



Accelerated dynamics with dry friction via time scaling and averaging of doubly nonlinear evolution equations *

Samir Adly, Hedy Attouch

► To cite this version:

Samir Adly, Hedy Attouch. Accelerated dynamics with dry friction via time scaling and averaging of doubly nonlinear evolution equations *. 2023. hal-04083299

HAL Id: hal-04083299

<https://cnrs.hal.science/hal-04083299>

Preprint submitted on 27 Apr 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.

Accelerated dynamics with dry friction via time scaling and averaging of doubly nonlinear evolution equations.*

Samir ADLY[†] and Hedy ATTOUCH[‡]

ABSTRACT. In a Hilbert framework, for convex differentiable optimization, we analyze the long-time behavior of inertial dynamics with dry friction. In classical approaches based on asymptotically vanishing viscous damping (in accordance with Nesterov’s method), the results are expressed in terms of rapid convergence of the values of the function to its minimum value. On the other hand, dry friction induces convergence towards an approximate minimizer, typically the system stops at x when a given threshold $\|\nabla f(x)\| \leq r$ is satisfied. We will obtain rapid convergence results in this direction. In our approach, we start from a doubly nonlinear first-order evolution equation involving two potentials: one is the differentiable function f to be minimized, which acts on the state of the system via its gradient, and the other is the nonsmooth potential dry friction $\varphi(x) = r\|x\|$ which acts on the velocity vector via its sub-differential. To highlight the central role played by $\nabla f(x)$, we will also argue with the dual formulation of this dynamics, which has a Riemannian gradient structure. We then rely on the general acceleration method recently developed by Attouch, Bot and Nguyen, which consists in applying the method of time scaling and then averaging to a continuous differential equation of the first order in time. We thus obtain fast convergence results for second-order time-evolution systems involving dry friction, asymptotically vanishing viscous damping, and Hessian-driven damping in the implicit form.

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

1	Introduction and preliminary results	2
1.1	Some historical facts	3
1.1.1	Viscous friction	4
1.1.2	Dry friction	4
1.1.3	Hessian-driven damping	6
1.2	Contents	7
2	Steepest descent with two potentials: general f (not necessarily convex)	7
2.1	Existence and uniqueness of solution trajectories	8
2.2	First convergence properties	9

* At the time of finalizing the writing of this paper, we learned of the passing of Professor Boris T. Polyak, initiator of the link between inertial dissipative dynamics and optimization. Therefore, we would like to dedicate this work to his memory.

[†]Laboratoire XLIM, Université de Limoges, 87060 Limoges, France. E-mail: samir.adly@unilim.fr

[‡]IMAG, Université Montpellier, CNRS UMR 5149, Place Eugène Bataillon, 34095 Montpellier CEDEX 5, France. E-mail: hedy.attouch@umontpellier.fr

3	Steepest descent method with two potentials: case f convex	10
3.1	Convergence rate of velocities	10
3.2	Dual dynamic, Riemannian structure	11
3.3	Some particular cases with fast convergence	14
4	Convergence properties of the inertial system (DRYSTAR)	16
4.1	Introducing (DRYSTAR) by scaling and averaging of (DRYGRAD)	16
4.2	Convergence rates of (DRYSTAR)	18
4.3	A fast convergence result	22
5	Inertial dynamic and convergence associated with the dual formulation	22
6	Numerical illustrations	24
7	Conclusion, perspective	29
A	Auxiliary results	30
A.1	Convergence of values for (DRYGRAD)	30
A.2	Technical lemmas	32

1 Introduction and preliminary results

Nonsmooth dynamical systems are a class of evolution problems in which nonsmoothness of the involved trajectories or the vector fields occur. These systems are present in a variety of fields, including engineering, physics, biology, finance, and economics, and are of significant interest for understanding the dynamic behavior of various systems in practical applications. Nonsmooth dynamical systems and hybrid dynamical systems are closely related, as both involve the study of systems that exhibit both continuous and discontinuous behavior. Hybrid dynamical systems are systems that involve both continuous and discrete dynamics, meaning that their state evolves continuously over time and also changes abruptly at certain discrete events. These systems often have a combination of continuous and discrete variables, and the dynamics can switch between different modes or subsystems. Nonsmooth dynamical systems, on the other hand, are systems that involve continuous dynamics, but have discontinuities or singularities in the system's behavior. These discontinuities can arise from changes in the system's parameters or from the presence of constraints or impacts or dry friction. Dry friction is known to cause instabilities and affect the performance and behavior of mechanical systems. Evolution differential inclusion, a type of unilateral dynamical system, can be used to describe these systems. However, due to the lack of smoothness, classical smooth mathematical methods have limited applicability and require new analytical and numerical approaches to be developed for nonsmooth systems.

In this paper, we are interested in exploring the connection between continuous nonsmooth dynamical systems and numerical optimization. The step which follows our study is the passage by temporal discretization from the continuous dynamics to the numerical optimization algorithms (this is a subject of later studies). We begin by examining the first-order evolution equation

$$(\text{DRYGRAD}) \quad \dot{z}(t) + \partial\phi(\dot{z}(t)) + \nabla f(z(t)) \ni 0, \quad (1.1)$$

that is doubly nonlinear and involves two potentials. We make the following standing assumptions on the two potentials f and ϕ .

$$(\mathcal{A}) \quad \begin{cases} f : \mathcal{H} \rightarrow \mathbb{R} \text{ is a continuously differentiable function which is bounded from below.} \\ \nabla f \text{ is Lipschitz continuous on the bounded sets of } \mathcal{H}. \\ \phi : \mathcal{H} \rightarrow \mathbb{R} \text{ satisfies } \phi(x) = r\|x\| \text{ for some } r > 0. \end{cases} \quad (1.2)$$

While for most gradient methods in optimization the basic starting dynamic is the steepest descent method, we start in this paper from the doubly nonlinear differential inclusion (1.1) which contains the additional term $\partial\phi(\dot{z}(t))$ attached to dry friction, hence the short terminology (DRYGRAD) for Dry friction Gradient system. The first potential, denoted as f , is a differentiable function that is to be minimized and acts on the state of the system through its gradient ∇f . The second potential, denoted as $\phi(\cdot) = r\|\cdot\|$ (with $r > 0$), acts on the velocity vector via its subdifferential. The presence of this nonsmooth dry friction potential ϕ changes drastically the asymptotic behavior analysis of the associated dynamics. In classical gradient methods based on the steepest descent the results are expressed in terms of convergence of the values of the function to its minimum value. On the other hand, dry friction induces convergence towards an approximate minimizer, typically the system stops at x when a given threshold $\|\nabla f(x)\| \leq r$ is satisfied. To emphasize the role played by the gradient, we also examine the dual approach that governs the evolution of $g(t) = \nabla f(x(t))$, and the corresponding evolution

$$\nabla^2 f^*(g(t))\dot{g}(t) + g(t) - \text{proj}_{\mathbb{B}(0,r)}(g(t)) = 0 \quad (1.3)$$

thus making appear the Riemannian structure associated with the Hessian of the convex Fenchel conjugate function f^* (when this function is smooth) associated with f . Here, $\text{proj}_{\mathbb{B}(0,r)}$ denotes the projection operator onto the closed ball $\mathbb{B}(0, r)$. Our first study concerns the convergence properties as $t \rightarrow +\infty$ of the trajectories generated by the primal evolution systems (DRYGRAD) and its dual one (1.3).

We then rely on the general acceleration method recently developed by Attouch, Bot and Nguyen in [17] which consists in applying the method of time scaling and then averaging to a continuous differential equation of the first order in time. When applied to (DRYGRAD), this provides the second-order in time evolution system

$$(\text{DRYSTAR}) \quad \ddot{x}(s) + \frac{\alpha}{s}\dot{x}(s) + \nabla f\left(x(s) + \frac{s}{\alpha-1}\dot{x}(s)\right) - \text{proj}_{\mathbb{B}(0,r)}\left(\nabla f\left(x(s) + \frac{s}{\alpha-1}\dot{x}(s)\right)\right) = 0,$$

that involves dry friction aspects (smoothly via the resolvents of $\partial\phi$), 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. When starting from the dual dynamic we obtain

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

called (DRYSTARDE) for DRYSTAR Dual Equation. Based on the time scaling and averaging method, we do not need to develop a Lyapunov analysis for these inertial systems. We simply exploit the convergence results for the initial first-order system (DRYGRAD) and its dual, and then use tools from differential and integral calculus. We thus obtain fast convergence results for second-order time-evolution systems involving dry friction, asymptotically vanishing viscous damping, and Hessian-driven damping in the implicit form.

1.1 Some historical facts

Let's explain the role and the importance of each of the damping terms that enter our inertial dynamics.

1.1.1 Viscous friction

B. Polyak initiated the use of inertial dynamics to accelerate the gradient method in optimization. In [60], based on the inertial system with a fixed viscous damping coefficient $\gamma > 0$

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

he introduced the Heavy Ball with Friction method. For a strongly convex function f , the Heavy-Ball Method (HBF) guarantees exponential convergence of $f(x(t))$ to $\min_{\mathcal{H}} f$. For general convex functions, the convergence rate of (HBF) is at best on the order of $1/t$ (in the worst case), which is no better than that of the steepest descent method. A significant contribution in the field was made by Su-Boyd-Candès [68], with the introduction of a vanishing viscous damping coefficient, represented by $\gamma(t) = \frac{\alpha}{t}$, where α is a positive parameter. The Su-Boyd-Candès dynamic is a continuous ordinary differential equation (ODE) that represents the Nesterov accelerated gradient (NAG) method and it is given by

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

For general convex functions it provides a continuous version of the accelerated gradient method of Nesterov. For $\alpha \geq 3$, each trajectory $x(\cdot)$ of $(AVD)_{\alpha}$ satisfies the asymptotic convergence rate of the values $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/t^2)$. As a specific feature, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes (tends to zero) as time t goes to infinity, hence the terminology Asymptotic Vanishing Damping. The convergence properties of the dynamic $(AVD)_{\alpha}$ have been the subject of many recent studies, see [16, 19, 20, 21, 23, 24, 28, 32, 33, 54, 68]. The case where the parameter $\alpha = 3$ is particularly significant, as it corresponds to Nesterov's historical algorithm. However, in this case, the question of whether the trajectories converge remains an open problem, except in one dimension where convergence has been proven [24]. For values of $\alpha > 3$, it has been shown by Attouch-Chbani-Peypouquet-Redont [23] that each trajectory converges weakly to a minimizer of f . The corresponding algorithmic result was obtained by Chambolle-Dossal [41]. Additionally, it has been demonstrated in [28] and [54] that for $\alpha > 3$, the asymptotic convergence rate of the values is actually $\mathcal{O}(1/t^2)$. The subcritical case where $\alpha \leq 3$ has been studied by Apidopoulos-Aujol-Dossal [16] and Attouch-Chbani-Riahi [24], and it has been found that the convergence rate of the objective values is $\mathcal{O}\left(t^{-\frac{2\alpha}{3}}\right)$. These rates are optimal, meaning they can be reached or approached arbitrarily closely.

1.1.2 Dry friction

From a historical point of view, it seems that Leonardo da Vinci is credited with conducting the first systematic study of friction, in which he discovered that the friction force is proportional to the load, opposes motion and is independent of the contact area. In the 17th century, Guillaume Amontons [15] formulated what is known as Amontons's three laws of dry friction. He found that friction is proportional to the force between two surfaces. Later in the 18th, Charles-Augustin de Coulomb [44] expanded Amontons' work by expressing the frictional force as a function of the load and the direction of sliding velocity. Coulomb also showed that the force of friction is independent of the contact area between the two surfaces. Arthur Morin found that the friction at zero sliding speed (static friction) is larger than dynamic friction [57]. Osborne Reynolds introduced the concept of viscous friction in lubricated contact [62] and Richard Stribeck observed that the friction force decreases with increasing sliding speed, transitioning from static friction to Coulomb friction [67]. Since then, many scientists and engineers have continued to study and refine dry friction models. Coulomb's dry friction has numerous applications in fields such as mechanical engineering, civil engineering and material science, where it is used to design and optimize systems that

involve sliding and surface contact. More recently, J.-J. Moreau and P. D. Panagiotopoulos developed a mathematical framework for studying nonsmooth mechanical problems, which has been applied to the study of friction and stick-slip phenomena using advanced tools from modern convex analysis and set-valued analysis [3, 2, 40, 55, 56, 58]. This approach using set-valued functions led to the development of rigorous mathematical models like differential inclusions, variational inequalities, and hemivariational inequalities, which can precisely and rigorously describe the friction force and the stick-slip phenomena. We refer to [4, 10] for comprehensive references on the mathematical modeling and analysis of friction problems in the context of contact mechanics.

Coming back to our inertial dynamical systems involving dry friction, the first results concerning the finite convergence property under the action of dry friction have been obtained by Adly-Attouch-Cabot [8] for the continuous dynamics

$$\ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0, \quad t \in [t_0, +\infty[. \quad (1.4)$$

Assuming that the potential friction function ϕ has a sharp minimum at the origin (dry friction), they showed that, generically with respect to the initial data, the solution trajectories converge in finite time to equilibria. Similar results for the corresponding proximal-based algorithms have been obtained by Baji-Cabot [34] and Adly-Attouch [6].

Let's specify the mathematical tools that will be useful for analyzing the set-valued term $\partial\phi(\dot{x}(t))$ in equation (1.4), which represents dry friction in the model. The friction potential function ϕ is supposed to satisfy the Dry Friction property (denoted by (DF))

$$(DF) \quad \begin{cases} \phi : \mathcal{H} \rightarrow \mathbb{R} \text{ is convex continuous;} \\ \min_{\xi \in \mathcal{H}} \phi(\xi) = \phi(0) = 0; \\ 0 \in \text{int}(\partial\phi(0)). \end{cases}$$

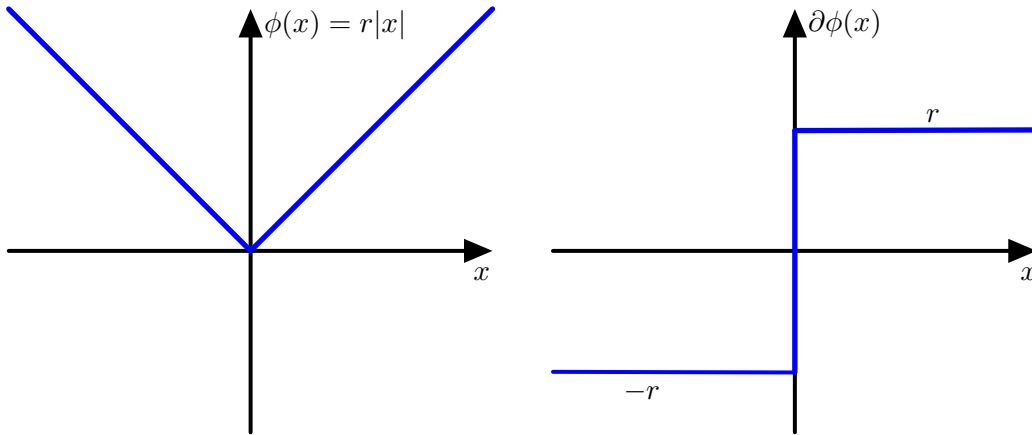


Figure 1: Dry friction potential and its associated Coulomb's friction law.

The particular case $\phi(\xi) = r\|\xi\|$, with $r > 0$, models dry friction (also called Coulomb friction) in mechanics. This simplified Coulomb friction model can be represented as a set-valued maximally monotone graph (see Figure 1), where the coefficient of dry friction is denoted as $r > 0$. However, this model has limitations in mechanics, as it is a monotone model and cannot capture stick-slip phenomena. To address this issue, a nonmonotone model called the Stribeck model is often used. In this work, we will focus on the analysis of dynamical systems with links to optimization algorithms, and for this purpose, we

will use the simplified monotone Coulomb friction law (DF).

The key assumption $0 \in \text{int}(\partial\phi(0))$ expresses that ϕ has a sharp minimum at the origin. This is specified in the following elementary lemma, see [2, Lemma 4.1 page 83], where, in item (iv), ϕ^* denotes the standard Legendre-Fenchel conjugate of ϕ defined by

$$\phi^*(p) = \sup_{x \in \mathcal{H}} \{ \langle p, x \rangle - \phi(x) \}.$$

Lemma 1.1 *Let $\phi : \mathcal{H} \rightarrow \mathbb{R}$ be a convex continuous function such that $\min_{\xi \in \mathcal{H}} \phi(\xi) = \phi(0) = 0$. Then, the following formulations of the dry friction are equivalent:*

- (i) $0 \in \text{int}(\partial\phi(0))$;
- (ii) *there exists some $r > 0$ such that $\mathbb{B}(0, r) \subset \partial\phi(0)$;*
- (iii) *there exists some $r > 0$ such that, for all $\xi \in \mathcal{H}$, $\phi(\xi) \geq r\|\xi\|$;*
- (iv) *there exists some $r > 0$ such that, $\|f\| \leq r \implies \partial\phi^*(f) \ni 0$,*

The parameter r is an important factor in our analysis. To better understand its role, we will say that the friction potential function ϕ satisfies the property (DF) $_r$ if ϕ satisfies the Dry Friction property (DF) with $\mathbb{B}(0, r) \subset \partial\phi(0)$. The property (iv) above expresses that, when the force f exerted on the system is less than a threshold $r > 0$, then the system stabilizes, *i.e.* the velocity $v = 0 \in \partial\phi^*(f)$. This contrasts with the viscous damping that can asymptotically produce many small oscillations.

The following lemma will play a key role in showing the finite convergence property. Indeed, this property gives the soft thresholding property satisfied by the proximal operator associated with a function ϕ having a sharp minimum at the origin. The proximal operator associated with the convex function $\phi : \mathcal{H} \rightarrow \mathbb{R}$ is defined as:

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

Lemma 1.2 *Let $\phi : \mathcal{H} \rightarrow \mathbb{R}$ be a convex continuous function that satisfies (DF) $_r$, *i.e.*, $\partial\phi(0) \supset B(0, r)$.*

Then the following implication holds: for $\lambda > 0$, and $x \in \mathcal{H}$

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

1.1.3 Hessian-driven damping

The inertial system

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

was introduced in [13]. In line with (HBF), a fixed positive friction coefficient γ is incorporated. The introduction of Hessian-driven damping in this method makes it possible to eliminate the transversal oscillations that are likely to occur with the traditional HBF method, as observed in [13] when using the Rosenbrock function. The importance of using geometric damping adapted to the specific function to be minimized, namely f , had already been considered by Alvarez in [12] when examining the inertial system

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

where $\Gamma : \mathcal{H} \rightarrow \mathcal{H}$ is a linear positive anisotropic operator. But still this damping operator is fixed. For a general convex function, the Hessian-driven damping in $(\text{DIN})_{\gamma,\beta}$ performs a similar operation in a closed-loop adaptive way. The terminology (DIN) stands shortly for Dynamical Inertial Newton. It refers to the natural link between this dynamic and the continuous Newton method. Recent studies have been devoted to the study of the inertial dynamic

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

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 [29], Attouch-Chbani-Fadili-Riahi [30], Shi-Du-Jordan-Su [65]. Related to this is the Inertial System with Implicit Hessian Damping

$$(\text{ISIHD}) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\left(x(t) + \beta(t)\dot{x}(t)\right) = 0, \quad (1.5)$$

considered by Alecsa-László-Pința in [11], see also Attouch-Fadili-Kungurtsev [25] in the perturbed case. The rationale justifying the use of the term “implicit” comes from the observation that by Taylor expansion (as $t \rightarrow +\infty$ we have $\dot{x}(t) \rightarrow 0$) one has

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

hence making the Hessian damping appear indirectly.

1.2 Contents

The paper is organized as follows. Sections 2 and 3 are devoted to the study of the first order evolution equation (DRYGRAD) respectively in the general case, and the convex case. In this last case, we analyse the dual dynamical Riemannian aspects. Then in Section 4, in the case f convex, we apply the time scale and averaging technique to (DRYGRAD) to obtain accelerated convergence results. In Section 5 we examine similar questions for the dual dynamic of (DRYGRAD). We complete this study with some numerical illustrations and perspectives.

2 Steepest descent with two potentials: general f (not necessarily convex)

Our approach is built on the Dry friction Gradient system

$$(\text{DRYGRAD}) \quad \dot{z}(t) + \partial\phi(\dot{z}(t)) + \nabla f(z(t)) \ni 0 \quad (2.1)$$

which is a doubly nonlinear evolution equation. The class of first-order evolution systems with two potentials models various nonlinear diffusion processes, and has therefore been extensively studied in the field of PDE's, [43], [64] and references therein.

Remark 2.1 Dry friction is a complex phenomenon that involves the interaction of multiple physical factors such as contact deformation, surface roughness, and adhesion. In order to accurately model dry friction, it is important to consider the system's dynamics and how they affect the contact forces.

In the case of a system undergoing dynamic motion, the mechanical interpretation of dry friction is more accurate when the term $\partial\phi$ is attached to a second-order dynamical system of the form:

$$\ddot{z}(t) + \gamma\dot{z}(t) + \partial\phi(\dot{z}(t)) + \nabla f(z(t)) \ni 0.$$

where $z(t)$ is the position, $\dot{z}(t)$ is the velocity, $\ddot{z}(t)$ is the acceleration, $\gamma > 0$ is a damping coefficient, $\partial\phi(\dot{z}(t))$ represents the dry frictional force, and $\nabla f(z(t))$ is the external force applied to the system. However, in quasi-static problems where the motion is very slow or negligible, the inertial effects can be neglected and only the position and velocity need to be considered. In this case, the mechanical interpretation of dry friction can be simplified and modeled by using (DRYGRAD). After devising the equation by $\gamma > 0$, this leads to a first-order dynamical system of the form:

$$\dot{z}(t) + \partial\phi(\dot{z}(t)) + \nabla f(z(t)) \ni 0,$$

where the second-order time derivative is absent, and the behavior of the system is determined solely by the position and velocity. The accuracy of the mechanical interpretation of dry friction depends on the dynamics of the system, and for quasi-static problems, (DRYGRAD) dynamic can be used to simplify the model and reduce it to a first-order dynamical system.

2.1 Existence and uniqueness of solution trajectories

Our approach is original due to the incorporation of the nonsmooth dry friction potential ϕ . Therefore, it requires a thorough and independent analysis. Many previous studies have focused mainly on the finite time existence of solutions, which can be a difficult topic for PDEs. In contrast, our approach simplifies this question and concentrates instead on the study of the asymptotic behavior of the solution trajectories. To study the existence and uniqueness of the solution trajectory to the Cauchy problem associated with (DRYGRAD), let us write it as

$$\dot{z}(t) - (I + \partial\phi)^{-1}(-\nabla f(z(t))) = 0.$$

Equivalently

$$\dot{z}(t) + T(z(t)) = 0$$

where $T = T_1 \circ T_2$ with

$$T_1(z) = -(I + \partial\phi)^{-1}(-z)$$

and

$$T_2(z) = \nabla f(z).$$

The operator T_1 is Lipschitz continuous, as being equal to the resolvent of the maximally monotone operator $\partial\phi$. Indeed, in our situation, it is equal to the soft thresholding operator $T_1(z) = z - \text{proj}_{\mathbb{B}(0,r)}(z)$. By assumption, the operator T_2 is Lipschitz continuous on the bounded sets. Therefore, the composition of the two operators is Lipschitz continuous on the bounded sets. The existence and uniqueness of a local solution to the corresponding Cauchy problem is a direct consequence of the classical Cauchy-Lipschitz theorem. The transition from a local solution to a global solution results from the following energy estimates. Note that we get a classical solution in the following sense. The trajectory z is of class C^1 but the velocity vector \dot{z} is only Lipschitz continuous. According to the above formulation of the soft thresholding operator T_1 , we obtain the following equivalent formulation of (DRYGRAD)

$$\dot{z}(t) + \nabla f(z(t)) - \text{proj}_{\mathbb{B}(0,r)}(\nabla f(z(t))) = 0. \quad (2.2)$$

It is worth noting that, even if f is convex, the operator $z \mapsto \nabla f(z) - \text{proj}_{\mathbb{B}(0,r)}(\nabla f(z))$ may not be monotone with $r > 0$, which makes this system relevant within the framework of doubly nonlinear equations.

2.2 First convergence properties

The convergence properties of the solution trajectories of (DRYGRAD) are summarized in the following theorem. Without loss of generality, we assume that $\|\nabla f(z_0)\| > r$, as we would have already reached an equilibrium otherwise.

Theorem 2.1 *Let $z : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global solution trajectory of (DRYGRAD). Then, $t \mapsto f(z(t))$ is a decreasing function, and the following properties are satisfied*

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

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

(iii) *the trajectory $t \mapsto z(t)$ converges strongly as $t \rightarrow +\infty$, and its limit z_∞ satisfies*

$$\|\nabla f(z_\infty)\| = r.$$

(iv) *Either the trajectory stabilizes in finite time at z_∞ such that $\|\nabla f(z_\infty)\| = r$, which is the general case, or else the trajectory satisfies for all $t \geq t_0$ the inequality $\|\nabla f(z(t))\| > r$.*

Proof. (i) and (ii) Take the scalar product of (DRYGRAD) with $\dot{z}(t)$. We get

$$\|\dot{z}(t)\|^2 + \langle \partial\phi(\dot{z}(t)), \dot{z}(t) \rangle + \frac{d}{dt} f(z(t)) = 0. \quad (2.3)$$

According to the properties of a dry friction potential we have

$$0 = \phi(0) \geq \phi(\dot{z}(t)) + \langle \partial\phi(\dot{z}(t)), -\dot{z}(t) \rangle$$

which gives

$$\langle \partial\phi(\dot{z}(t)), \dot{z}(t) \rangle \geq \phi(\dot{z}(t)) \geq r\|\dot{z}(t)\|.$$

Combining the above property with (2.3) we obtain

$$\|\dot{z}(t)\|^2 + r\|\dot{z}(t)\| + \frac{d}{dt} f(z(t)) \leq 0. \quad (2.4)$$

Therefore $\frac{d}{dt} f(z(t)) \leq 0$, which gives that $t \mapsto f(z(t))$ is a nonincreasing function. Integrating (2.4), and using the fact that f is bounded from below gives the integral estimates (i) and (ii).

(iii) Let us now come to the convergence properties of the solution trajectories of (DRYGRAD). Since $\int_{t_0}^{+\infty} \|\dot{z}(t)\| dt < +\infty$, it is immediate to verify that $z(t)$ satisfies the Cauchy property, and therefore converges strongly as $t \rightarrow +\infty$. Let us call z_∞ its limit. Since ∇f is continuous we have that, as $t \rightarrow +\infty$

$$\nabla f(z(t)) \rightarrow \nabla f(z_\infty).$$

From the equivalent formulation of (DRYGRAD)

$$\dot{z}(t) = (I + \partial\phi)^{-1} (-\nabla f(z(t)))$$

and the continuity of the resolvent operator we get

$$\lim_{t \rightarrow +\infty} \dot{z}(t) = (I + \partial\phi)^{-1} (-\nabla f(z_\infty)).$$

Since $\int_{t_0}^{+\infty} \|\dot{z}(t)\| dt < +\infty$ we deduce that $\lim_{t \rightarrow +\infty} \dot{z}(t) = 0$ and

$$(I + \partial\phi)^{-1}(-\nabla f(z_\infty)) = 0.$$

By definition of ϕ , this gives $\|\nabla f(z_\infty)\| \leq r$.

(iv) Let us now show that $\|\nabla f(z_\infty)\| = r$, and examine the finite time convergence property. We have proved that the trajectory $t \mapsto z(t)$ converges strongly as $t \rightarrow +\infty$ to a limit z_∞ , and that $\|\nabla f(z_\infty)\| \leq r$. There are two different cases to examine:

a) The trajectory $t \mapsto z(t)$ satisfies for all $t \geq t_0$

$$\|\nabla f(z(t))\| > r.$$

Then at the limit we have $\|\nabla f(z_\infty)\| = r$, and the trajectory $z(\cdot)$ reaches z_∞ after an infinite time.

b) There exists a finite time $t_1 > t_0$ such that

$$\|\nabla f(z(t_1))\| = r.$$

Indeed at time t_1 the vector field which governs the dynamic is equal to zero, which by Cauchy-Lipschitz formula implies that the trajectory remains fixed at this point. So we have finite convergence.

Thus, in all cases we have at the limit that $\|\nabla f(z_\infty)\| = r$. ■

Remark 2.2 Actually, the case a) mentioned above is quite exceptional. It could occur, for instance, when we examine the function $f(x) = \frac{1}{2}\|x\|^2$. For simplicity take $\mathcal{H} = \mathbb{R}$, and start with $z_0 > r$. In this case the dynamic (DRYGRAD) reduces to

$$\dot{z}(t) + z(t) - r = 0 \tag{2.5}$$

which gives

$$z(t) = e^{t_0-t} z_0 + r(1 - e^{t_0-t}) = r + e^{t_0-t}(z_0 - r).$$

Thus in this case we have convergence within infinite time as in case a).

3 Steepest descent method with two potentials: case f convex

3.1 Convergence rate of velocities

Let us first complete the results of Theorem 2.1 by examining the properties specific to the case where f is convex.

Theorem 3.1 *Let $z : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global solution trajectory of (DRYGRAD) where f is assumed to be convex. Then, the following properties are satisfied*

- (i) $t \mapsto \|\dot{z}(t)\|$ is decreasing;
- (ii) $\|\dot{z}(t)\| = o\left(\frac{1}{t}\right)$ as $t \rightarrow +\infty$.

Proof. (i) Let us use the specific form of the dry friction potential ϕ . As long as $\dot{z}(t) \neq 0$, we have

$$\dot{z}(t) + r \frac{\dot{z}(t)}{\|\dot{z}(t)\|} + \nabla f(z(t)) = 0. \tag{3.1}$$

After derivation of (3.1) with respect to t , we get

$$\ddot{z}(t) + r \frac{\dot{z}(t)}{\|\dot{z}(t)\|} - r \dot{z}(t) \frac{1}{\|\dot{z}(t)\|^2} \frac{d}{dt} (\|\dot{z}(t)\|) + \nabla^2 f(z(t)) \dot{z}(t) = 0. \quad (3.2)$$

Taking the scalar product with $\dot{z}(t)$ we get

$$\left(1 + \frac{r}{\|\dot{z}(t)\|}\right) \langle \ddot{z}(t), \dot{z}(t) \rangle - r \frac{d}{dt} (\|\dot{z}(t)\|) + \langle \nabla^2 f(z(t)) \dot{z}(t), \dot{z}(t) \rangle = 0. \quad (3.3)$$

According to the convexity of f we get

$$\frac{1}{2} \left(1 + \frac{r}{\|\dot{z}(t)\|}\right) \frac{d}{dt} (\|\dot{z}(t)\|^2) - r \frac{d}{dt} (\|\dot{z}(t)\|) \leq 0. \quad (3.4)$$

Equivalently

$$\|\dot{z}(t)\| \frac{d}{dt} (\|\dot{z}(t)\|) \leq 0, \quad (3.5)$$

which gives $\frac{d}{dt} (\|\dot{z}(t)\|) \leq 0$. Therefore $t \mapsto \|\dot{z}(t)\|$ is decreasing.

(ii) According to (ii) in Theorem 2.1, and $t \mapsto \|\dot{z}(t)\|$ is decreasing, we deduce that for all $t \geq t_0$

$$(t - t_0) \|\dot{z}(t)\| \leq \int_{t_0}^t \|\dot{z}(u)\| du \leq \int_{t_0}^{+\infty} \|\dot{z}(u)\| du < +\infty$$

which gives, that for some positive constant C

$$\|\dot{z}(t)\| \leq \frac{C}{t}.$$

Indeed we have

$$\|\dot{z}(t)\| = o\left(\frac{1}{t}\right)$$

thanks to the following lemma (apply it with $h(t) = \|\dot{z}(t)\|$).

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 $\lim_{t \rightarrow +\infty} th(t) = 0$.*

Proof. The nonincreasing property of h implies that $\frac{d}{dt} (th(t)) = h(t) + t\dot{h}(t) \leq h(t)$ for every $t \geq t_0$. Since $h \in \mathbb{L}^1([t_0, +\infty[)$, the result follows from [1, Lemma 5.2]. ■

The proof of Theorem 3.1 is thereby complete. ■

3.2 Dual dynamic, Riemannian structure

To better understand the convergence properties of gradients, which are crucial in (DRYGRAD), we should examine the dual dynamical approach. To begin, let's rewrite the (DRYGRAD) differential inclusion in a more compact form

$$\partial\psi(\dot{z}(t)) + \nabla f(z(t)) \ni 0, \quad (3.6)$$

where $\psi(v) := \phi(v) + \frac{1}{2}\|v\|^2$. Indeed, according to the classical additive subdifferential rule, we have $\partial\psi(v) = \partial\phi(v) + v$. Then, we apply the Attouch-Théra duality principle [31] to (3.6), which is governed by the sum of two operators. Set

$$g(t) := \nabla f(z(t))$$

which becomes our dual variable, and express all the involved quantities in terms of $g(t)$. According to the Fenchel conjugaison formula we have

$$\partial\psi(\dot{z}(t)) \ni -g(t) \iff \dot{z}(t) \in \partial\psi^*(-g(t))$$

and

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

Therefore, we get

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

Let us compute the above expression $-\partial\psi^*(-g(t))$. We have

$$\psi^* = (\phi + \frac{1}{2}\|\cdot\|^2)^* = (\phi^*)_1, \text{ the Moreau-Yosida regularization of } \phi^* \text{ of index 1.}$$

Since $\phi^* = \delta_{\mathbb{B}(0,r)}$, we get

$$\psi^*(\xi) = \frac{1}{2}\text{dist}^2(\xi, \mathbb{B}(0, r)),$$

and

$$-\partial\psi^*(-g(t)) = g(t) - \text{proj}_{\mathbb{B}(0,r)}(g(t)).$$

Therefore the dynamic which governs the evolution of $g(t) = \nabla f(z(t))$ is given by

$$\frac{d}{dt}(\partial f^*(g(t))) + g(t) - \text{proj}_{\mathbb{B}(0,r)}(g(t)) \ni 0. \quad (3.7)$$

This type of doubly nonlinear equation has also been the subject of active research. Particularly interesting is the fact that, under additional assumption of f , it can be interpreted as a Riemannian gradient flow. For that, we assume that f^* is a Legendre function of class C^2 . The dynamic can be equivalently written

$$\nabla^2 f^*(g(t))\dot{g}(t) + g(t) - \text{proj}_{\mathbb{B}(0,r)}(g(t)) = 0 \quad (3.8)$$

thus making appear the Riemannian structure associated with the Hessian of the convex function f^* . The second-order differentiability of convex functions has been explored in the literature from various perspectives. In connection with (3.8), the second-order differentiability of the Legendre function f^* , given a smooth convex function f , will be discussed in Remark 3.1.

Let us summarize the above results in the following statement, and establish the convergence rates of the gradients.

Theorem 3.2 *Let $z : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global solution trajectory of (DRYGRAD). Then $g(t) := \nabla f(z(t))$ is a solution trajectory of the generalized Riemannian gradient flow*

$$\frac{d}{dt}(\partial f^*(g(t))) + g(t) - \text{proj}_{\mathbb{B}(0,r)}(g(t)) \ni 0. \quad (3.9)$$

The following convergence properties are satisfied:

- (i) $t \mapsto \text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r))$ is a decreasing function, and the following convergence rate is satisfied: as $t \rightarrow +\infty$

$$\text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r)) = o\left(\frac{1}{t}\right);$$

(ii) For any $g_\infty \in \mathbb{B}(0, r)$, the function $t \mapsto D(g(t), g_\infty)$ is decreasing where

$$D(g(t), g_\infty) = f^*(g_\infty) - f^*(g(t)) - \langle \nabla f^*(g(t)), g_\infty - g(t) \rangle. \quad (3.10)$$

Proof. The dissipative properties associated with the above Riemannian gradient flow are obtained by using the Bregman distance defined by (3.10). By derivating this expression we get

$$\begin{aligned} \frac{d}{dt} D(g(t), g_\infty) &= -\frac{d}{dt} (f^*(g(t))) - \left\langle \frac{d}{dt} \nabla f^*(g(t)), g_\infty - g(t) \right\rangle + \langle \nabla f^*(g(t)), \dot{g}(t) \rangle \\ &= -\left\langle \frac{d}{dt} \nabla f^*(g(t)), g_\infty - g(t) \right\rangle \\ &= \left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), g_\infty - g(t) \right\rangle \end{aligned} \quad (3.11)$$

where the last equality comes from (3.7). Equivalently,

$$\frac{d}{dt} D(g(t), g_\infty) + \left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), g(t) - g_\infty \right\rangle = 0.$$

According to the convexity of the function $\psi^*(\xi) = \frac{1}{2} \text{dist}^2(\xi, \mathbb{B}(0, r))$, and $g_\infty \in \mathbb{B}(0, r)$ we have

$$0 \geq \frac{1}{2} \text{dist}^2(g(t), \mathbb{B}(0, r)) + \left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), g_\infty - g(t) \right\rangle.$$

Therefore,

$$\frac{d}{dt} D(g(t), g_\infty) + \frac{1}{2} \text{dist}^2(g(t), \mathbb{B}(0, r)) \leq 0.$$

So, we get that $t \mapsto D(g(t), g_\infty)$ is decreasing and

$$\int_{t_0}^{+\infty} \text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r)) dt < +\infty. \quad (3.12)$$

Let us now pass from the above integral estimate to a pointwise estimate. To this hand, let us take the scalar product of (3.7) with $\dot{g}(t)$. We get

$$\left\langle \frac{d}{dt} (\partial f^*(g(t))), \dot{g}(t) \right\rangle + \left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), \dot{g}(t) \right\rangle = 0. \quad (3.13)$$

On the one hand we have

$$\frac{d}{dt} \psi^*(g(t)) = \left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), \dot{g}(t) \right\rangle.$$

On the other hand

$$\left\langle \frac{d}{dt} (\partial f^*(g(t))), \dot{g}(t) \right\rangle \geq 0.$$

Indeed, in the case where f^* is \mathcal{C}^2 the above inequality is equivalent to

$$\langle \nabla^2 f^*(g(t)) \dot{g}(t), \dot{g}(t) \rangle \geq 0,$$

which is satisfied by convexity of f^* . In the general case, it is obtained by exploiting the monotonicity of ∂f^* at $g(t)$ and $g(t + h)$. Combining the above properties we get

$$\frac{d}{dt} \psi^*(g(t)) \leq 0. \quad (3.14)$$

Therefore $t \mapsto \psi^*(g(t))$ is a decreasing function, which, according to (3.12), satisfies the following integral estimate $\int_{t_0}^{+\infty} \psi^*(g(t))dt < +\infty$. We now apply lemma 3.1 to obtain

$$2\psi^*(g(t)) = \text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r)) = o\left(\frac{1}{t}\right), \quad (3.15)$$

and the proof is thereby completed. ■

Remark 3.1 Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup +\infty$ be a convex, proper and lower semicontinuous function. It is known that the Legendre conjugate f^* associated with f is related to the subdifferential of f as follows:

$$\begin{aligned} y \in \partial f(x) &\iff f(x) + f^*(y) = \langle y, x \rangle \\ &\iff x \in \partial f^*(y). \end{aligned}$$

Thus, the inverse of the subdifferential of f , ∂f , as a set-valued operator is ∂f^* , i.e. $\partial f^* = (\partial f)^{-1}$. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and of class \mathcal{C}^1 , then

$$y \in \partial f(x) \iff y = \nabla f(x).$$

Suppose that $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is nonsingular with a continuous inverse. Then f^* is also of class \mathcal{C}^1 and

$$\nabla f^* = (\nabla f)^{-1}.$$

Furthermore, if f is convex and twice continuously differentiable, i.e. f is of class \mathcal{C}^2 , with a nonsingular Hessian matrix $\nabla^2 f(x)$ at every point $x \in \mathbb{R}^n$, then by using the inverse function theorem, we have

$$(\nabla^2 f(x))^{-1} = \nabla^2 f^*(y), \text{ with } y = \nabla f(x).$$

We note that this formula can be rewritten in terms of associated quadratic forms as follows

$$y \in \partial f(x) \implies Q_{f^*}(y) = (Q_f(x))^*,$$

where $Q_f(x) = \frac{1}{2} \langle \nabla^2 f(x)z, z \rangle$ is the associated quadratic form with the Hessian matrix of f .

The following result linking conjugation and second-order properties of convex functions was proved by G. Gorni in [49]. Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup +\infty$ be a convex, proper and lower semicontinuous function. If f has a second-order Taylor expansion at the origin

$$f(x) = \frac{1}{2} \langle Ax, x \rangle + o(\|x\|^2) \text{ as } x \rightarrow 0,$$

with A symmetric and positive definite, then f^* has also a second-order Taylor expansion at the origin, which is given by

$$f^*(y) = \frac{1}{2} \langle A^{-1}y, y \rangle + o(\|y\|^2) \text{ as } y \rightarrow 0.$$

3.3 Some particular cases with fast convergence

Let us consider the case where f is a convex quadratic positively defined function.

Theorem 3.3 Suppose that $f(x) = \frac{1}{2}\langle Ax, x \rangle$ where $A : \mathcal{H} \rightarrow \mathcal{H}$ is a linear continuous operator which is symmetric and positive definite. Then for any solution trajectory $t \mapsto g(t)$ with $g(t) = \nabla f(z(t))$ of the dual version of (DRYGRAD), there is exponential convergence rate, i.e. for some $\mu > 0$

$$\text{dist}(\nabla f(z(t)), \mathbb{B}(0, r)) \leq Ce^{-\mu t},$$

and the trajectory takes infinite time to reach this equilibria. Precisely, starting with $\|\nabla f(z(t_0))\| > r$, then for some $\nu > 0$ we have

$$\text{dist}(\nabla f(z(t)), \mathbb{B}(0, r)) \geq Ce^{-\nu t}.$$

As a consequence for any trajectory z of (DRYGRAD), the equilibria is only reached after an infinite time.

Proof. According to $f(x) = \frac{1}{2}\langle Ax, x \rangle$, we have $f^*(x) = \frac{1}{2}\langle A^{-1}x, x \rangle$, where A^{-1} is the inverse operator of A . Then, the dual version of (DRYGRAD) writes as follows

$$A^{-1}\dot{g}(t) + g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)) = 0. \quad (3.16)$$

Set briefly $h(t) = \frac{1}{2}\text{dist}^2(g(t), \mathbb{B}(0, r))$. We have

$$\begin{aligned} \dot{h}(t) &= \left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), \dot{g}(t) \right\rangle \\ &= -\left\langle g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)), A(g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t))) \right\rangle. \end{aligned}$$

Let $\mu > 0$ and $\nu > 0$ such that for all $\xi \in \mathcal{H}$

$$\nu\|\xi\|^2 \geq \langle A\xi, \xi \rangle \geq \mu\|\xi\|^2.$$

Then

$$\dot{h}(t) + \mu\|g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t))\|^2 \leq 0 \leq \dot{h}(t) + \nu\|g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t))\|^2.$$

Let us now observe that

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

Therefore we get

$$\dot{h}(t) + 2\mu h(t) \leq 0 \leq \dot{h}(t) + 2\nu h(t).$$

This immediately implies that

$$C_1 e^{-2\nu t} \leq \text{dist}^2(g(t), \mathbb{B}(0, r)) = 2h(t) \leq C_2 e^{-2\mu t}.$$

and hence

$$C_3 e^{-\nu t} \leq \text{dist}(\nabla f(z(t)), \mathbb{B}(0, r)) \leq C_4 e^{-\mu t}.$$

As a consequence, according to the definition of (DRYGRAD) we have $\dot{z}(t) \neq 0$ for any $t \geq t_0$, which means that the equilibrium is reached only after an infinite time. ■

4 Convergence properties of the inertial system (DRYSTAR)

4.1 Introducing (DRYSTAR) by scaling and averaging of (DRYGRAD)

Let us make the change of time variable $t = \tau(s)$ in the first-order evolution equation (DRYGRAD), where $\tau(\cdot)$ is an increasing function from \mathbb{R}^+ to \mathbb{R}^+ , which is continuously differentiable, and which satisfies $\lim_{s \rightarrow +\infty} \tau(s) = +\infty$. Set

$$y(s) := z(\tau(s)).$$

On the one hand, by the derivation chain rule, we have

$$\dot{y}(s) = \dot{\tau}(s)\dot{z}(\tau(s)). \quad (4.1)$$

On the other hand, setting $t = \tau(s)$ in (DRYGRAD) gives

$$\dot{z}(\tau(s)) + \partial\phi(\dot{z}(\tau(s))) + \nabla f(z(\tau(s))) = 0. \quad (4.2)$$

According to (4.1) and (4.2), we obtain

$$\dot{y}(s) + \dot{\tau}(s)\partial\phi\left(\frac{1}{\dot{\tau}(s)}\dot{y}(s)\right) + \dot{\tau}(s)\nabla f(y(s)) \ni 0. \quad (4.3)$$

According to the positive homogeneity of degree zero of $\partial\phi$, we get

$$\dot{y}(s) + \dot{\tau}(s)\partial\phi(\dot{y}(s)) + \dot{\tau}(s)\nabla f(y(s)) \ni 0. \quad (4.4)$$

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

$$\dot{x}(s) + \frac{1}{\dot{\tau}(s)}(x(s) - y(s)) = 0, \quad (4.5)$$

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

$$y(s) = x(s) + \dot{\tau}(s)\dot{x}(s). \quad (4.6)$$

By temporal derivation of (4.6) we get

$$\dot{y}(s) = \dot{x}(s) + \ddot{\tau}(s)\dot{x}(s) + \dot{\tau}(s)\ddot{x}(s). \quad (4.7)$$

Replacing $\dot{y}(s)$ as given by (4.7) in (4.4) we get

$$\dot{\tau}(s)\ddot{x}(s) + (1 + \ddot{\tau}(s))\dot{x}(s) + \dot{\tau}(s)\partial\phi(\dot{y}(s)) + \dot{\tau}(s)\nabla f(y(s)) = 0. \quad (4.8)$$

After dividing by $\dot{\tau}(s) > 0$, and according to (4.6), we obtain

$$\ddot{x}(s) + \frac{1 + \ddot{\tau}(s)}{\dot{\tau}(s)}\dot{x}(s) + \partial\phi(\dot{y}(s)) + \nabla f(x(s) + \dot{\tau}(s)\dot{x}(s)) = 0. \quad (4.9)$$

Let us express this equation only in terms of x and its derivatives. According to (4.4), we have

$$\dot{y}(s) = (I + \dot{\tau}(s)\partial\phi)^{-1}(-\dot{\tau}(s)\nabla f(y(s))). \quad (4.10)$$

Therefore,

$$\begin{aligned}
 \partial\phi(\dot{y}(s)) &= -\frac{1}{\dot{\tau}(s)} \left((I + \dot{\tau}(s)\partial\phi)^{-1}(-\dot{\tau}(s)\nabla f(y(s))) + \dot{\tau}(s)\nabla f(y(s)) \right) \\
 &= \frac{1}{\dot{\tau}(s)} \left(-\dot{\tau}(s)\nabla f(y(s)) - (I + \dot{\tau}(s)\partial\phi)^{-1}(-\dot{\tau}(s)\nabla f(y(s))) \right) \\
 &= \nabla\phi_{\dot{\tau}(s)} \left(-\dot{\tau}(s)\nabla f \left(x(s) + \dot{\tau}(s)\dot{x}(s) \right) \right). \tag{4.11}
 \end{aligned}$$

Combining (4.9) with (4.11) we get

$$\ddot{x}(s) + \frac{1 + \ddot{\tau}(s)}{\dot{\tau}(s)} \dot{x}(s) + \nabla\phi_{\dot{\tau}(s)} \left(-\dot{\tau}(s)\nabla f \left(x(s) + \dot{\tau}(s)\dot{x}(s) \right) \right) + \nabla f \left(x(s) + \dot{\tau}(s)\dot{x}(s) \right) = 0. \tag{4.12}$$

According to the general equality, which is valid for any $\lambda > 0$ and $p \in \mathcal{H}$

$$\nabla\phi_{\lambda}(-\lambda p) + p = -(I + \partial\phi)^{-1}(-p) = p - \text{proj}_{\mathbb{B}(0,r)}(p),$$

we get the equivalent formulation

$$\ddot{x}(s) + \frac{1 + \ddot{\tau}(s)}{\dot{\tau}(s)} \dot{x}(s) + \nabla f \left(x(s) + \frac{s}{\alpha - 1} \dot{x}(s) \right) - \text{proj}_{\mathbb{B}(0,r)} \left(\nabla f \left(x(s) + \frac{s}{\alpha - 1} \dot{x}(s) \right) \right) = 0. \tag{4.13}$$

In doing so, we passed from the first-order differential equation (4.3) to the second-order differential equation (4.13), with the advantage that now the coefficient in front of the gradient is fixed.

Let us now particularize the time scale $\tau(\cdot)$. According to the Su, Boyd and Candès [68] model of the Nesterov method, we consider the case where the viscous damping coefficient in (4.9) satisfies

$$\frac{1 + \ddot{\tau}(s)}{\dot{\tau}(s)} = \frac{\alpha}{s}. \tag{4.14}$$

Set $\theta(s) = \dot{\tau}(s)$. We are led to solve the first-order linear differential equation

$$\dot{\theta}(s) - \frac{\alpha}{s}\theta(s) = -1. \tag{4.15}$$

After multiplication by $s^{-\alpha}$, we get equivalently

$$\frac{d}{ds} (s^{-\alpha}\theta(s)) = -s^{-\alpha}. \tag{4.16}$$

By integrating we get

$$s^{-\alpha}\theta(s) = -\frac{1}{-\alpha + 1} s^{-\alpha+1} + C_3, \tag{4.17}$$

which gives

$$\theta(s) = \frac{s}{\alpha - 1} + C_3 s^{\alpha}. \tag{4.18}$$

According to $\theta(s) = \dot{\tau}(s)$, we finally get

$$\tau(s) = \frac{s^2}{2(\alpha - 1)} + C_4 s^{\alpha+1} + C_5. \tag{4.19}$$

This leads to the choice $\alpha > 1$. As a particular simple situation we take $C_4 = C_5 = 0$, which gives

$$\tau(s) = \frac{s^2}{2(\alpha - 1)}. \tag{4.20}$$

Replacing τ by the above value respectively in (4.4)-(4.5) and (4.12) gives the following result.

Theorem 4.1 *Time scaling and averaging of the steepest descent method with two potentials (DRY-GRAD) and with time scale $\tau(s) = \frac{s^2}{2(\alpha - 1)}$ give the two equivalent differential systems:*

(i) *The couple of variables (y, x) is solution of the following first-order differential system :*

$$\begin{cases} \dot{y}(s) + \frac{s}{\alpha - 1} \partial \phi(\dot{y}(s)) + \frac{s}{\alpha - 1} \nabla f(y(s)) \ni 0 \\ \dot{x}(s) + \frac{\alpha - 1}{s} (x(s) - y(s)) = 0. \end{cases} \quad (4.21)$$

(ii) *x is solution of the inertial system*

$$\ddot{x}(s) + \frac{\alpha}{s} \dot{x}(s) - (I + \partial \phi)^{-1} \left[-\nabla f \left(x(s) + \frac{s}{\alpha - 1} \dot{x}(s) \right) \right] = 0, \quad (4.22)$$

that is

$$(\text{DRYSTAR}) \quad \ddot{x}(s) + \frac{\alpha}{s} \dot{x}(s) + \nabla f \left(x(s) + \frac{s}{\alpha - 1} \dot{x}(s) \right) - \text{proj}_{\mathbb{B}(0, r)} \left(\nabla f \left(x(s) + \frac{s}{\alpha - 1} \dot{x}(s) \right) \right) = 0.$$

The introduction of this new dynamic requires several comments.

Comments. a) Note that in (4.22) we have to deal with a classical differential equation. Indeed the nonsmooth potential ϕ which models dry friction now enters the dynamic via its resolvent operator, which is a Lipschitz continuous operator.

b) In doing so, we passed from the first-order differential equation

$$\dot{y}(s) + \frac{s}{\alpha - 1} \partial \phi(\dot{y}(s)) + \frac{s}{\alpha - 1} \nabla f(y(s)) \ni 0 \quad (4.23)$$

to the second-order differential equation (DRYSTAR). The benefit of this transformation is that the coefficient in front of the gradient is now fixed. However, it is important to demonstrate that the fast convergence rates are still present.

c) When $\phi = 0$, *i.e.* without dry friction (which corresponds to $r = 0$), the above dynamic (DRYSTAR) is related to the Inertial System with Implicit Hessian Damping

$$(\text{ISIHD}) \quad \ddot{x}(s) + \frac{\alpha}{s} \dot{x}(s) + \nabla f \left(x(s) + \beta(s) \dot{x}(s) \right) = 0, \quad (4.24)$$

considered by Alecsa, László, and Pința in [11], see also Attouch, Fadili, and Kungurtsev [25] in the perturbed case. The reason for referring to the approach as “implicit” stems from the fact that by utilizing Taylor expansion, as s approaches infinity and $\dot{x}(s)$ approaches zero, one can approximate:

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

resulting in the appearance of the Hessian damping indirectly.

4.2 Convergence rates of (DRYSTAR)

We will assess the convergence rates of (DRYSTAR) trajectories through the use of time scale and averaging arguments, without performing any additional Lyapunov analysis. To this end, we will rely on the convergence results established in Theorem 2.1 for (DRYGRAD). As before, we begin by selecting an initial point x_0 such that $\|\nabla f(x_0)\| > r$, otherwise the trajectory would reach an equilibrium at the origin.

Theorem 4.2 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a convex \mathcal{C}^1 function, whose gradient is Lipschitz continuous on the bounded sets and such that $S = \operatorname{argmin} f \neq \emptyset$. Suppose that $\alpha > 1$. Let $x : [s_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of (DRYSTAR) that is*

$$\ddot{x}(s) + \frac{\alpha}{s} \dot{x}(s) + \nabla f\left(x(s) + \frac{s}{\alpha-1} \dot{x}(s)\right) - \operatorname{proj}_{\mathbb{B}(0,r)}\left(\nabla f\left(x(s) + \frac{s}{\alpha-1} \dot{x}(s)\right)\right) = 0. \quad (4.25)$$

Then the following statements are satisfied:

$$(i) \quad \|\dot{x}(s)\| = o\left(\frac{1}{s}\right) \text{ as } s \rightarrow +\infty.$$

$$(ii) \quad \int_{s_0}^{+\infty} \|\dot{x}(s)\| ds < +\infty.$$

$$(iii) \quad x(s) \text{ converges strongly as } s \rightarrow +\infty, \text{ and its limit } x_\infty \text{ satisfies } \|\nabla f(x_\infty)\| = r.$$

Proof. Following [17], let us interpret the transition from y to x as an averaging process. Next, using techniques from integral calculus, we aim to transfer the properties of y , obtained from Theorem 2.1 through time scaling, to x . To achieve this, we recast the second equation of (4.21) as follows:

$$s\dot{x}(s) + (\alpha - 1)x(s) = (\alpha - 1)y(s). \quad (4.26)$$

After multiplication of (4.26) by $s^{\alpha-2}$, we get equivalently

$$s^{\alpha-1}\dot{x}(s) + (\alpha - 1)s^{\alpha-2}x(s) = (\alpha - 1)s^{\alpha-2}y(s), \quad (4.27)$$

that is

$$\frac{d}{ds} (s^{\alpha-1}x(s)) = (\alpha - 1)s^{\alpha-2}y(s). \quad (4.28)$$

By integrating (4.28) from s_0 to s , and according to $x(s_0) = y(s_0)$, we obtain

$$x(s) = \frac{s_0^{\alpha-1}}{s^{\alpha-1}}y(s_0) + \frac{\alpha-1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2}y(u)du. \quad (4.29)$$

Then, observe that $x(s)$ can be simply written as follows

$$x(s) = \int_{s_0}^s y(u) d\mu_s(u), \quad (4.30)$$

where μ_s is the positive Radon measure on $[s_0, s]$ 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 δ_{s_0} is the Dirac measure at s_0 . We have that μ_s is a positive Radon measure on $[s_0, s]$ whose total mass is equal to 1. It is therefore a probability measure, and $\int_{s_0}^s y(u) d\mu_s(u)$ is obtained by **averaging** the trajectory $y(\cdot)$ on $[s_0, s]$ with respect to μ_s . From there, we can deduce fast convergence properties for the solution trajectories of (DRYSTAR).

i) Let us first consider the convergence rate of the velocities. According to Theorem (3.1)(ii), there exists a function $\varepsilon_0(\cdot)$ with $\lim_{t \rightarrow +\infty} \varepsilon_0(t) = 0$ such that

$$\|\dot{z}(t)\| = \frac{\varepsilon_0(t)}{t}.$$

Therefore

$$\|\dot{z}(\tau(s))\| = \frac{\varepsilon_0(\tau(s))}{\tau(s)}. \quad (4.31)$$

Sinc $\dot{y}(s) = \dot{\tau}(s)\dot{z}(\tau(s))$ we deduce that

$$\|\dot{y}(s)\| = \frac{\varepsilon_0(\tau(s))\dot{\tau}(s)}{\tau(s)}. \quad (4.32)$$

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

$$\|\dot{y}(s)\| = \frac{\varepsilon(s)}{s}, \quad (4.33)$$

with $\varepsilon(s) = 2\varepsilon_0(\tau(s))$ which tends to zero as $s \rightarrow +\infty$. Let us now establish a similar estimate for $\|\dot{x}(s)\|$. According to the definition of x

$$x(s) = \frac{s_0^{\alpha-1}}{s^{\alpha-1}}x_0 + \frac{\alpha-1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2}y(u)du. \quad (4.34)$$

Let us derivate this expression. We get

$$\dot{x}(s) = -\frac{(\alpha-1)s_0^{\alpha-1}}{s^\alpha}x_0 - \frac{(\alpha-1)^2}{s^\alpha} \int_{s_0}^s u^{\alpha-2}y(u)du + (\alpha-1)\frac{1}{s}y(s). \quad (4.35)$$

Let us reformulate this expression in terms of $\dot{y}(s)$, which is the quantity whose speed of convergence is known by (4.33). Indeed, by integration by parts we have

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

After multiplication of the above expression by $\frac{\alpha-1}{s^\alpha}$, and according to $y(s_0) = x_0$, we get

$$\frac{\alpha-1}{s^\alpha} \int_{s_0}^s u^{\alpha-1}\dot{y}(u)du = (\alpha-1)\frac{1}{s}y(s) - \frac{(\alpha-1)s_0^{\alpha-1}}{s^\alpha}x_0 - \frac{(\alpha-1)^2}{s^\alpha} \int_{s_0}^s u^{\alpha-2}y(u)du. \quad (4.36)$$

Comparing (4.35) and (4.36) we get

$$\dot{x}(s) = \frac{\alpha-1}{s^\alpha} \int_{s_0}^s u^{\alpha-1}\dot{y}(u)du. \quad (4.37)$$

After multiplication of (4.37) by s , and according to (4.33), we obtain successively

$$\begin{aligned} s\dot{x}(s) &= \frac{\alpha-1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2}u\dot{y}(u)du \\ s\|\dot{x}(s)\| &\leq \frac{\alpha-1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2}\varepsilon(u)du. \end{aligned}$$

Since $\varepsilon(s)$ tends to zero as $s \rightarrow +\infty$, by applying Lemma A.1 we obtain $\lim_{s \rightarrow +\infty} \|s\dot{x}(s)\| = 0$.

ii) According to Theorem 2.1 ii), we have

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

i.e. , the trajectory $z(\cdot)$ has a finite length. This property being invariant by change of time variable, we have that

$$\int_{s_0}^{+\infty} \|\dot{y}(s)\| ds < +\infty.$$

According to (4.37) we have that

$$\|\dot{x}(s)\| \leq \frac{\alpha - 1}{s^\alpha} \int_{s_0}^s u^{\alpha-1} \|\dot{y}(u)\| du. \quad (4.38)$$

Therefore

$$\int_{s_0}^{+\infty} \|\dot{x}(s)\| ds \leq \int_{s_0}^{+\infty} \left(\frac{\alpha - 1}{s^\alpha} \int_{s_0}^s u^{\alpha-1} \|\dot{y}(u)\| du \right) ds \quad (4.39)$$

$$= \int_{s_0}^{+\infty} \|\dot{y}(u)\| du < +\infty \quad (4.40)$$

where the last equality follows from a simple calculus using the Fubini-Tonelli inversion theorem for the multiple integral of a nonnegative function.

iii) Therefore, we have that the trajectory $x(\cdot)$ converges strongly to some x_∞ . Let us verify that

$$\lim_{t \rightarrow +\infty} z(t) = \lim_{s \rightarrow +\infty} y(s) = \lim_{s \rightarrow +\infty} x(s).$$

Indeed, the first equality is an immediate consequence of $y(s) = z(\tau(s))$ and $\tau(s) \rightarrow +\infty$ as $s \rightarrow +\infty$. The second equality follows again from the averaging property, and from the general property which says that convergence entails ergodic convergence. Let us make this precise. By definition of $x(s)$

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

Let $y_\infty = \lim_{s \rightarrow +\infty} y(s)$. Since μ_s is a probability measure

$$x(s) - y_\infty = \int_{s_0}^s (y(u) - y_\infty) d\mu_s(u) = \frac{s_0^{\alpha-1}}{s^{\alpha-1}} (y(s_0) - y_\infty) + \frac{\alpha - 1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2} (y(u) - y_\infty) du.$$

Therefore

$$\|x(s) - y_\infty\| \leq \frac{s_0^{\alpha-1}}{s^{\alpha-1}} \|y(s_0) - y_\infty\| + \frac{\alpha - 1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2} \|y(u) - y_\infty\| du.$$

The equality $\lim_{s \rightarrow +\infty} y(s) = \lim_{s \rightarrow +\infty} x(s)$ follows from a direct application of Lemma A.1. According to Theorem 2.1 iii) we thus have that $x(s)$ converges strongly as $s \rightarrow +\infty$, and its limit x_∞ satisfies

$$\|\nabla f(x_\infty)\| = r,$$

which completes the proof. ■

Remark 4.1 As a general rule, the trajectories $z(\cdot)$ and $y(\cdot)$ ultimately reach the same equilibrium in finite time. As a result, for some finite time \bar{s} , we can conclude that $y(\bar{s}) = x_\infty$. Using the average formula in equation (4.34), we can compute that for all $s \geq \bar{s}$, the following expression holds:

$$x(s) = x_\infty + \frac{C}{s^{\alpha-1}}. \quad (4.41)$$

Thus, finite-time stabilization is no longer valid for $x(\cdot)$, but it is nearly true when α is large.

4.3 A fast convergence result

A direct adaptation of Theorem 3.3 gives the following result.

Theorem 4.3 *Suppose that $f(x) = \frac{1}{2}\langle Ax, x \rangle$ where $A : \mathcal{H} \rightarrow \mathcal{H}$ is a linear continuous operator which is symmetric and positive definite. Then for any solution trajectory $x : [s_0, +\infty[\rightarrow \mathcal{H}$ of (DRYSTAR), we have the following accelerated exponential convergence rate: for some $\mu > 0$*

$$\text{dist} \left(\nabla f \left(x(s) + \frac{s}{\alpha - 1} \dot{x}(s) \right), \mathbb{B}(0, r) \right) \leq C e^{-\mu s^2}.$$

5 Inertial dynamic and convergence associated with the dual formulation

Let us apply the time scale and averaging technique to the dual formulation of the (DRYGRAD) dynamic that we recall below

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

where ψ^* is the even continuously differentiable function

$$\psi^*(g) = \frac{1}{2} \text{dist}^2(g, \mathbb{B}(0, r)),$$

whose gradient is equal to

$$\nabla \psi^*(g) = g - \text{proj}_{\mathbb{B}(0, r)}(g).$$

Equivalently, when f^* is smooth

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

thus making appear the Riemannian gradient structure.

Let us make the change of time variable $t = \tau(s)$ and set $v(s) = g(\tau(s))$. We obtain

$$\nabla^2 f^*(v(s)) \dot{v}(s) + \dot{\tau}(s) \nabla \psi^*(v(s)) = 0.$$

Next, we proceed with the averaging process and introduce $w(\cdot)$, defined as follows:

$$\dot{w}(s) + \frac{1}{\dot{\tau}(s)} (w(s) - v(s)) = 0, \tag{5.1}$$

with $w(s_0) = v(s_0) = x_0$ given in \mathcal{H} . We get

$$v(s) = w(s) + \dot{\tau}(s) \dot{w}(s)$$

which, by differentiation, gives

$$\dot{v}(s) = \dot{\tau}(s) \ddot{w}(s) + (1 + \ddot{\tau}(s)) \dot{w}(s).$$

Combining the above relations we get

$$\nabla^2 f^* \left(w(s) + \dot{\tau}(s) \dot{w}(s) \right) \left(\dot{\tau}(s) \ddot{w}(s) + (1 + \ddot{\tau}(s)) \dot{w}(s) \right) + \dot{\tau}(s) \nabla \psi^* \left(w(s) + \dot{\tau}(s) \dot{w}(s) \right) = 0.$$

After dividing by $\dot{\tau}(s)$, and using that the Hessian acts as a linear operator, we obtain

$$\nabla^2 f^* \left(w(s) + \dot{\tau}(s) \dot{w}(s) \right) \left(\ddot{w}(s) + \frac{1 + \ddot{\tau}(s)}{\dot{\tau}(s)} \dot{w}(s) \right) + \nabla \psi^* \left(w(s) + \dot{\tau}(s) \dot{w}(s) \right) = 0.$$

In accordance with the developments of section 4, take $\alpha > 1$ and

$$\tau(s) = \frac{s^2}{2(\alpha - 1)}. \quad (5.2)$$

This gives $\frac{1 + \ddot{\tau}(s)}{\dot{\tau}(s)} = \frac{\alpha}{s}$. Therefore the dynamic becomes

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

called (DRYSTARDE) for DRYSTAR Dual Equation.

Let us now examine the convergence rate properties of this dynamic. According to (3.15) we have

$$\psi^*(g(t)) = \frac{1}{2} \text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r)) = o\left(\frac{1}{t}\right). \quad (5.3)$$

Therefore the time scaling $\tau(s) = \frac{s^2}{2(\alpha - 1)}$, and $v(s) = g(\tau(s))$ gives

$$\psi^*(v(s)) = \psi^*(g(\tau(s))) = o\left(\frac{1}{s^2}\right). \quad (5.4)$$

Then use the interpretation of w as an average of v , that is

$$w(s) = \int_{s_0}^s v(u) d\mu_s(u), \quad (5.5)$$

where μ_s is the probability measure on $[s_0, s]$ 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 δ_{s_0} is the Dirac measure at s_0 . According to the convexity of ψ^* and the Jensen inequality we obtain

$$\psi^*(w(s)) = \text{dist}^2(w(s), \mathbb{B}(0, r)) = o\left(\frac{1}{s^2}\right). \quad (5.6)$$

Equivalently,

$$\text{dist}(w(s), \mathbb{B}(0, r)) = o\left(\frac{1}{s}\right). \quad (5.7)$$

Let us summarize the above results in the following statement.

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

$$\frac{d}{dt} (\partial f^*(g(t))) + g(t) - \text{proj}_{\mathbb{B}(0, r)}(g(t)) \ni 0. \quad (5.8)$$

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

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

Then w satisfies the following inertial system

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

with $\nabla \psi^*(w(s)) = 2w(s) - 2\text{proj}_{\mathbb{B}(0, r)}(w(s))$ and the following convergence properties are satisfied: as $s \rightarrow +\infty$

$$\text{dist}^2 \left(w(s) + \frac{s}{\alpha - 1} \dot{w}(s), \mathbb{B}(0, r) \right) = o\left(\frac{1}{s^2}\right) \text{ and } \text{dist}^2(w(s), \mathbb{B}(0, r)) = o\left(\frac{1}{s^2}\right).$$

6 Numerical illustrations

In this section, we will use adapted standard Runge-Kutta integrators to solve the involved continuous dynamics and conduct a series of numerical experiments to demonstrate the theoretical results discussed in the previous sections. Nesterov's accelerated gradient method is focused on achieving fast convergence of the objective function values in optimization problems. In contrast, our approach is centered on achieving fast convergence of the gradients of the objective function. By using a Riemannian gradient structure, we are able to obtain fast convergence rates for both the system and its dual, which governs the evolution of the gradients. This allows us to efficiently r -minimize, in the sense $\|\nabla f(z_\infty)\| \leq r$, the differentiable function f subject to the dry friction potential $\varphi(z) = r\|z\|$ and achieve our optimization goals with greater speed and accuracy. We set the asymptotic vanishing parameter α to a fixed value of 5 in all numerical experiments.

Example 6.1 Let us begin this section by considering an example to illustrate the dynamic (DRYSTAR) 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 $x_0 = (2, 2)$ and $x_1 = (1, 1)$. Note that f is the form $f(x) = \langle x, Qx \rangle$ with $Q = \text{diag}([a, b])$. We take $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$, $x = (x_1, x_2) \mapsto \phi(x) = r\|x\|_2$, with $r > 0$.

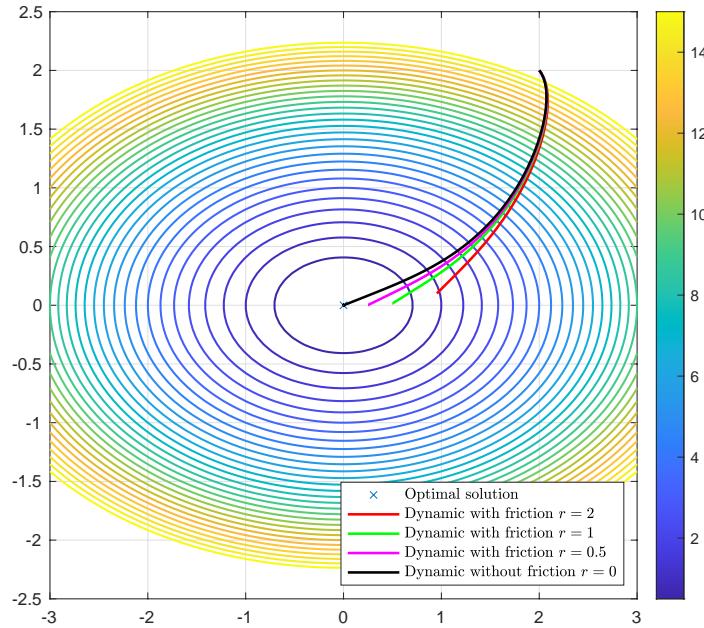


Figure 2: Illustration of the second-order evolution system (DRYSTAR) for different values of the dry friction coefficient r .

In Figure 2, we have plotted trajectories of the dynamic (DRYSTAR) for various values of the dry friction coefficient r , including the case where there is no friction (i.e. $r = 0$), the black trajectory. We observe that the dynamic (DRYSTAR) stops on an r -stationary point that satisfies: $\|\nabla f(z_\infty)\| \leq r$. On the other hand, the trajectory of the dynamic without friction (shown in black color) converges to the unique stationary point $z_\infty = (0, 0)$ of f . Let us make the following remark concerning the r -stationary point condition $\|\nabla f(z_\infty)\| \leq r$.

Remark 6.1 Consider any iterative method, such as gradient descent, for finding the critical points of a function f . Suppose that we have a sequence $(x_k)_{k \in \mathbb{N}}$ generated by the method. There are several stopping criteria that can be used, such as $\|x_{k+1} - x_k\| \leq \text{tolerance}$ or $\|\nabla f(x_k)\| \leq \text{tolerance}$, where the tolerance is a user-defined parameter. When the stopping criterion is based on the norm of the gradient of the function f , i.e., $\|\nabla f(x_k)\| \leq \text{tolerance}$, we implicitly assume that $\|\nabla f(x_\infty)\| \leq r$, where r is the tolerance. Theorem 4.2 shows that if the tolerance $r > 0$ is chosen appropriately, the trajectories of the system with dry friction (4.21) and (4.22), denoted by $s \mapsto x(s)$ and $s \mapsto y(s)$, satisfy $\lim_{s \rightarrow +\infty} x(s) = \lim_{s \rightarrow +\infty} y(s) = x_\infty$, where x_∞ satisfies $\|\nabla f(x_\infty)\| = r$. Therefore, the condition $\|\nabla f(x_\infty)\| \leq r$, known as the r -stationary point condition, appears to be a natural choice for practical use.

Example 6.2 In this example, we illustrate Theorem 2.1 on a quadratic function. It states that if $z : [t_0, +\infty[\rightarrow \mathbb{R}^2$ is a global solution trajectory of (DRYGRAD), where the gradient of the function f is projected onto the ball of radius r , then the function $t \mapsto f(z(t))$ is decreasing. Additionally, the theorem provides four important properties satisfied by the trajectory:

- The integral of the norm of the velocity squared is finite.
- The integral of the norm of the velocity is finite.
- The trajectory converges to a limit z_∞ as $t \rightarrow +\infty$, and the norm of the gradient of f at the limit point is equal to r .

In figure 3, we plotted four figures that show the behavior of a trajectory of the dynamic (DRYGRAD). The first figure shows the value of the objective function f along the trajectory as a function of time. It demonstrates that the function value decreases over time. The second figure shows the norm of the velocity vector $\|\dot{z}\|$ as a function of time. It demonstrates that the velocity vector converges to zero over time. The third figure shows the trajectory of the system in the x - y plane. It demonstrates the path taken by the trajectory starting from $z_0 = (2, 0.5)$. The fourth figure shows the norm of the gradient vector $\|\nabla f(z(t))\|$ as a function of time. It demonstrates that the gradient vector converges to the boundary of the ball $\mathbb{B}(0, r)$ with radius r . In the numerical test, we considered the function $f(x, y) = \frac{1}{2}x^2 + 5y^2$, and choose $r = 0.1$ and $z_0 = (2, \frac{1}{2})$.

Example 6.3 In this example we illustrate the exponential convergence rate stated in Theorem 3.3 using a quadratic function of the form $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ of the form

$$f(x) = \frac{1}{2} \langle Qx, x \rangle,$$

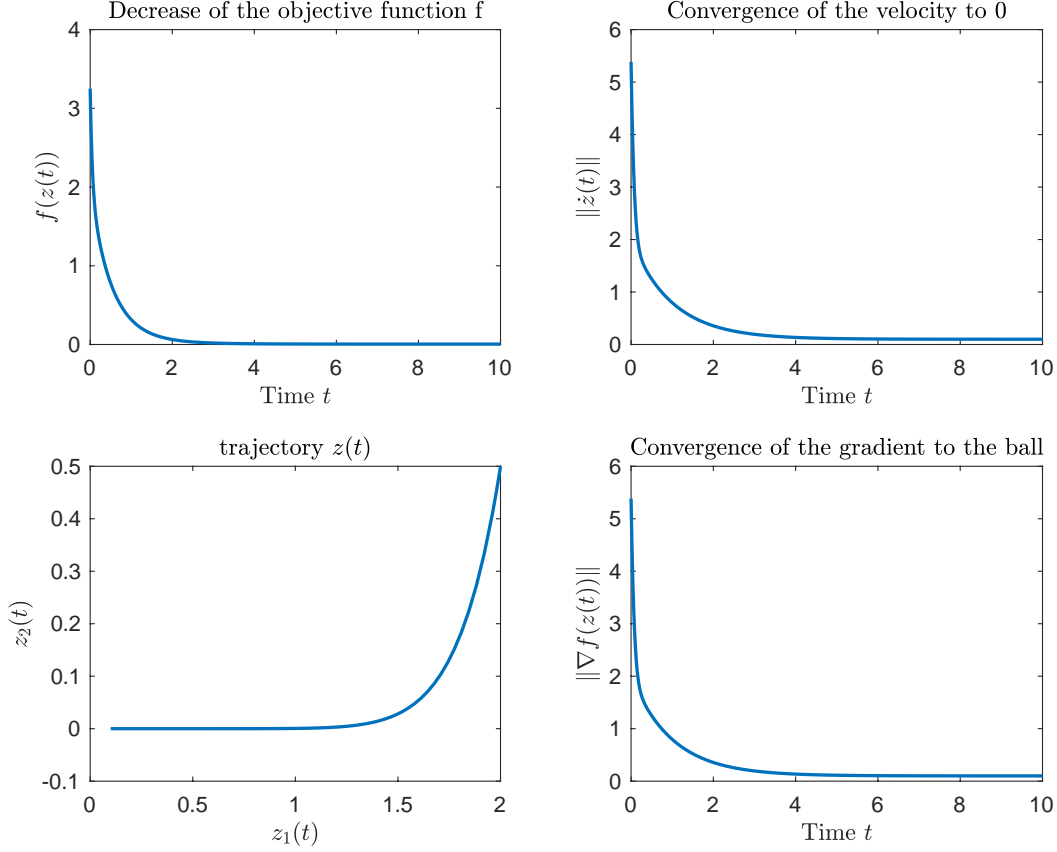
where $Q = \text{diag}([a, b])$ with $0 < a \leq b$. Let us remind that the first order dual dynamic can be rewritten as

$$\dot{g}(t) = -Qg(t) + Q\text{proj}_{\mathbb{B}(0, r)}(g(t)), \quad (6.1)$$

with $g(t) = \nabla f(z(t))$. Here $\mu = a$ and $\nu = 1/b$. We observe in Figure 4 that

$$\text{dist}(\nabla f(z(t)), \mathbb{B}(0, r)) \geq Ce^{-\nu t}.$$

Therefore, for any trajectory $t \mapsto z(t)$ of (DRYGRAD), the equilibrium is only reached after an infinite amount of time.


 Figure 3: Illustration of Theorem 2.1 with $r = 0.1$ and $z_0 = (2, 0.5)$.

Example 6.4 In this example, we aim to compare the two dual dynamics: first-order Riemannian gradient flow and second-order dynamic (DRYSTARDE) on a quadratic function that has the same form as the one in Example 6.3. It is worth noting that in this case, (DRYSTARDE) can be expressed as follows:

$$\ddot{w}(s) + \frac{\alpha}{s}\dot{w}(s) + \frac{2s}{\alpha-1}Q\dot{w}(s) + 2Qw(s) - 2Q\text{proj}_{\mathbb{B}(0,r)}\left(w(s) + \frac{s}{\alpha-1}\dot{w}(s)\right) = 0, \quad (6.2)$$

with $w(t_0) = (2, 2)$ and $\dot{w}(t_0) = (0, 0)$. We compare the distance of the gradient to the ball $\mathbb{B}(0, r)$. We denote by $t \mapsto g(t) = \nabla f(z(t))$ the solution of the first-order Riemannian gradient flow (6.1) and $s \mapsto w(s)$ the solution of the accelerated second-order dynamic (6.2). Using Theorem 3.2, we know that the function $t \mapsto \text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r))$ is a decreasing function, and the following convergence rate is satisfied: as $t \rightarrow +\infty$

$$\text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r)) = o\left(\frac{1}{t}\right).$$

While using Theorem 5.7, we know that

$$\text{dist}^2\left(w(s) + \frac{s}{\alpha-1}\dot{w}(s), \mathbb{B}(0, r)\right) = o\left(\frac{1}{s^2}\right) \text{ and } \text{dist}^2(w(s), \mathbb{B}(0, r)) = o\left(\frac{1}{s^2}\right).$$

In Figure 5, we have plotted (in the semilogy scale plot) four functions: $t \mapsto \text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r))$ in black, $s \mapsto \text{dist}^2(w(s), \mathbb{B}(0, r))$ in red, $s \mapsto \text{dist}^2\left(w(s) + \frac{s}{\alpha-1}\dot{w}(s), \mathbb{B}(0, r)\right)$ in green and $s \mapsto \frac{1}{s^2}$

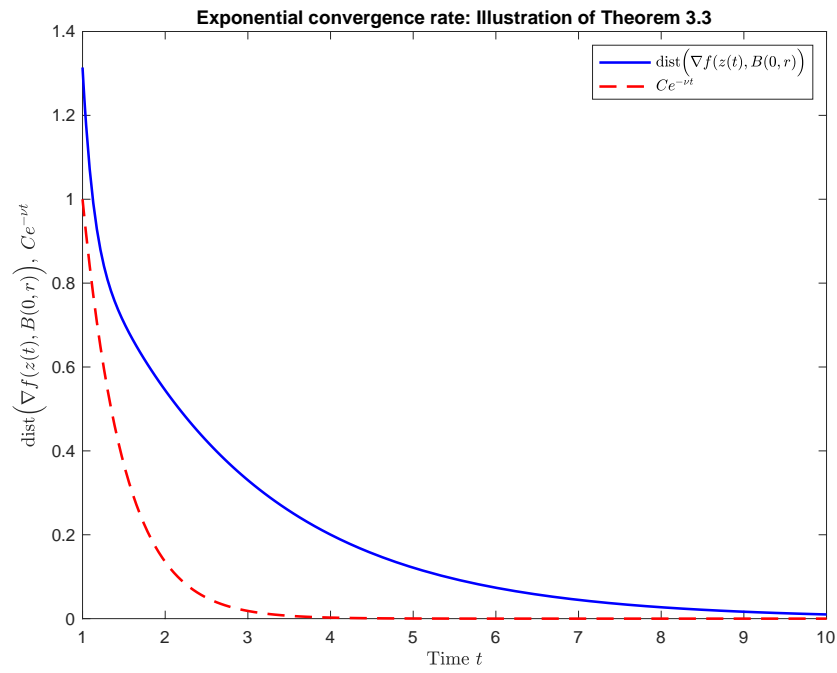
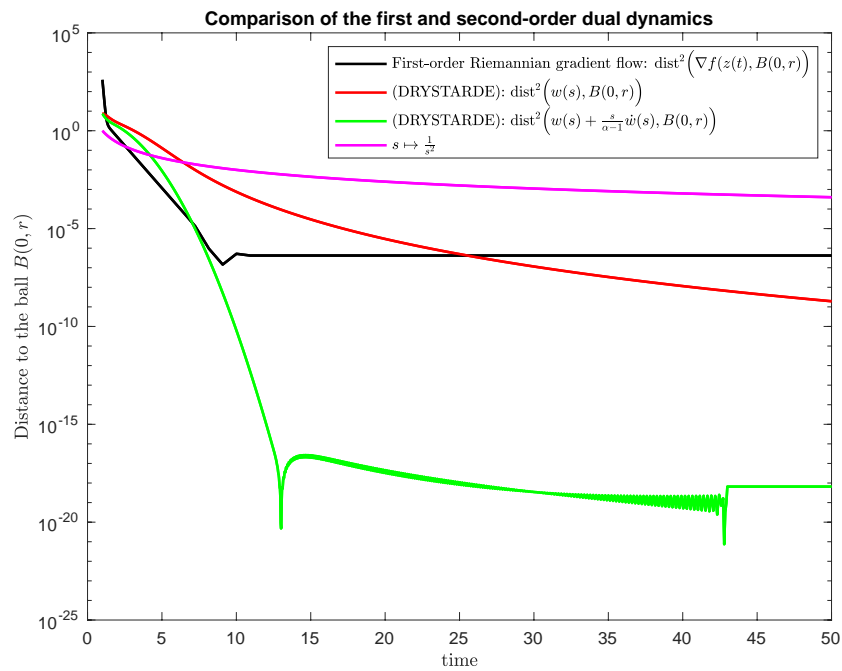

 Figure 4: Illustration of Theorem 2.1 with $r = 0.1$ and $z_0 = (2, 0.5)$.


Figure 5: Illustration of Theorem 5.7

in magenta. We observe in this example that the second-order dynamic (DRYSTARDE) outperforms the first-order Riemannian dynamic in terms of convergence rate, as the distance to the gradient converges quickly. This contrasts with Nesterov's accelerated gradient method, which aims for fast convergence of objective function values, whereas our approach focuses on rapid convergence of the objective function's gradients. By utilizing a Riemannian gradient structure, we can achieve fast convergence rates for both the system and its dual, which governs the evolution of the gradients. This enables us to effectively minimize the differentiable function f subject to the dry friction potential $\varphi(x) = r\|x\|$, resulting in faster and more precise optimization outcomes, especially when the distance between the gradient ∇f along the trajectory and the boundary of the ball is taken into consideration as a criterion.

Example 6.5 Let us consider a least square objective function of the form:

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2,$$

with $A \in \mathbb{R}^{m \times n}$ matrix such that $Q = A^T A$ is positive definite and $b \in \mathbb{R}^m$ a random vector with two random initial conditions x_0 and x_1 in \mathbb{R}^n . In the numerical test, we selected a matrix A with dimensions $m = 690$ and $n = 14$.

To solve the two dynamics (6.1) and (6.2) using the objective function $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, we need to express the gradient and Hessian of f and f^* in terms of the matrix A and the vector b . We have,

$$\nabla f(x) = A^T(Ax - b), \quad \nabla^2 f(x) = Q = A^T A, \quad \nabla f^*(y) = Q^{-1}y + Q^{-1}A^T b \text{ and } \nabla^2 f^*(y) = Q^{-1}.$$

Therefore, we can rewrite the dynamic (3.8) as

$$\dot{g}(t) = -Qg(t) + Q \operatorname{proj}_{\mathbb{B}(0,r)}(g(t)), \quad (6.3)$$

where $g(t) = \nabla f(z(t)) = Qz(t) - A^T b$ and $z(t)$ is the solution of (DRYGRAD) with the initial condition $z(0) = x_0$.

Similarly, we can rewrite the dynamic (6.2) as

$$\ddot{w}(s) + \frac{\alpha}{s}\dot{w}(s) + \frac{2s}{\alpha-1}Q\dot{w}(s) + 2Qw(s) - 2Q \operatorname{proj}_{\mathbb{B}(0,r)}\left(w(s) + \frac{s}{\alpha-1}\dot{w}(s)\right) = 0, \quad (6.4)$$

with initial condition $w(s_0) = x_0$ and $\dot{w}(s_0) = 0$.

Four functions are plotted in Figure 6. As a reference the magenta curve represents $s \mapsto \frac{1}{s^2}$. The black curve represents $t \mapsto \operatorname{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r))$, the red curve represents $s \mapsto \operatorname{dist}^2(w(s), \mathbb{B}(0, r))$, the green curve represents $s \mapsto \operatorname{dist}^2\left(w(s) + \frac{s}{\alpha-1}\dot{w}(s), \mathbb{B}(0, r)\right)$. In accordance with Theorem 5.7, we observe in Figure 6 that

$$\operatorname{dist}^2(w(s), \mathbb{B}(0, r)) = o\left(\frac{1}{s^2}\right) \text{ and } \operatorname{dist}^2\left(w(s) + \frac{s}{\alpha-1}\dot{w}(s), \mathbb{B}(0, r)\right) = o\left(\frac{1}{s^2}\right).$$

In this example, we notice that the convergence rate is faster for the second-order dynamic (DRYSTARDE) than for the first-order Riemannian dynamic, since the distance of w (the average of the gradients of f) to the ball decreases rapidly.

As a summary of this section, the Dual Dynamic (DRYSTARDE) is an interesting system that offers additional convergence results compared to other methods. Moreover, its acceleration is even more promising as it allows for rapid convergence towards an average of gradients $w(s)$ (as per formula (5.5)). Despite this, understanding the rapid convergence of gradients remains a challenge, and further research is needed to fully comprehend this phenomenon. Nonetheless, the Dual Dynamic represents a significant step towards resolving these delicate issues, as it aims to achieve rapid convergence towards an extremal point in the sense of Ekeland, which is closely related to the concept of Pareto optimality.

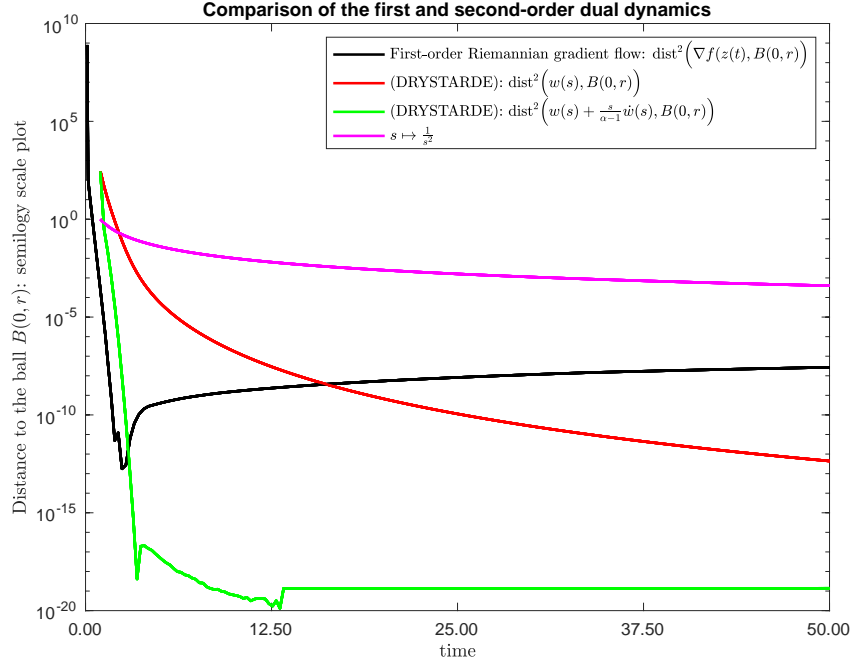


Figure 6: Semilogy scale plots of the functions $t \mapsto \text{dist}^2(\nabla f(z(t), B(0, r)))$, $s \mapsto \text{dist}^2(w(s), B(0, r))$, $s \mapsto \text{dist}^2\left(w(s) + \frac{s}{\alpha-1} \dot{w}(s), B(0, r)\right)$ and $s \mapsto \frac{1}{s^2}$

7 Conclusion, perspective

In this paper, we analyze the large-time behavior of inertial dynamics with dry friction in a Hilbert setting for convex differentiable optimization problems. The analysis relied on the acceleration method developed by Attouch, Bot and Nguyen [17] for the continuous steepest descent method, which combines time scaling and averaging. By starting from a doubly nonlinear first-order evolution equation, a time scaling and averaging technique is applied 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 convergence rates do not require developing a Lyapunov analysis for inertial systems, but instead relies on the convergence results for the original first-order system which is studied carefully and tools from differential and integral calculus. However, many questions still remain open and need further investigations. One important area that requires further study is the nonsmooth case where f is assumed to be convex (not necessarily smooth). Given $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ a closed, convex and proper function, a natural extension of (DRYGRAD) is obtained by replacing the gradient of f at z by $\partial f(z)$, which is the subdifferential of f at z in the sense of convex analysis. In doing so, we obtain a first-order differential inclusion with two nonsmooth potentials ϕ and f . We can expect that the scaling and averaging technique of this new differential inclusion would provide a second-order dynamic of the form (DRYSTAR) with ∂f instead of ∇f . This last differential inclusion provides a continuous model of the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). In fact, for the additive composite structure $f = g + h$ with g smooth and h nonsmooth, an implicit temporal discretization with respect to the nonsmooth term ∂h and explicit discretization with respect to the smooth term ∇g gives a proximal-gradient algorithm of the FISTA type. FISTA was introduced by [36], based on the Nesterov Accelerated Gradient method. Since the velocity \dot{x}

could be discontinuous in this case, the acceleration \ddot{x} might be a Radon measure (shock solutions). This kind of problems can be studied within the framework of measure differential inclusions. In this case, the concept of solutions must be adapted. In the particular important case where $g(x) = \frac{1}{2}\|Ax - b\|^2$ (with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$), and the function h is prox-friendly, then a judicious change of metric brings us back to the use of gradient-based techniques, replacing the nonsmooth potential by its Moreau's envelope in the new metric. Due to space limitations, we will not study the differential inclusions with nonsmooth potential in this paper, but we can expect to get results similar to the smooth case.

An other important issue is the dual formulation of (DRYGRAD) and its generalized Riemannian gradient flow counterpart (3.9). The differentiability or 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. These questions highlight the need for further research in the field of convex optimization involving dry friction. The numerical experiments indicate that (DRYSTARDE) exhibits good convergence results and promising acceleration towards an average of gradients. However, comprehending the rapid convergence of gradients remains a challenge, requiring further research.

In this paper, we have focused only on the continuous-time case, with a particular interest in the analysis of the associated dynamical systems. However, it would be important to explore the temporal discretization of these dynamics and investigate the convergence of the associated algorithms. Furthermore, it would be interesting to test these new algorithms on a wide range of optimization problems of varying sizes. By doing so, we can gain a deeper understanding of these algorithms and their effectiveness in optimization.

A Auxiliary results

A.1 Convergence of values for (DRYGRAD)

As we have already pointed out, the most meaningful quantity in (DRYGRAD) is the gradient term $\nabla f(z(t))$ for which we have the minimization principle

$$\text{dist}^2(\nabla f(z(t)), \mathbb{B}(0, r)) \downarrow 0$$

and the corresponding convergence rates. Although convergence rates of values can also be calculated, they are less relevant and are not associated with a minimization principle. Unlike classical steepest descent, (DRYGRAD) does not converge to the minimum value of f , but instead reaches a point where $\|\nabla f(z_\infty)\| \leq r$. This concept is made precise in the following statement.

Theorem A.1 *Let $z : [t_0, +\infty[\rightarrow \mathcal{H}$ be a global solution trajectory of (DRYGRAD). Then, $t \mapsto f(z(t))$ is a decreasing function, and the following convergence rate is satisfied: as $t \rightarrow +\infty$*

$$f(z(t)) - f_* = o\left(\frac{1}{\ln t}\right),$$

where $f_* = \inf\{f(z(t)) : t \geq t_0\}$.

Proof. According to (2.3), we have

$$\|\dot{z}(t)\|^2 + \frac{d}{dt}f(z(t)) \leq 0. \quad (\text{A.1})$$

So $\frac{d}{dt}f(z(t)) \leq 0$, which implies that (DRYGRAD) is a descent method. We analyze the case where convergence holds in infinite time. The case of finite convergence requires a different analysis. So let us introduce

$$f_* = \inf\{f(z(t)) : t \geq t_0\}.$$

So we have $f(z(t)) \downarrow f_*$ as $t \rightarrow +\infty$. Let us rewrite (DRYGRAD) as the non-autonomous gradient system

$$\dot{z}(t) + \gamma(t)\nabla f(z(t)) = 0, \quad (\text{A.2})$$

where

$$\gamma(t) = \frac{1}{1 + \frac{r}{\|\dot{z}(t)\|}} = \frac{\|\dot{z}(t)\|}{r + \|\dot{z}(t)\|}.$$

We have already proved that $\|\dot{z}(t)\| = o\left(\frac{1}{t}\right)$. Therefore, given $\epsilon > 0$, for t sufficiently large

$$\gamma(t) = \frac{\|\dot{z}(t)\|}{r + \|\dot{z}(t)\|} \leq \frac{\epsilon}{t}.$$

Let us now interpret (A.2) as a time rescaled steepest descent. Set $t = \theta(s)$ where θ is an increasing function that will be fixed later. Set $y(s) = z(\theta(s))$. We have

$$\dot{y}(s) + \gamma(\theta(s))\dot{\theta}(s)\nabla f(y(s)) = 0. \quad (\text{A.3})$$

Take θ such that

$$\gamma(\theta(s))\dot{\theta}(s) = 1.$$

Let us integrate this differential equation, and introduce $\Gamma = \int \gamma$ a primitive of γ . Therefore $\Gamma' = \gamma \geq 0$ which implies that Γ is an increasing function. We thus have

$$\frac{d}{ds}\Gamma(\theta(s)) = 1,$$

which gives

$$\Gamma(\theta(s)) = s.$$

So, θ is the inverse function of Γ , which implies that it is an increasing function. A careful inspection shows that $\lim_{t \rightarrow +\infty} \Gamma(t) = \Gamma_\infty < +\infty$, because $x(\cdot)$ has finite length. So the time scaling is between $t \in [s_0, +\infty[$ and $s = \theta^{-1}(t) = \Gamma(t) \in [s_0, \Gamma_\infty[$. Since $\gamma(t) \leq \frac{\epsilon}{t}$, and $\Gamma' = \gamma$, we have

$$\Gamma(s) \leq \epsilon \ln(s).$$

Therefore

$$s = \Gamma(\theta(s)) \leq \epsilon \ln(\theta(s)),$$

which gives

$$\theta(s) \geq \exp\left(\frac{1}{\epsilon}s\right).$$

On the other hand, according to a classical convergence property of the classical descent method

$$\dot{y}(s) + \nabla f(y(s)) = 0, \quad (\text{A.4})$$

we have

$$f(y(s)) - f_* \leq f(y(s)) - \inf f = o\left(\frac{1}{s}\right).$$

Since $y(s) = z(\theta(s))$, we get

$$f(z(\theta(s))) - f_* = o\left(\frac{1}{s}\right).$$

Since for z , we have a descent method and $\theta(s) \geq \exp(\frac{1}{\epsilon}s)$ we deduce that

$$f(z(\exp(\frac{1}{\epsilon}s))) - f_* \leq f(z(\theta(s))) - f_* = o\left(\frac{1}{s}\right).$$

This gives

$$f(z(t)) - f_* = o\left(\frac{1}{\ln t}\right),$$

which completes the proof. ■

A.2 Technical lemmas

Lemma A.1 *Let $a : [s_0, +\infty[\rightarrow \mathbb{R}$ be a positive real valued function which verifies $\lim_{u \rightarrow +\infty} a(u) = 0$. Take $\alpha > 1$. Then $\lim_{s \rightarrow +\infty} A(s) = 0$, where*

$$A(s) = \frac{1}{s^{\alpha-1}} \int_{s_0}^s u^{\alpha-2} a(u) du.$$

Proof. Given $\epsilon > 0$, let us take T_ϵ such that $s_0 < T_\epsilon$ and $a(u) \leq \epsilon$ for $t \geq T_\epsilon$. For $s > T_\epsilon$, let us write

$$\begin{aligned} A(s) &= \frac{1}{s^{\alpha-1}} \int_{s_0}^{T_\epsilon} u^{\alpha-2} a(u) du + \frac{1}{s^{\alpha-1}} \int_{T_\epsilon}^s u^{\alpha-2} a(u) du \\ &\leq \frac{1}{s^{\alpha-1}} \int_{s_0}^{T_\epsilon} u^{\alpha-2} a(u) du + \frac{\epsilon}{\alpha - 1}. \end{aligned}$$

Letting s go to $+\infty$, we obtain

$$\limsup_{s \rightarrow +\infty} A(s) \leq \frac{\epsilon}{\alpha - 1}.$$

This being true for any $\epsilon > 0$, we infer that $\lim_{s \rightarrow +\infty} A(s) = 0$, which gives the claim. ■

Acknowledgements. This research benefited from the support of the FMJH “Program Gaspard Monge for optimization and operations research and their interactions with data science”, and from the support from EDF, Thales and Orange.

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, *A variational approach to nonsmooth dynamics: applications in unilateral mechanics and electronics*, Springer Briefs in Mathematics, 2017.
- [3] S. Adly, D. Goeleven, *A stability theory for second order nonsmooth dynamical systems with application to friction problems*, Journal de Mathématiques Pures et Appliquées, 83 (2004), pp. 17–51.
- [4] S. Adly, D. Goeleven. *A nonsmooth approach for the modelling of a mechanical rotary drilling system with friction*,. Evolution Equations and Control Theory, no. 4, 915-934 (2020).

- [5] 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.
- [6] S. Adly, H. Attouch, *First-order inertial algorithms involving dry friction damping*, Math. Program., 193(1) (2022), pp. 405–445.
- [7] 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.
- [8] 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.
- [9] S. Adly, B. Brogliato, B.K. Le, *Well-posedness, robustness and stability analysis of a set-valued controller for Lagrangian systems*, SIAM J. Control Optim., 51 (2013), pp. 1592–1614.
- [10] L.X. Ahn, *Dynamics of mechanical systems with Coulomb friction*, Springer-Verlag, Berlin, 2003.
- [11] 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., 84 (2021), pp. 1687–1716.
- [12] 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.
- [13] F. Álvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., 81 (8) (2002), pp. 747–779.
- [14] 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.
- [15] G. Amontons, *On the Resistance Originating in Machines*, Proceedings of the French Royal Academy of Sciences, 1699, pp. 206–222.
- [16] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov’s rule*, Mathematical Programming, 180 (2020), 137–156.
- [17] 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.
- [18] H. Attouch, G. Buttazzo, G. Michaille, *Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization*. Second Edition, MOS/SIAM Series on Optimization, MO 17, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, (2014).
- [19] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, 263 (2017), pp. 5412–5458.
- [20] H. Attouch, A. Cabot, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., 28 (1) (2018), pp. 849–874.
- [21] 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.

- [22] H. Attouch, Z. Chbani, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim., 29 (3) 2019, pp. 2227–2256.
- [23] 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.
- [24] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$* , ESAIM-COCV, 25 (2019), Article Number 2.
- [25] H. Attouch, J. Fadili, V. Kungurtsev, *On the effect of perturbations, errors in first-order optimization methods with inertia and Hessian driven damping*, EECT, 2023, Volume 12, Issue 1, pp. 71–117. Doi: 10.3934/eect.2022022
- [26] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system*, Commun. Contemp. Math., 2 (1) (2000), pp. 1–34.
- [27] H. Attouch, P.E. Maingé, P. Redont, *A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws*, Differential Equations and Applications, 4 (1) (2012), pp. 27–65.
- [28] 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.
- [29] H. Attouch, J. Peypouquet, P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, 261 (2016), pp. 5734–5783.
- [30] 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.
- [31] H. Attouch, M. Théra, *A general duality principle for the sum of two operators*, J. Convex Analysis, 3 (1996), pp. 1–24.
- [32] 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.
- [33] J.-F. Aujol, Ch. Dossal, *Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for $b > 0$* , 2017, <https://hal.inria.fr/hal-01547251v2>.
- [34] B. Baji and A. Cabot, *An inertial proximal algorithm with dry friction: finite convergence results*, Set-Valued Analysis, 9 (1) (2006), pp. 1–23.
- [35] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
- [36] A. Beck, M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (1) (2009), pp. 183–202.
- [37] R. I. Bot, E. R. Csetnek, S.C. László, *A second order dynamical approach with variable damping to nonconvex smooth minimization*, Appl. Anal., 99(3) (2020), pp. 361–378.
- [38] R. I. Bot, E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., 54 (3) (2016), pp. 1423–1443.

- [39] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, 1972.
- [40] B. Brogliato, *Nonsmooth mechanics. Models, dynamics and control*. Third edition. Communications and Control Engineering Series. Springer, 2016.
- [41] 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.
- [42] A. Chambolle, T. Pock, *An introduction to continuous optimization for imaging*, Acta Numerica, 25 (2016), pp. 161–319.
- [43] P. Colli, A. Visintin, *On a class of doubly nonlinear evolution equations*, Communications in Partial Differential Equations, 15 (5) (1990), pp. 737–756.
- [44] C.A. Coulomb, *Théorie des machines simples, en ayant egard aufrottement de leurs parties, et à la roideur des cordages*, Mem. Math Phys., Paris, X, 1785, pp. 161–342.
- [45] L. Da Vinci, (1519), *The Notebooks*, Ed. Jean Paul Richter, New York: Dover Pub. Inc., 1970.
- [46] 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.
- [47] E. D. Dolan and J. J. Moré, *Benchmarking Optimization Software with Performance Profiles*, Math. Program., 91 (2002), pp. 201–213.
- [48] E. Ghadimi, H. R. Feyzmahdavian, M. Johansson, *Global convergence of the heavy-ball method for convex optimization*, in 2015 European Control Conference, July 2015, pp. 310–315.
- [49] G. Gorni, *Conjugation and second-order properties of convex functions*, J. Math. Anal. Appl., **158**, 293–315 (1991).
- [50] A. Haraux, M. A. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, J. Differential Equations, 144 (2) (1998), pp. 313–320.
- [51] A. Haraux, M. A. Jendoubi, *The convergence problem for dissipative autonomous systems*, Classical methods and recent advances, January 30, Springer, 2015.
- [52] C. Lemarechal, C. Sagastizábal. *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*. SIAM J. Optim. 7 (1997), no. 2, 367–385.
- [53] J. Liang, J. Fadili, G. Peyré, *Local linear convergence of forward-backward under partial smoothness*, Advances in Neural Information Processing Systems, 2014, pp. 1970–1978.
- [54] 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.
- [55] J.J. Moreau, *Unilateraly and dry friction in the dynamics of rigid body collections*, Proc. Contact Mechanics Int. Symp., (Ed. Curnier A.), Presses Polytechniques et Universitaires Romandes, 31–48, 1992.
- [56] J.J. Moreau, *Unilateral contact and dry friction in finite freedom dynamics*, Non-Smooth Mechanics and Applications, **302** of CISM Courses and Lectures (Eds. Moreau J.J. and Panagiotopoulos P.D.), 1–82, Springer, Wien, 1988.

- [57] A. J. Morin, *New Friction Experiments carried out at Metz in 1831-1833*, Proceedings of the French Royal Academy of Sciences, 4 1833, pp. 1-128.
- [58] P.D. Panagiotopoulos, *Non-Convex Superpotentials in the Sense of F.H. Clarke and Applications*, Mech. Res. Comm., 8 (1981), pp. 335–340.
- [59] J. Peypouquet, S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. Convex Anal, 17 (3-4) (2010), pp. 1113–1163.
- [60] B.T. Polyak, *Some methods of speeding up the convergence of iterative methods*, Z. Vysht Math. Fiz., 4 (1964), pp. 1–17.
- [61] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).
- [62] O. Reynolds, *On the Theory of Lubrication and its application to Mr. Beauchamp Tower's experiments, including an experimental determination of the viscosity of olive oil*, Phil. Trans. Royal Soc., 177 (1886), pp. 157–234.
- [63] R.T. Rockafellar and R. Wets, *Variational analysis*, Springer, Berlin, 1998.
- [64] 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.
- [65] 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.
- [66] J. W. Siegel, *Accelerated first-order methods: Differential equations and Lyapunov functions*, arXiv: Optimization and Control (math.OC):1903.05671v1 (2019).
- [67] R. Stribeck, *Die Wesentlichen Eigenschaften der Gleit-und Rollenlager*, Z. Verein. Deut. Ing. , 46(38) (1902), pp. 1341–1348.
- [68] 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.