to  $z$  as an averaging process. Precisely, rewriting the relation between  $y$  and  $z$  we obtain

$$s\dot{z}(s) + (\alpha - 1)z(s) = (\alpha - 1)y(s)$$

Multiplying this equation with  $s^{\alpha-2}$ , we get

$$s^{\alpha-1}\dot{z}(s) + (\alpha - 1)s^{\alpha-2}z(s) = (\alpha - 1)s^{\alpha-2}y(s),$$

which is equivalent to

$$\frac{d}{ds}(s^{\alpha-1}z(s)) = (\alpha - 1)s^{\alpha-2}y(s).$$

Integrating this equation from  $s_0$  to  $s$ , we obtain

$$z(s) = \frac{s_0^{\alpha-1}}{s^{\alpha-1}}y(s_0) + \frac{\alpha - 1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2}y(u)du,$$

which can be written abstractly as

$$z(s) = \int_{s_0}^s y(u)d\mu_s(u),$$

where  $\mu_s$  is the positive Radon measure  $[s_0, s]$  defined by

$$\mu_s = \frac{s_0^{\alpha-1}}{s^{\alpha-1}}\delta_{s_0} + (\alpha - 1)\frac{u^{\alpha-2}}{s^{\alpha-1}}du,$$

where  $\delta_{s_0}$  denotes the Dirac measure at  $s_0$ . Since  $\mu_s$  is positive and has the integral over  $[s_0, s]$  being 1, it is a probability measure. It is clear that  $z(s)$  can be seen as the average of trajectory  $y(\cdot)$  on  $[s_0, s]$  with respect to  $\mu_s$ .

For the first item of the theorem, we use the Lipschitz continuity of the gradient of  $f$

$$f(z(s)) - \inf_{\mathcal{H}} f \leq f\left(\int_{s_0}^s y(u)d\mu_s(u)\right) - \inf_{\mathcal{H}} f + \mathcal{O}(1/s^2).$$

What remains is to show that

$$f\left(\int_{s_0}^s y(u)d\mu_s(u)\right) - \inf_{\mathcal{H}} f = \mathcal{O}(1/s^2),$$

which can be achieved by making use of the convexity of  $f$ , and hence the Jensen inequality.

For the second item, since  $\nabla f(x_*) = 0$  and according to [35, Theorem 2.1.5], we have

$$\frac{1}{2L}\|f(z(s))\|^2 \leq f(z(s)) - \inf_{\mathcal{H}} f.$$

This inequality combined with the first item gives us the result.

For the third item, according to Corollary 3.1, we have for any solution trajectory of (DRYAD) that

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



Making a change of variable associated with the time scaling step

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

we obtain

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

where  $y(s) = x(\tau(s))$ . In light of the averaging step, we replace  $y(s)$  with  $x(s) + \frac{s}{\alpha-1}\dot{x}(s)$  and obtain the result of this item.

For the last item which concerns the weak convergence of  $z(s)$ , we argue as follows. We know that the solution trajectory of the steepest descent dynamic converges weakly to a solution  $x_* \in S = \operatorname{argmin} f$ . This immediately gives that  $y(s) = x(\tau(s))$  converges weakly to  $x_*$ . To pass from this result to the result of  $z(\cdot)$ , we use the interpretation of  $z$  as an average of  $y$

$$z(s) = \frac{s_0^{\alpha-1}}{s^{\alpha-1}} y(s_0) + \frac{\alpha-1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2} y(u) du.$$

In order to have the weak convergence of  $z(\cdot)$ , meaning  $\langle z(s), v \rangle \rightarrow \langle x_*, v \rangle$  as  $s \rightarrow \infty$  for all  $v \in \mathcal{H}$ , after elementary calculus, it is sufficient to require that if  $a(\cdot)$  is a positive real valued function which satisfies  $\lim_{r \rightarrow \infty} a(r) = 0$ , then  $\lim_{s \rightarrow \infty} A(s) = 0$ , where

$$A(s) = \frac{\alpha-1}{s^{\alpha-1}} \int_{s_0}^s r^{\alpha-2} a(r) dr,$$

which indeed can be proven to be true. ■

## 5 Applying the time scaling and averaging techniques to the dual system (DDRYAD)

Let us recall the dual dynamical system

$$\frac{d}{dt}(\partial f^*(g(t))) + \nabla \Psi_\beta^*(g(t)) \ni 0$$

Performing similarly to the case of (DRYAD), we can apply the time scaling and averaging to the dual system to obtain the following second order dynamic

$$(iDDRYAD) \quad \nabla^2 f^* \left( w(s) + \frac{s}{\alpha-1} \dot{w}(s) \right) \left( \ddot{w}(s) + \frac{\alpha}{s} \dot{w}(s) \right) + \nabla \Psi_\beta^* \left( w(s) + \frac{s}{\alpha-1} \dot{w}(s) \right) = 0,$$

where the relation of  $g(t)$  with  $w(t)$  is as follows

$$\begin{cases} v(s) = g(\tau(s)), \\ v(s) = w(s) + \dot{\tau}(s) + \dot{w}(s) \end{cases}, \text{ with } \tau(s) = \frac{s^2}{2(\alpha-1)}$$

Here  $v$  is associated with the scaling step and  $w$  is associated with the averaging step. Recall that for the dual dynamic, we have

$$\lim_{t \rightarrow \infty} t \|g(t)\|^2 = 0$$

As  $\tau(s) \rightarrow \infty$  as  $s \rightarrow \infty$ , we can replace  $t$  with  $\tau(s)$  in the above limit to obtain

$$\lim_{s \rightarrow \infty} s \|v(s)\| = 0, \text{ or } \|v(s)\| = o(1/s)$$

The differential equation connecting  $v$  and  $w$  gives us the interpretation of  $w$  as an average of  $v$  as follows:

$$w(s) = \int_{s_0}^s v(u) d\mu_s(u),$$

where  $\mu_s = \frac{s_0^{\alpha-1}}{s^{\alpha-1}} \delta_{s_0} + (\alpha-1) \frac{u^{\alpha-2}}{s^{\alpha-1}} du$  is a probability measure on  $[s_0, s]$ . Here  $\delta_{s_0}$  denotes the Dirac measure at  $s_0$ .

According to the convexity of  $\|\cdot\|$  and the Jensen inequality, we obtain that

$$\|w(s)\| = o(1/s)$$

The following theorem summarizes the results we just showed for this second order dual dynamic

**Theorem 5.1** *Let  $x : [t_0, \infty) \rightarrow \mathcal{H}$  be a global solution trajectory of (DRYAD). Then  $g(t) := \nabla f(x(t))$  is the solution trajectory of the generalized Riemannian flow*

$$\frac{d}{dt}(\partial f^*(g(t))) + \frac{\gamma\beta + 1}{\gamma} g(t) - \frac{1}{\gamma} \text{proj}_{B(0,r)}(g(t)) \ni 0.$$

Set  $\tau(s) = \frac{s^2}{2(\alpha-1)}$  with  $\alpha > 1$ , and  $v(s) = g(\tau(s))$ . Define  $w$  as the solution of the differential equation

$$\dot{w}(s) + \frac{1}{\tau(s)}(w(s) - v(s)) = 0, \text{ with } w(s_0) = v(s_0) = x_0.$$

Then  $w$  satisfies the following inertial system

$$\nabla^2 f^* \left( w(s) + \frac{s}{\alpha-1} \dot{w}(s) \right) \left( \ddot{w}(s) + \frac{\alpha}{s} \dot{w}(s) \right) + \nabla \Psi_\beta^* \left( w(s) + \frac{s}{\alpha-1} \dot{w}(s) \right) = 0,$$

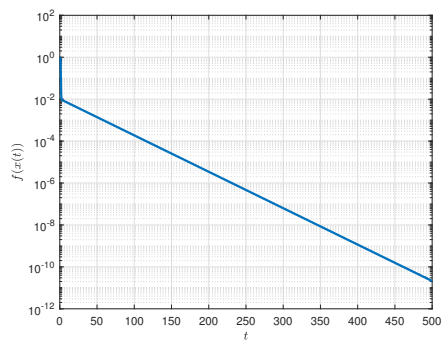
and we have  $\|w(s)\| = o(1/s)$  as  $s \rightarrow \infty$

## 6 Numerical results

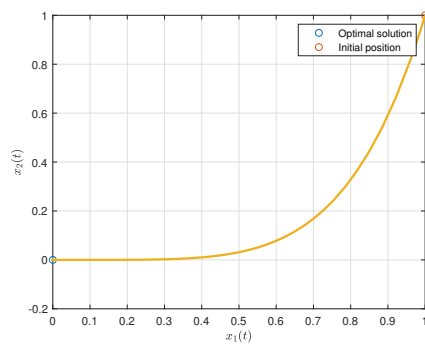
In this section, we will use adapted standard Runge-Kutta methods to solve numerically the involved continuous dynamics and conduct a series of numerical illustrative experiments to illustrate the theoretical results discussed in the previous sections.

**Example 6.1** Let us begin this section by considering an example to illustrate the dynamic (DRYAD) in dimension 2 in the case of a convex and quadratic function. More precisely, let us set  $f(x_1, x_2) = ax_1^2 + bx_2^2$  with  $0 \leq a < b$  and the initial condition  $x(1) = (1, 1)$  and  $\dot{x}(1) = (0, 0)$ . Note that  $f$  is of the form  $f(x) = \langle x, Qx \rangle$  with  $Q = \text{diag}([a, b])$ . We take  $\varphi(x) = r\|x\|_2$ , with  $r = 0.1$ .

In Figure 1 we illustrate the behaviors of several quantities associated with (DRYAD). Figure 1(a) shows the value of the objective function  $f$  along the trajectory as a function of time. We can see that the function value decreases over time which is in accordance with the theoretical result. Figure 1(b) shows the trajectory of the system starting from an initial position and finally ending up at the unique minimizer of  $f$ . The last two figures display the convergences towards zero of the velocity and gradient vectors.



(a) Behavior of the objective



(b) Solution trajectory

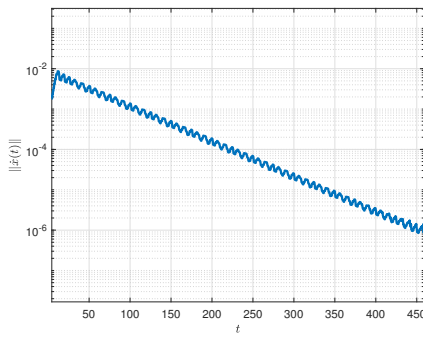
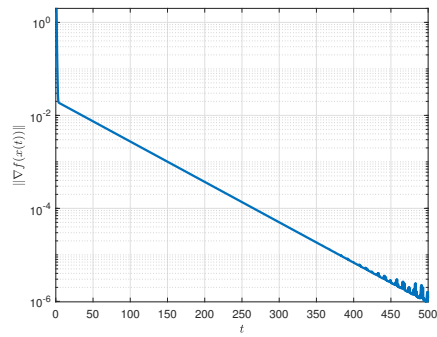
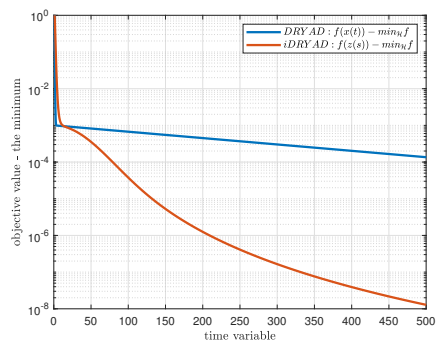
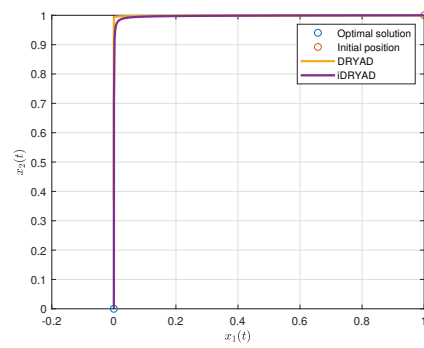
(c) Behavior of  $\|\dot{x}(t)\|$ (d) Behavior of  $\|\nabla f(x(t))\|$ 

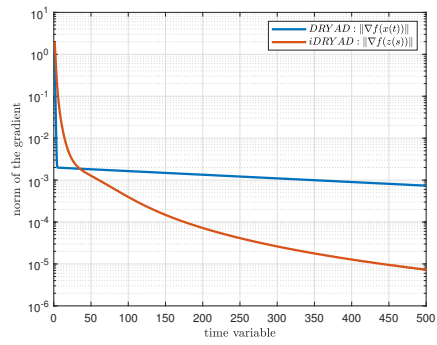
Figure 1: Illustration of the convergence results of (DRYAD).



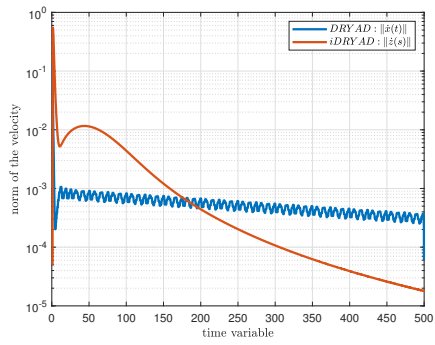
(a) Behavior of the objective



(b) Solution trajectory



(c) Behavior of the gradient



(d) Behavior of the velocity

Figure 2: Comparison between (DRYAD) and (iDRYAD)

**Example 6.2** Let us now compare the two primal dynamics (DRYAD) and (iDRYAD) on a quadratic function in dimension 2.

Similarly to Example 6.1, we consider 4 comparison criteria in Figure 2, namely the objective value, solution trajectories, norm of the gradient, and norm of the velocity. As can be seen from the figures, (iDRYAD) displays a superior performance compared to (DRYAD) which confirms our theoretical results.

**Example 6.3** We conduct the same numerical experiments as the previous example to compare the two dual dynamics (DDRYAD) and (iDDRYAD).

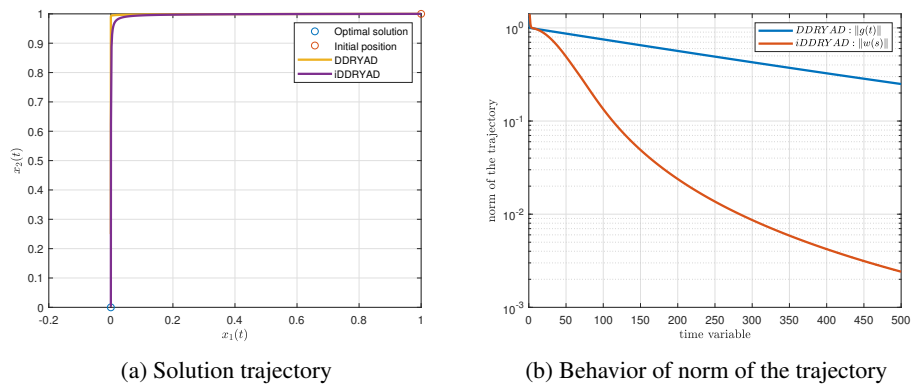


Figure 3: Comparison between (DDRYAD) and (iDDRYAD)

In this comparison, in addition to providing the solution trajectories of (DDRYAD) and (iDDRYAD) on the plane, we also present the evolutions of the norm of their trajectories, which are supposed to converge to zero according to the theoretical results. Clearly, (iDDRYAD) outperforms (DDRYAD).

**Example 6.4** Let us conclude the numerical tests with this example where we bring together the 4 dynamics, namely (DRYAD), (iDRYAD), (DDRYAD), and (iDDRYAD) into one plot. To this end, we will display the norms of the gradients of the two primal dynamics' trajectories and the norms of the two dual dynamics' trajectories. We will see the evolutions of these 4 quantities which are supposed to converge to zero as time tends to infinity. We use a quadratic problem for this numerical test.

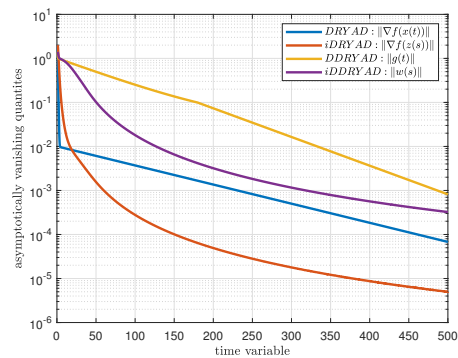


Figure 4: Comparing the four dynamics

If we look at Figure 4, in addition to the observation that the second order dynamics have faster convergences than the first order ones, which has been seen in previous examples, we can also say that the primal dynamics seem to have slightly better performances than the dual dynamics in terms of convergence to zero of their respective considered quantities

In summary, based on the above examples and the theoretical results, it is evident that inertial dynamics exhibit accelerated convergence rates for objective values, gradient norms and velocity vectors in both primal and dual contexts. These enhanced convergence properties present important advantages in the optimization field by facilitating more effective and efficient optimization processes. The ability to achieve faster convergence means that optimal or near-optimal solutions can be obtained more quickly, reducing computation time and improving resource utilization. These benefits not only improve overall optimization efficiency, but also the performance of various optimization applications. Consequently, the enhanced convergence properties of inertial dynamics make them extremely valuable and desirable tools in the optimization field.

## 7 Conclusion

In this paper, we study the long-time behavior of inertial dynamics with dry friction in a Hilbert setting for convex differentiable optimization problems. The analysis made use of the the time scaling and averaging techniques developed by Attouch, Bot and Nguyen [14] to accelerate first order dynamical systems. We initially study a doubly nonlinear first-order evolution equation, and subsequently adopt the mentioned acceleration method to obtain a second-order in time evolution system involving dry friction, asymptotically vanishing viscous damping, and a damping driven by the Hessian in the implicit form. The obtained accelerated convergence rates of the inertial dynamic do not require developing a Lyapunov analysis, but instead relies on the convergence results of the original first-order system and tools from differential and integral calculus. However, there are several questions that require more research regarding this topic. One natural direction is to study the case where the objective function  $f$  is a nonsmooth convex function by replacing the gradient terms in (DRYAD) by the subdifferential of  $f$ . Another important area is to design from (DRYAD) the associated first order dynamic representing the minimization problem of additive functions where  $f$  is the sum of a smooth and a nonsmooth function. Regarding the dual formulation of (DRYAD), the double differentiability of the Legendre-Fenchel transform  $f^*$  of the convex function  $f$  plays a crucial role in the formulation of the dual dynamic approach. This poses an issue since this assumption may not be available in practice. The numerical experiments highlight the accelerated convergence properties of inertial dynamics as opposed to their first order counterparts. While we have focused on continuous-time scenarios in this paper, it is important to explore the temporal discretization of these dynamics and examine the convergence properties of the associated algorithms. In addition, it would be interesting to carry out tests on various optimization problems at different scales. This research will enable us to gain a deeper understanding of these algorithms, their optimization efficiency and their applicability to a wider range of problem sizes.

## A Asymptotic convergence rates for the perturbed gradient flow

Let us provide asymptotic convergence rates for the perturbed gradient flow, which we have rely on in previous sections.

**Theorem A.1** *Suppose that  $f: \mathcal{H} \rightarrow \mathbb{R}$  a convex differentiable function that satisfies  $S = \operatorname{argmin} f \neq \emptyset$ . Let  $z: [t_0, +\infty[ \rightarrow \mathcal{H}$  be a solution trajectory of*

$$\dot{z}(t) + \nabla f(z(t)) = g(t) \tag{A.1}$$

where  $g: [t_0, +\infty[ \rightarrow \mathcal{H}$  is such that

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

Then the following statements are satisfied:

- (i) (convergence of gradients towards zero)  $\|\nabla f(z(t))\| = o\left(\frac{1}{\sqrt{t}}\right)$  as  $t \rightarrow +\infty$ .
- (ii) (integral estimate of the gradients)  $\int_{t_0}^{+\infty} t \|\nabla f(z(t))\|^2 dt < +\infty$ .
- (iii) (convergence of values)  $f(z(t)) - \inf_{\mathcal{H}} f = o\left(\frac{1}{t}\right)$  as  $t \rightarrow +\infty$ .
- (iv) The solution trajectory  $z(\cdot)$  converges weakly as  $t \rightarrow +\infty$ , and its limit belongs to  $S = \operatorname{argmin} f$ .

**Proof.** For sake of completeness, let us recall some of the arguments used in the asymptotic analysis of the perturbed steepest descent system when the perturbation  $g$  satisfies (A.2). Given  $z_* \in S$ , let  $T > t_0$  be fixed and for every  $t_0 \leq t \leq T$  consider

$$\mathcal{E}_T(t) := t \left( f(z(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2} \|z(t) - z_*\|^2 + \int_t^T \langle z(\tau) - z_* + \tau \dot{z}(\tau), g(\tau) \rangle d\tau.$$

Differentiating  $\mathcal{E}_T$  gives for all  $T \geq t \geq t_0$

$$\begin{aligned} \frac{d}{dt} \mathcal{E}_T(t) &= f(z(t)) - \inf_{\mathcal{H}} f + t \langle \nabla f(z(t)), \dot{z}(t) \rangle + \langle z(t) - z_*, \dot{z}(t) - g(t) \rangle - t \langle \dot{z}(t), g(t) \rangle \\ &= f(z(t)) - \inf_{\mathcal{H}} f - t \|\dot{z}(t)\|^2 - \langle z(t) - z_*, \nabla f(z(t)) \rangle \\ &\leq -t \|\dot{z}(t)\|^2, \end{aligned} \quad (\text{A.3})$$

where the second equality comes from (A.1), and the last inequality follows from the convexity of  $f$ . By integration from  $t_0$  to  $t$ , we deduce that

$$\begin{aligned} &t \left( f(z(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2} \|z(t) - z_*\|^2 \\ &\leq C + \int_{t_0}^t \langle z(\tau) - z_*, g(\tau) \rangle d\tau + \int_{t_0}^t \tau \langle \dot{z}(\tau), g(\tau) \rangle d\tau - \int_{t_0}^t \tau \|\dot{z}(\tau)\|^2 d\tau \\ &\leq C + \int_{t_0}^t \|z(\tau) - z_*\| \|g(\tau)\| d\tau + \frac{1}{2} \int_{t_0}^{+\infty} \tau \|g(\tau)\|^2 d\tau - \frac{1}{2} \int_{t_0}^t \tau \|\dot{z}(\tau)\|^2 d\tau. \end{aligned} \quad (\text{A.4})$$

We obtain the following estimate (recall that  $C$  denotes a generic constant), satisfied for all  $t \geq t_0$

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

According to Gronwall Lemma, we conclude that for all  $t \geq t_0$

$$\|z(\tau) - z_*\| \leq \sqrt{2C} + \int_{t_0}^t \|g(\tau)\| d\tau \leq \sqrt{2C} + \int_{t_0}^{+\infty} \|g(\tau)\| d\tau < +\infty.$$

The trajectory is therefore bounded. Using this property and (A.2) allows us to assert from (A.4) that

$$t \left( f(z(t)) - \inf_{\mathcal{H}} f \right) + \frac{1}{2} \|z(t) - z_*\|^2 + \frac{1}{2} \int_{t_0}^t \tau \|\dot{z}(\tau)\|^2 d\tau \leq C. \quad (\text{A.5})$$

The above estimate does not depend on  $T$ , so it is satisfied for all  $t \geq t_0$ . It immediately gives the convergence rate of the values for the solution trajectories of the perturbed (SD)

$$f(z(t)) - \inf_{\mathcal{H}} f \leq \frac{C}{t}.$$

Since  $t(f(z(t)) - \inf_{\mathcal{H}} f) + \frac{1}{2} \|z(t) - z_*\|^2 \geq 0$  for  $t \geq t_0$ , letting  $t$  goes to  $+\infty$  in (A.5), we get

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

According to the constitutive equation (A.1) we get

$$\int_{t_0}^{+\infty} t \|\nabla f(z(t))\|^2 dt \leq 2 \int_{t_0}^{+\infty} t \|\dot{z}(t)\|^2 dt + 2 \int_{t_0}^{+\infty} t \|g(t)\|^2 dt < +\infty.$$

Let us differentiate the anchor function, which is another classical ingredient of the Lyapunov analysis

$$\begin{aligned} \frac{d}{dt} \left( \frac{1}{2} \|z(t) - z_*\|^2 \right) &= \langle z(t) - z_*, \dot{z}(t) \rangle \\ &= - \langle z(t) - z_*, \nabla f(z(t)) \rangle - \langle z(t) - z_*, g(t) \rangle \\ &\leq - \left( f(z(t)) - \inf_{\mathcal{H}} f \right) + \sup_{t \geq t_0} \|z(t) - z_*\| \cdot \|g(t)\| \end{aligned} \quad (\text{A.6})$$

$$\leq \sup_{t \geq t_0} \|z(t) - z_*\| \cdot \|g(t)\|. \quad (\text{A.7})$$

Recall that the trajectory  $z(\cdot)$  is bounded. According to assumption (A.2), we deduce that the right hand side of (A.7) belongs to  $\mathbb{L}^1([t_0, +\infty[)$ . Therefore, from [1, Lemma 5.1] we obtain that

$$\lim_{t \rightarrow +\infty} \|z(t) - z_*\|^2 \in \mathbb{R} \text{ exists}$$

and so  $\lim_{t \rightarrow +\infty} \|z(t) - z_*\| \in \mathbb{R}$  does. In other words, the first condition of Opial's lemma is fulfilled. Furthermore, since  $\lim_{t \rightarrow +\infty} f(z(t)) = \inf_{\mathcal{H}} f$  and  $f$  is convex and weakly lower semicontinuous, the second condition of Opial's lemma is also fulfilled. This gives the weak convergence of the trajectory  $z(t)$  as  $t \rightarrow +\infty$  to an element in  $S = \operatorname{argmin} f$ .

Now let us show that in fact

$$\lim_{t \rightarrow +\infty} t \left( f(z(t)) - \inf_{\mathcal{H}} f \right) = 0,$$

meaning that the convergence rate of  $f(z(t)) - \inf_{\mathcal{H}} f$  is actually  $o(1/t)$ . To see this, we integrate (A.6) from  $t_0$  to  $t > t_0$  and then let  $t$  converge to  $+\infty$ . This yields

$$\int_{t_0}^{+\infty} \frac{1}{t} t \left( f(z(t)) - \inf_{\mathcal{H}} f \right) dt = \int_{t_0}^{+\infty} \left( f(z(t)) - \inf_{\mathcal{H}} f \right) dt < +\infty \quad (\text{A.8})$$



and thus  $\liminf_{t \rightarrow +\infty} t(f(z(t)) - \inf_{\mathcal{H}} f) = 0$ . It remains to show that this limit exists. To this end we compute the time derivative of  $t(f(z(t)) - \inf_{\mathcal{H}} f)$  and apply once again [1, Lemma 5.1]

$$\begin{aligned} \frac{d}{dt} \left( t \left( f(z(t)) - \inf_{\mathcal{H}} f \right) \right) &= f(z(t)) - \inf_{\mathcal{H}} f + t \langle \nabla f(z(t)), \dot{z}(t) \rangle \\ &= f(z(t)) - \inf_{\mathcal{H}} f - t \|\dot{z}(t)\|^2 - t \langle g(t), \dot{z}(t) \rangle \\ &\leq f(z(t)) - \inf_{\mathcal{H}} f + \frac{1}{4} t \|g(t)\|^2. \end{aligned}$$

Statement (iii) follows from assumption (A.2) and (A.8).

Let  $L$  be the Lipschitz constant of  $\nabla f$  on a ball containing the trajectory  $z(\cdot)$ . It follows from [22, Lemma 1] that for every  $t \geq t_0$

$$0 \leq \frac{t}{2L} \|\nabla f(z(t))\|^2 \leq t(f(z(t)) - \inf_{\mathcal{H}} f),$$

which implies that  $\lim_{t \rightarrow +\infty} t \|\nabla f(z(t))\|^2 = 0$  and proves (i). ■

**Acknowledgements.** This work was supported by a public grant from the Fondation Mathématique Jacques Hadamard.

## Declarations

### Data availability statements

Not relevant for our topic.

### Conflict of interest

The authors declare that they have no conflict of interest.

### Competing interests

The authors declare that they have no competing interests.

### Authors contributions

All authors contributed equally in writing this article. All authors read and approved the final manuscript.

## References

- [1] B. Abbas, H. Attouch, B.F. Svaiter, *Newton-like dynamics and forwardbackward methods for structured monotone inclusions in Hilbert spaces*, *J. Optim. Theory Appl.*, **161** (2) (2014), 331–360.
- [2] S. Adly, H. Attouch, *Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping*, *SIAM J. Optim.*, 30(3) (2020), pp. 2134–2162.
- [3] S. Adly, H. Attouch, *First-order inertial algorithms involving dry friction damping*, *Math. Program.*, 193(1) (2022), pp. 405–445.

- [4] S. Adly, H. Attouch, Accelerated dynamics with dry friction via time scaling and averaging of doubly nonlinear evolution equations. *To appear in Nonlinear Analysis: Hybrid Systems*. HAL-CNRS-04083299
- [5] S. Adly, H. Attouch, Finite time stabilization of continuous inertial dynamics combining dry friction with Hessian-driven damping, *J. Conv. Analysis*, 28 (2) (2021), pp. 281310.
- [6] S. Adly, H. Attouch, A. Cabot, *Finite time stabilization of nonlinear oscillators subject to dry friction*, Nonsmooth Mechanics and Analysis, Adv. Mech. Math., 12 (2006), Springer, New York, pp. 289–304.
- [7] C.D. Alecsa, S. László, T. Pinta, *An extension of the second order dynamical system that models Nesterov's convex gradient method*, Appl. Math. Optim., (2020), <https://doi.org/10.1007/s00245-020-09692-1>
- [8] F. Alvarez, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM J. Control Optim., 38 (4) (2000), pp. 1102–1119.
- [9] F. Álvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., 81 (2002), No. 8, pp. 747–779.
- [10] H. Amann, J. I. Díaz, *A note on the dynamics of an oscillator in the presence of strong friction*, Nonlinear Analysis, 55 (2003), pp. 209–216.
- [11] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule*, Mathematical Programming, 180 (2020), 137–156.
- [12] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality*, Mathematics of Operations Research, 35 (2) (2010), pp. 438–457.
- [13] H. Attouch, J. Bolte, B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, regularized Gauss-Seidel methods*, Math. Program., 137(1) (2013), pp. 91–129.
- [14] H. Attouch, R.I. Bot, D.-K. Nguyen, *Fast convex optimization via time scale and averaging of the steepest descent*, arXiv:2208.08260v2 [math.OC] 30 Aug 2022.
- [15] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, 263 (2017), pp. 5412–5458.
- [16] H. Attouch, A. Cabot, Z. Chbani, H. Riahi, *Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient*, Evolution Equations and Control Theory, 7 (3) (2018), pp. 353–371.
- [17] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. Ser. B 168 (2018), pp. 123–175.
- [18] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case  $\alpha \leq 3$* , arXiv:1706.05671v1 [math.OC] 2017, ESAIM-COCV (2019) published electronically, DOI 10.1051/cocv/2017083.

- [19] H. Attouch, J. Fadili, V. Kungurtsev, *On the effect of perturbations, errors in first-order optimization methods with inertia and Hessian driven damping*, to appear in EECT, arXiv:2106.16159v1 [math.OC] 30 Jun 2021.
- [20] H. Attouch, J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than  $1/k^2$* , SIAM J. Optim., 26 (3) (2016), pp. 1824–1834.
- [21] H. Attouch, J. Peypouquet, P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, 261 (2016), pp. 5734–5783.
- [22] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, Math. Program., 193 (2022), pp. 113–155.
- [23] H. Attouch, M. Théra, *A general duality principle for the sum of two operators*, J. Convex Analysis, 3 (1996), pp. 1–24.
- [24] J.-F. Aujol, Ch. Dossal, *Stability of over-relaxations for the Forward-Backward algorithm, application to FISTA*, SIAM J. Optim., 25 (4) (2015), pp. 2408–2433.
- [25] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
- [26] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, 1972.
- [27] A. Chambolle, Ch. Dossal, *On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm*, J. Optim. Theory Appl., 166 (2015), pp. 968–982.
- [28] P. Colli, A. Visintin, *On a class of doubly nonlinear evolution equations*, Communications in Partial Differential Equations, 15 (5) (1990), pp. 737–756.
- [29] J. I. Díaz, A. Liñán, *On the asymptotic behavior of a damped oscillator under a sublinear friction term*, Rev. R. Acad. Cien. Serie A. Mat., 95 (1) (2001), pp. 155–160.
- [30] A. Haraux, M. Ghisi, M. Gambino, *Local and global smoothing effects for some linear hyperbolic equations with a strong dissipation*, Trans. Amer. Math. Soc. 368 (2016), 2039–2079.
- [31] S. Lojasiewicz, *Sur les trajectoires de gradient d'une fonction analytique*, Seminari di Geometria 1982–1983. Università di Bologna, Dipartimento di Matematica, pp 115–117, 1984.
- [32] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Math., 41 (3) (2017), pp. 681–685.
- [33] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).
- [34] B.T. Polyak, *Some methods of speeding up the convergence of iterative methods*, Z. Vysht Math. Fiz., 4 (1964), pp. 1–17.
- [35] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Applied Optimization.
- [36] R. Rossi, A. Mielke, G. Savaré, *A metric approach to a class of doubly nonlinear evolution equations and applications*, Ann. Scuola Norm. Sup. Pisa, (5) Vol 7, (2008), pp. 97–69.
- [37] B. Shi, S. S. Du, M. I. Jordan, W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, Math. Program., 195(1) (2022), pp. 79–148.

- [38] W. Su, S. Boyd, E. J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method*, Journal of Machine Learning Research, 17 (2016), pp. 1–43.