J. Eur. Math. Soc. (Online first) DOI 10.4171/JEMS/1231

Published by EMS Press

Hedy Attouch · Radu Ioan Boţ · Ernö Robert Csetnek

Fast optimization via inertial dynamics with closed-loop damping

Received September 10, 2020

Abstract. In a real Hilbert space $\mathcal{H}$, in order to develop fast optimization methods, we analyze the asymptotic behavior, as time t tends to infinity, of a large class of autonomous dissipative inertial continuous dynamics. The function $f : \mathcal{H} \to \mathbb{R}$ to be minimized (not necessarily convex) enters the dynamic via its gradient, which is assumed to be Lipschitz continuous on bounded subsets of $\mathcal{H}$. This results in autonomous dynamical systems with nonlinear damping and nonlinear driving force. We first consider the case where the damping term $\partial\phi(\dot{x}(t))$ acts as closed-loop control of the velocity. The damping potential $\phi : \mathcal{H} \to \mathbb{R}_+$ is a convex continuous function which reaches its minimum at the origin. We show the existence and uniqueness of a global solution to the associated Cauchy problem. We analyze the asymptotic convergence and the convergence rates of the trajectories generated by this system. To do this, we use techniques from optimization, control theory, and PDEs: Lyapunov analysis based on the decreasing property of an energy-like function, quasi-gradient and Kurdyka–Łojasiewicz theory, and monotone operator theory for wave-like equations. Convergence rates are obtained based on the geometric properties of the data f and ϕ. We put forward minimal hypotheses on the damping potential ϕ guaranteeing the convergence of trajectories, thus showing the dividing line between strong and weak damping. When f is strongly convex, we give general conditions on the damping potential ϕ which provide exponential convergence rates. Then, we extend the results to the case where additional Hessian-driven damping enters the dynamic, which reduces the oscillations. Finally, we consider a new inertial system where damping jointly involves the velocity $\dot{x}(t)$ and the gradient $\nabla f(x(t))$. This study naturally leads to similar results for proximal-gradient algorithms obtained by temporal discretization; some of them are studied in the article. In addition to its original results, this work surveys numerous works devoted to the interaction between damped inertial continuous dynamics and numerical optimization algorithms, with an emphasis on autonomous systems, adaptive procedures, and convergence rates.

Keywords. Closed-loop damping, convergence rates, damped inertial gradient systems, Hessian damping, quasi-gradient systems, Kurdyka–Łojasiewicz inequality, maximally monotone operators

Hedy Attouch: IMAG, Université Montpellier, CNRS, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France; hedy.attouch@umontpellier.fr

Radu Ioan Boţ: Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria; radu.bot@univie.ac.at

Ernö Robert Csetnek: Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria; robert.csetnek@univie.ac.at

Mathematics Subject Classification (2020): 37N40, 46N10, 49M15, 65K05, 65K10, 90C25

1. Introduction

Our work is part of the active research stream which studies the close link between continuous dissipative dynamical systems and optimization algorithms obtained by temporal discretization. In this context, second-order evolution equations provide a natural and intuitive way to speed up algorithms. Then, the optimization properties come from the damping term. It is the skill of the mathematician to design this term to obtain rapidly converging trajectories and algorithms (ideally, to obtain optimal convergence rates).

More precisely, we will consider the following system (ADIGE) which covers a large number of situations. Let $\mathcal{H}$ be a real Hilbert space endowed with a scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Let $f : \mathcal{H} \to \mathbb{R}$ be a differentiable function (not necessarily convex) whose gradient $\nabla f : \mathcal{H} \to \mathcal{H}$ is Lipschitz continuous on bounded subsets of $\mathcal{H}$, and such that $\inf_{\mathcal{H}} f > -\infty$ (when considering the Hessian of f, we will assume that f is twice differentiable). Our objective is to study from the optimization point of view the Autonomous Damped Inertial Gradient Equation

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

where the damping term $\mathcal{G}(\dot{x}(t), \nabla f(x(t)), \nabla^2 f(x(t)))$ acts as a closed-loop control. Under suitable assumptions, this term will induce dissipative effects, which tend to stabilize asymptotically (i.e. as $t \to +\infty$) the trajectories to critical points of f (minimizers in the case where f is convex). We will use the generic terminology *damped inertial continuous dynamics* to designate second-order evolution systems which have a strict Lyapunov function. To be specific we will refer to (ADIGE) or to some of its particular cases. From this, we can distinguish two distinct classes of dynamics and algorithms, depending on whether the damping term involves coefficients which are given a priori as functions of time (open-loop damping, nonautonomous dynamic), or is a feedback of the current state of the system (closed-loop damping, adaptive methods, autonomous dynamic). We will use these terms describing the two classes of dynamics indiscriminately, but to be precise they correspond to the cases of autonomous and nonautonomous dynamics respectively. Indeed, one of our objectives is to understand if the closed-loop damping can achieve (and possibly improve) the fast convergence properties of the accelerated gradient method of Nesterov. Recall that, in convex optimization, the accelerated gradient method of Nesterov (which is associated with a nonautonomous damped inertial dynamic) provides convergence rate of order $1/t^2$, which is optimal for first-order methods (involving only evaluations of ∇f at iterates). This justifies the importance of inertial dynamics for developing fast optimization methods (recall that the continuous steepest descent, which is a first-order evolution equation, only guarantees the convergence rate of $1/t$ for general convex functions). Closely related questions concern the impact of geometric properties of data (damping term, objective function) on the convergence rates of trajectories and iterations. This is a wide subject which involves continuous optimization, as well as the study of the stabilization of oscillating systems in mechanics and physics. Due to the highly nonlinear characteristics of (ADIGE) (nonlinearity occurs both in the damping term and in the gradient of f), our convergence analysis will mainly rely on the combination of the quasi-gradient approach for inertial systems initiated by Bégout–Bolte–Jendoubi [44] with the theory of Kurdyka–Łojasiewicz. The price to pay is that some of the results are only valid in finite-dimensional Hilbert spaces. It should be noted that the relative simplicity of the functional framework (single function space, differentiable objective function) does not allow direct application to the corresponding PDEs. Our objective is mainly the study of optimization problems, but the Lyapunov analysis developed in the article can be a very useful guide for its extension to the PDE framework, as was done in [9, 40, 56].

1.1. Presentation of the results

For each of the following systems, we will prove existence and uniqueness of the solution of the Cauchy problem, and study its asymptotic behavior.

1.1.1. ADIGE-V. Our study mainly concerns the differential inclusion

$$\text{(ADIGE-V)} \qquad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)), \tag{1}$$

where $\phi : \mathcal{H} \to \mathbb{R}$ is a convex continuous function which achieves its minimum at the origin, and the operator $\partial\phi : \mathcal{H} \to 2^{\mathcal{H}}$ is its convex subdifferential. The damping term $\mathcal{G}$ depends only on the velocity, which is reflected by the suffix V. This model encompasses several classic situations:

• The case $\phi(u) = \frac{\gamma}{2}\|u\|^2$ corresponds to the Heavy Ball with Friction method

$$\text{(HBF)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) = 0, \tag{2}$$

introduced by B. Polyak [85, 86] and further studied by Attouch–Goudou–Redont [28] (exploration of local minima), Alvarez [8] (convergence in the convex case), Haraux–Jendoubi [70, 71] (convergence in the analytic case), Bégout–Bolte–Jendoubi [44] (convergence based on the Kurdyka–Łojasiewicz property), to cite part of the rich literature devoted to this subject.

• The case $\phi(u) = r\|u\|$ corresponds to the dry friction effect. Then (ADIGE-V) is a differential inclusion (because ϕ is nondifferentiable), which, when $\dot{x}(t)$ is not equal to zero, reads

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

The importance of this case in optimization comes from the finite time stabilization property of trajectories, which is satisfied generically with respect to initial data. The rigorous mathematical treatment of this case was given by Adly–Attouch–Cabot [5] and Amann–Diaz [12]; see Adly–Attouch [2–4] for recent developments.

• Taking $\phi(u) = \frac{1}{p}\|u\|^p$ with $p \geq 1$ allows one to treat these questions in a unifying way. We will pay particular attention to the role played by the parameter p in the asymptotic convergence analysis. For $p > 1$ the dynamic reads $\ddot{x}(t) + \|\dot{x}(t)\|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0$. We will see that the case $p = 2$ separates weak damping ($p > 2$) from strong damping ($p < 2$), hence the importance of this case.

1.1.2. ADIGE-VH. We will extend the previous results to the differential inclusion

$$\text{(ADIGE-VH)} \quad \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,$$

which, besides a damping potential ϕ as above acting on the velocity, also involves geometric damping driven by the Hessian of f, hence the terminology. The inertial system

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

was introduced in [11]. In the same spirit as (HBF), the dynamic $\text{(DIN)}_{\gamma,\beta}$ contains a *fixed* positive viscous friction coefficient $\gamma > 0$. The introduction of Hessian-driven damping allows one to damp the transversal oscillations that might arise with (HBF), as observed in [11] in the case of the Rosenbrock function. The need to consider geometric damping adapted to f had already been observed by Alvarez [8] who considered the inertial system

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

where $\Gamma : \mathcal{H} \to \mathcal{H}$ is a linear positive anisotropic operator (see also [47]). But still this damping operator is fixed. For a general convex function, Hessian-driven damping in $\text{(DIN)}_{\gamma,\beta}$ acts similarly, in an adaptive way. Here (DIN) stands for Dynamic Inertial Newton system. It refers to the natural link between this dynamic and the continuous Newton method (see Attouch–Svaiter [37]). Recent studies have been devoted to the 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 oscillatory aspects; see Attouch–Peypouquet–Redont [35], Attouch–Chbani–Fadili–Riahi [24], and Shi–Du–Jordan–Su [89].

1.1.3. ADIGE-VGH. Finally, we will consider the new dynamical system

$$\text{(ADIGE-VGH)} \quad \ddot{x}(t) + \partial\phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,$$

where the damping term $\partial\phi(\dot{x}(t) + \beta\nabla f(x(t)))$ involves both the velocity vector and the gradient of the potential function f. The parameter $\beta \geq 0$ is attached to the geometric damping induced by the Hessian. As previously, ϕ is a damping potential function. Assuming that f is convex and ϕ is a sharp function at the origin, that is, $\phi(u) \geq r\|u\|$ for some $r > 0$, we will show that, for each trajectory generated by (ADIGE-VGH),

(i) $x(\cdot)$ converges weakly as $t \to +\infty$, and its limit belongs to $\operatorname{argmin}_{\mathcal{H}} f$;

(ii) $\dot{x}(t)$ and $\nabla f(x(t))$ converge strongly to zero as $t \to +\infty$;

(iii) after a finite time, $x(\cdot)$ follows the steepest descent dynamic.

1.2. Contents

The paper is organized in accordance with the above presentation. In Section 2, we recall some classical facts concerning the Heavy Ball with Friction method, the Su–Boyd–Candès dynamic approach to the Nesterov method, Hessian-driven damping, and dry friction. Then, we successively examine each of the cases considered above: Sections 3–7 are devoted to the closed-loop control of the velocity, which is the main part of our study. We show the existence and uniqueness of a global solution for the Cauchy problem, the exponential convergence rate for f strongly convex, the effect of weak damping, and finally analyze the convergence under the Kurdyka–Łojasiewicz (KŁ) property. Section 8 concerns some related algorithmic results. Section 9 is devoted to closed-loop damping with Hessian-driven damping. Section 10 is devoted to closed-loop damping involving the velocity and the gradient. We conclude by mentioning several lines of research for the future.

2. Classical facts

Let us recall some classical facts which will serve as comparison tools.

2.1. (HBF) dynamic system

The *Heavy Ball with Friction system*

$$\ddot{x}(t) + r\dot{x}(t) + \nabla f(x(t)) = 0 \tag{HBF$_r$}$$

was introduced by B. Polyak [85,86]. It involves a fixed viscous friction coefficient $r > 0$. Assuming that f is a convex function such that $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$, we know, by Alvarez's theorem [8], that each trajectory of $(\mathrm{HBF})_r$ converges weakly, and its limit belongs to $\operatorname{argmin}_{\mathcal{H}} f$. In addition, we have the following convergence rates, the proof of which (see [19]) is based on the decrease property of the Lyapunov function

$$\mathcal{E}(t) := \frac{1}{r^2}(f(x(t)) - \min\nolimits_{\mathcal{H}} f) + \frac{1}{2}\left\| x(t) - x^* + \frac{1}{r}\dot{x}(t)\right\|^2,$$

where $x^* \in \operatorname{argmin}_{\mathcal{H}} f$.

Theorem 1. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function of class $\mathcal{C}^1$ such that* $\operatorname{argmin} f \neq \emptyset$, *and let r be a positive parameter. Let $x(\cdot) : [0, +\infty[\to \mathcal{H}$ be a solution of* $(\mathrm{HBF})_r$. *Set $x(0) = x_0$ and $\dot{x}(0) = x_1$. Then*

(i) $\displaystyle\int_0^{+\infty} (f(x(t)) - \min\nolimits_{\mathcal{H}} f)\, dt < +\infty, \quad \int_0^{+\infty} t\|\dot{x}(t)\|^2\, dt < +\infty;$

(ii) $\displaystyle f(x(t)) - \min\nolimits_{\mathcal{H}} f \leq \frac{C(x_0, x_1)}{t}, \quad \|\dot{x}(t)\| \leq \frac{\sqrt{2C(x_0, x_1)}}{\sqrt{t}},$ *where*

$$C(x_0, x_1) =: \frac{3}{2r}(f(x_0) - \min\nolimits_{\mathcal{H}} f) + r\,\mathrm{dist}(x_0, \operatorname{argmin} f)^2 + \frac{5}{4r}\|x_1\|^2;$$

(iii) $\displaystyle f(x(t)) - \min\nolimits_{\mathcal{H}} f = o\left(\frac{1}{t}\right)$ *and* $\displaystyle\|\dot{x}(t)\| = o\left(\frac{1}{\sqrt{t}}\right)$ *as $t \to +\infty$.*

Let us now consider the case of a strongly convex function. Recall that a function $f : \mathcal{H} \to \mathbb{R}$ is *μ-strongly convex* for some $\mu > 0$ if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex. We have the exponential convergence rate, whose proof relies on the decrease properties of the Lyapunov function

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

where x^* is the unique minimizer of f.

Theorem 2. *Suppose that $f : \mathcal{H} \to \mathbb{R}$ is a function of class $\mathcal{C}^1$ which is μ-strongly convex for some $\mu > 0$. Let $x(\cdot) : [0, +\infty[\to \mathcal{H}$ be a solution of*

$$\ddot{x}(t) + 2\sqrt{\mu}\,\dot{x}(t) + \nabla f(x(t)) = 0. \tag{3}$$

Set $x(0) = x_0$ and $\dot{x}(0) = x_1$. Then for all $t \geq 0$,

$$f(x(t)) - \min_{\mathcal{H}} f \leq Ce^{-\sqrt{\mu}t},$$

where $C := f(x_0) - \min_{\mathcal{H}} f + \mu\,\mathrm{dist}(x_0, S)^2 + \|x_1\|^2$.

A recent account on the best tuning of the damping coefficient can be found in Aujol–Dossal–Rondepierre [39]. The above results show the important role in convergence rates played by the geometric properties of the data. Apart from the convex case, the first convergence result for (HBF) was obtained by Haraux–Jendoubi [70] in the case where $f : \mathbb{R}^n \to \mathbb{R}$ is a real-analytic function. They have shown the central role played by Łojasiewicz's inequality (see also [60]). Then, on the basis of Kurdyka's work in real algebraic geometry, Łojasiewicz's inequality was extended in [45] by Bolte–Daniilidis–Ley–Mazet to a large class of tame functions, possibly nonsmooth. This is the Kurdyka–Łojasiewicz inequality, (KŁ) for short. The convergence of first and second-order proximal-gradient dynamical systems in the context of the (KŁ) property was obtained by Boţ–Csetnek [48] and Boţ–Csetnek–László [50]. The (KŁ) property will be a key tool for obtaining convergence rates based on the geometric properties of the data. Note that this theory only works in the finite-dimensional setting[1] (the infinite-dimensional setting is a difficult topic which is the subject of current research), and only for autonomous systems. This explains why working with autonomous systems is important: it allows us to use the powerful (KŁ) theory.

2.2. Su–Boyd–Candès dynamic approach to Nesterov accelerated gradient method

The nonautonomous system

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

[1] In the field of PDEs, the Łojasiewicz–Simon theory [62] makes it possible to deal with certain classes of problems, such as semi-linear equations.

will serve as a reference to compare our results with the open-loop damping approach. It was introduced in the context of convex optimization by Su–Boyd–Candès [90]. As a specific feature, the viscous damping coefficient α/t vanishes (tends to zero) as t goes to infinity, hence the terminology "Asymptotic Vanishing Damping". This contrasts with (HBF) where the viscous friction coefficient is fixed, which precludes obtaining fast convergence of the values for general convex functions. Recall the main results concerning the asymptotic behavior of the trajectories generated by $(\text{AVD})_\alpha$.

- For $\alpha \geq 3$, for each trajectory $x(\cdot)$ of $(\text{AVD})_\alpha$, $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/t^2)$ as $t \to +\infty$.
- For $\alpha > 3$, each trajectory converges weakly to a minimizer of f [25]. In addition, it is shown in [33] and [78] that $f(x(t)) - \inf_{\mathcal{H}} f = o(1/t^2)$ as $t \to +\infty$.
- For $\alpha \leq 3$, we have $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(t^{-2\alpha/3})$ [13, 26].
- $\alpha = 3$ is a critical value.[2] It corresponds to the historical case studied by Nesterov [80, 81].

The implicit time discretization of $(\text{AVD})_\alpha$ provides an inertial proximal algorithm that enjoys the same properties as the continuous dynamics. Replacing the proximal step by a gradient step gives the following Nesterov accelerated gradient method (illustrated in Figure 1):

$$\begin{cases} y_k = x_k + (1-\alpha/k)(x_k - x_{k-1}), \\ x_{k+1} = y_k - s\nabla f(y_k), \end{cases}$$

which still enjoys the same properties when the step size s is less than the inverse of the Lipschitz constant of ∇f. Based on the dynamic approach above, many recent studies have been devoted to the convergence properties of the sequences (x_k) and (y_k), which has led to a better understanding and improvement of Nesterov's accelerated gradient algorithm [13, 20, 24–26, 33, 59, 90], and of the Ravine algorithm [64, 87].

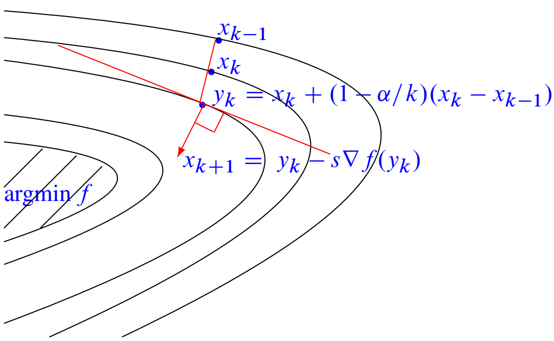

Fig. 1. Nesterov accelerated gradient method.

[2] The convergence of the trajectories is an open question in this case.

2.2.1. Optimal convergence rates. In the above results the convergence rates are optimal, that is, they can be reached, or approached arbitrarily close, as shown by the following example from [25]. Let us show that $\mathcal{O}(1/t^2)$ is the worst possible case for the rate of convergence of the values for $(\text{AVD})_\alpha$ trajectories when $\alpha \geq 3$. It is attained as a limit in the following example. Take $\mathcal{H} = \mathbb{R}$ and $f(x) = c|x|^\gamma$, where c and γ are positive parameters. We look for nonnegative solutions of $(\text{AVD})_\alpha$ of the form $x(t) = 1/t^\theta$ with $\theta > 0$. This means that the trajectory is not oscillating, it is completely damped. Let us determine the values of c, γ and θ that provide such solutions. We have

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = \theta(\theta + 1 - \alpha)\frac{1}{t^{\theta+2}}, \quad \nabla f(x(t)) = c\gamma|x(t)|^{\gamma-2}x(t) = c\gamma\frac{1}{t^{\theta(\gamma-1)}}.$$

Thus, $x(t) = 1/t^\theta$ is solution of $(\text{AVD})_\alpha$ if, and only if,

(i) $\theta + 2 = \theta(\gamma - 1)$, which is equivalent to $\gamma > 2$ and $\theta = \frac{2}{\gamma-2}$; and

(ii) $c\gamma = \theta(\alpha - \theta - 1)$, which is equivalent to $\alpha > \frac{\gamma}{\gamma-2}$ and $c = \frac{2}{\gamma(\gamma-2)}(\alpha - \frac{\gamma}{\gamma-2})$.

We have $\min f = 0$ and $f(x(t)) = \frac{2}{\gamma(\gamma-2)}(\alpha - \frac{\gamma}{\gamma-2})\frac{1}{t^{2\gamma/(\gamma-2)}}$.

The speed of convergence of $f(x(t))$ to 0 depends on the parameter γ. The exponent $\frac{2\gamma}{\gamma-2}$ is greater than 2, and tends to 2 when γ tends to $+\infty$. This limiting situation is obtained by taking a function f which becomes very flat around the set of its minimizers. Therefore, without any other geometric assumptions on f, we cannot expect a convergence rate better than $\mathcal{O}(1/t^2)$. This means that it is not possible to obtain a rate $\mathcal{O}(1/t^r)$ with $r > 2$, which holds for all convex functions. Hence, when $\alpha \geq 3$, $\mathcal{O}(1/t^2)$ is sharp. This does not contradict the rate $o(t^{-2})$ obtained when $\alpha > 3$.

2.3. Hessian-driven damping

The inertial system

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

was introduced in [11]. In line with (HBF), the viscous friction coefficient γ is a *fixed* positive real number. The introduction of Hessian-driven damping makes it possible to neutralize the oscillations likely to occur with (HBF), a key property for numerical optimization purposes.

To accelerate this system, several studies considered the case where the viscous damping is vanishing. As a model example, which is based on the Su–Boyd–Candès continuous model for the Nesterov accelerated gradient method, we have

$$(\text{DIN-AVD})_{\alpha,\beta} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0. \tag{4}$$

In connection with this system, let us cite Attouch–Peypouquet–Redont [35], Attouch–Chbani–Fadili–Riahi [24], Boţ–Csetnek–László [52], Castera–Bolte–Févotte–Pauwels [58], Kim [76], Lin–Jordan [79], Shi–Du–Jordan–Su [89]. While preserving the convergence properties of $(\text{AVD})_\alpha$, the above system provides fast convergence to zero of the

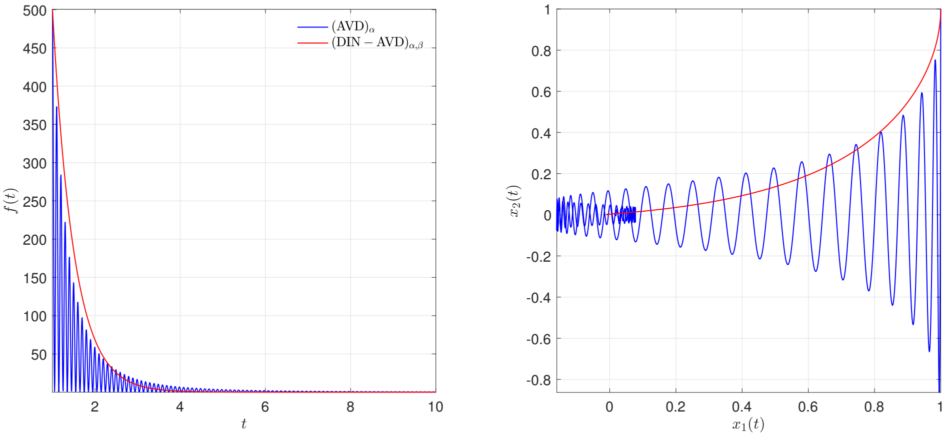

Fig. 2. Evolution of the objective (left) and trajectories (right) for $(\text{AVD})_\alpha$ ($\alpha = 3.1$) and $(\text{DIN-AVD})_{\alpha,\beta}$ ($\alpha = 3.1, \beta = 1$) on an ill-conditioned quadratic problem in $\mathbb{R}^2$.

gradients, namely $\int_{t_0}^{\infty} t^2 \|\nabla f(x(t))\|^2\, dt < +\infty$ for $\alpha \geq 3$ and $\beta > 0$, and reduces the oscillatory aspects.

To illustrate the remarkable effect of Hessian-driven damping, let us compare the two dynamics $(\text{AVD})_\alpha$ and $(\text{DIN-AVD})_{\alpha,\beta}$ on a simple ill-conditioned quadratic minimization problem. In the following example of [24], the trajectories can be computed in closed form. Take $\mathcal{H} = \mathbb{R}^2$ and $f(x_1, x_2) = \frac{1}{2}(x_1^2 + 1000x_2^2)$. We take parameters $\alpha = 3.1$, $\beta = 1$, to obey the condition $\alpha > 3$. Starting with the initial conditions $(x_1(1), x_2(1)) = (1, 1)$, $(\dot{x}_1(1), \dot{x}_2(1)) = (0, 0)$, we have the trajectories displayed in Figure 2. We observe that the wild oscillations of $(\text{AVD})_\alpha$ are neutralized by the presence of Hessian-driven damping in $(\text{DIN-AVD})_{\alpha,\beta}$.

At first glance, the presence of the Hessian may seem to cause numerical difficulties. However, this is not the case because the Hessian intervenes in the above ODE in the form $\nabla^2 f(x(t))\dot{x}(t)$, which is nothing other than the derivative with respect to time of $\nabla f(x(t))$. Thus, the temporal discretization of this dynamic provides first-order algorithms which, by comparison with the accelerated gradient method of Nesterov, contain a correction term which is equal to the difference of the gradients at two consecutive steps. The following closely related inertial system was recently introduced by Alecsa–László–Pinta [7]:

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

The link with $(\text{DIN-AVD})_{\alpha,\beta}$ results from Taylor expansion: as $t \to +\infty$ we have $\dot{x}(t) \to 0$, and so $\nabla f(x(t) + \beta\dot{x}(t)) \approx \nabla f(x(t)) + \beta\nabla^2 f(x(t))\dot{x}(t)$.

2.3.1. Hessian-driven damping and unilateral mechanics. Another motivation for the study of $(\text{DIN})_{\gamma,\beta}$ comes from mechanics, and modeling of damped shocks. In [31], Attouch–Maingé–Redont consider the inertial system with Hessian-driven damping

$$\ddot{x}(t) + \gamma\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) + \nabla g(x(t)) = 0, \tag{5}$$

where $g : \mathcal{H} \to \mathbb{R}$ is a smooth real-valued function. An interesting property of this system is that, after the introduction of an auxiliary variable y, it can be equivalently written as a first-order system involving only the time derivatives $\dot{x}(t)$, $\dot{y}(t)$ and the gradient terms $\nabla f(x(t))$, $\nabla g(x(t))$. More precisely, the system (5) is equivalent to the first-order differential equation

$$\begin{cases} \dot{x}(t) + \beta\nabla f(x(t)) + ax(t) + by(t) = 0, \\ \dot{y}(t) - \beta\nabla g(x(t)) + ax(t) + by(t) = 0, \end{cases} \tag{6}$$

where a and b are real numbers such that $a + b = \gamma$ and $\beta b = 1$. Note that (6) is different from the classical Hamiltonian formulation, which would still involve the Hessian of f. In contrast, the formulation (6) uses only first-order information on the function f (no occurrence of the Hessian of f). Replacing ∇f by ∂f in (6) allows us to extend the analysis to the case of a convex lower semicontinuous function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, and so to introduce constraints in the model. When $f = \delta_K$ is the indicator function of a closed convex set $K \subset \mathcal{H}$, the subdifferential operator ∂f takes account of the contact forces, while ∇g takes account of the driving forces. In this setting, by playing with the geometric damping parameter β, one can describe nonelastic shock laws with restitution coefficient (for more details we refer to [31] and references therein). Combination of dry friction ($\phi(u) = r\|u\|$) with Hessian damping has been considered by Adly–Attouch [3,4].

2.4. *Inertial dynamics with dry friction*

Although dry friction (also called Coulomb friction) plays a fundamental role in mechanics, its use in optimization has only recently been analyzed. Due to the nonsmooth character of the associated damping function $\phi(u) = r\|u\|$, the dynamics is a differential inclusion, which, when the speed is not equal to zero, is given by

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

In this case, the energy estimate gives $\int_0^{+\infty} \|\dot{x}(t)\|\, dt < +\infty$. Therefore, the trajectory has finite length, and it converges strongly. The limit x_∞ of the trajectory $x(\cdot)$ satisfies

$$\|\nabla f(x_\infty)\| \le r.$$

Thus, x_∞ is an "approximate" critical point of f. In practice, for optimization purposes, we choose a small $r > 0$. This amounts to solving the optimization problem $\min_{\mathcal{H}} f$ with the variational principle of Ekeland, instead of the Fermat rule. The importance of this case in optimization comes from the finite time stabilization property of trajectories, which is satisfied generically with respect to initial data. The rigorous mathematical treatment of this case has been undertaken by Adly–Attouch–Cabot [5]; see Adly–Attouch [2–4] for recent developments. Corresponding PDE results have been obtained by Amann–Diaz [12] for the nonlinear wave equation, and by Carles–Gallo [57] for the nonlinear Schrödinger equation.

2.5. Closed-loop versus open-loop damping

In the strongly convex case, the autonomous system (HBF) provides an exponential rate of convergence. On the other hand, the $(\mathrm{AVD})_\alpha$ system provides a convergence rate of order $1/t^\alpha$. Thus, in this case, closed-loop damping behaves better than open-loop damping. For general convex functions (i.e. in the worst case), we have the opposite situation: $(\mathrm{AVD})_\alpha$ provides a convergence rate $1/t^2$, while (HBF) gives only $1/t$. In this paper, we will study the impact of the choice of the damping potential on the rate of convergence. A related question is: using autonomous systems, can we obtain for general convex functions a convergence rate of order $1/t^2$, i.e. as good as the Nesterov accelerated gradient method? As we will see, to answer these questions, we will have to study different types of closed-loop damping, and rely on the geometric properties of the data. These questions fall within the framework of an active research current: Apidopoulos–Aujol–Dossal–Rondepierre [14] (geometrical properties of the data), Iutzeler–Hendricx [75] (online acceleration), Lin–Jordan [79] (control perspective on high-order optimization), Poon–Liang [88] (geometry of first-order methods and adaptive acceleration), to cite only some recent works.

3. Damping via closed-loop velocity control: existence and uniqueness

In this section, we will first introduce the notion of damping potential, and then prove the existence and uniqueness of the solution of the corresponding Cauchy problem.

3.1. Damping potential

We consider the differential inclusion

$$\text{(ADIGE-V)} \qquad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)),$$

where ϕ is a convex damping potential, as defined below.

Definition 1. A function $\phi : \mathcal{H} \to \mathbb{R}_+$ is a *damping potential* if

(i) ϕ is a nonnegative convex continuous function;

(ii) $\phi(0) = 0 = \min_{\mathcal{H}} \phi$;

(iii) the minimal section of $\partial\phi$ is bounded on bounded sets, that is, for any $R > 0$,

$$\sup_{\|u\| \le R} \|(\partial\phi)^0(u)\| < +\infty.$$

In the above, $(\partial\phi)^0(u)$ is the element of minimal norm of the closed convex nonempty set $\partial\phi(u)$ [53, Proposition 2.6]. Note that when $\mathcal{H}$ is finite-dimensional, property (iii) is automatically satisfied. Indeed, in this case, $\partial\phi$ is bounded on bounded sets [43, Proposition 16.17].

The concept of damping potential is flexible, and allows one to cover various situations. For example,

$$\phi_1(u) = \frac{\gamma}{2}\|u\|^2 + r\|u\|, \quad \phi_2(u) = \max\left\{\frac{\gamma}{2}\|u\|^2, r\|u\|\right\}$$

are damping potentials which combine dry friction with viscous damping [3].

3.2. *Existence and uniqueness results*

In this section, we study the existence and uniqueness of the solution of the Cauchy problem associated with (ADIGE-V), where ϕ is a convex damping potential. No convexity assumption is made on the function f, which is supposed to be differentiable. Since we work with autonomous (dissipative) systems, we can take an arbitrary initial time t_0. As is common, we take $t_0 = 0$, and hence work on the time interval $[0, +\infty[$.

Let us specify the notion of strong solution.

Definition 2. The trajectory $x : [0, +\infty[\to \mathcal{H}$ is said to be a *strong global solution* of (ADIGE-V) if

(i) $x \in \mathcal{C}^1([0, +\infty[; \mathcal{H})$;

(ii) $\dot{x} \in \mathrm{Lip}(0, T; \mathcal{H})$, $\ddot{x} \in L^\infty(0, T; \mathcal{H})$ for all $T > 0$;

(iii) for almost all $t > 0$, $0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t))$.

Note that since $\dot{x} \in \mathrm{Lip}(0, T; \mathcal{H})$, it is absolutely continuous on bounded time intervals, so its distribution derivative coincides with its derivative almost everywhere (which exists). Thus, the acceleration $\ddot{x}$ belongs to $L^\infty(0, T; \mathcal{H})$ for all $T > 0$, but it is not necessarily continuous. See [53, Appendix] for further details on vector-valued Lebesgue and Sobolev spaces.

Let us prove the following existence and uniqueness result for the associated Cauchy problem.

Theorem 3. *Let $f : \mathcal{H} \to \mathbb{R}$ be a differentiable function whose gradient is Lipschitz continuous on bounded subsets of $\mathcal{H}$, and such that $\inf_{\mathcal{H}} f > -\infty$. Let $\phi : \mathcal{H} \to \mathbb{R}_+$ be a damping potential (see Definition 1). Then, for any $x_0, x_1 \in \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\to \mathcal{H}$ of* (ADIGE-V) *such that $x(0) = x_0$ and $\dot{x}(0) = x_1$, that is,*

$$\begin{cases} 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)), \\ x(0) = x_0, \quad \dot{x}(0) = x_1. \end{cases}$$

Proof. We first consider the case where ∇f is Lipschitz continuous over the whole space, and then the case where it is Lipschitz continuous only on bounded sets. In both cases, the idea is to mix the existence results for ODEs which are based on the Cauchy–Lipschitz theorem with those based on the theory of maximally monotone operators. We treat the two cases independently because the proof is much simpler in the first case.

Case (a): ∇f is Lipschitz continuous on the whole space. The Hamiltonian formulation of (ADIGE-V) gives the equivalent first-order differential inclusion in the product space $\mathcal{H} \times \mathcal{H}$:

$$0 \in \dot{z}(t) + \partial\Phi(z(t)) + F(z(t)), \tag{7}$$

where $z(t) = (x(t), \dot{x}(t)) \in \mathcal{H} \times \mathcal{H}$, and

- $\Phi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is the convex function defined by $\Phi(x,u) = \phi(u)$;
- $F : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is defined by $F(x,u) = (-u, \nabla f(x))$.

Since ∇f is Lipschitz continuous on the whole space $\mathcal{H}$, we immediately see that F is a Lipschitz continuous mapping on $\mathcal{H} \times \mathcal{H}$. So, we can apply a result on evolution equations governed by Lipschitz perturbations of convex subdifferentials [53, Proposition 3.12] to conclude that (7) has a unique strong global solution with initial data $z(0) = (x_0, x_1)$.

Case (b): ∇f is Lipschitz continuous on bounded sets. The major difficulty in (ADIGE-V) is the presence of the term $\partial\phi(\dot{x}(t))$, which involves a possibly nonsmooth operator $\partial\phi$. A natural idea is to regularize this operator, and thus obtain a classical evolution equation. To this end, we use Moreau–Yosida regularization. Let us recall some basic facts concerning this regularization procedure. For any $\lambda > 0$, the *Moreau envelope* of ϕ of index λ is the function $\phi_\lambda : \mathcal{H} \to \mathbb{R}$ defined by

$$\phi_\lambda(u) = \min_{\xi \in \mathcal{H}} \left\{ \phi(\xi) + \frac{1}{2\lambda}\|u - \xi\|^2 \right\} \quad \text{for all } u \in \mathcal{H}.$$

The function ϕ_λ is convex, of class $\mathcal{C}^{1,1}$, and satisfies $\inf_{\mathcal{H}} \phi_\lambda = \inf_{\mathcal{H}} \phi$, $\operatorname{argmin}_{\mathcal{H}} \phi_\lambda = \operatorname{argmin}_{\mathcal{H}} \phi$. One can consult [18, Section 17.2.1] and [38, 43, 53] for an in-depth study of the properties of the Moreau envelope in a Hilbert framework. In our context, since $\phi : \mathcal{H} \to \mathbb{R}$ is a damping potential, we can easily verify that ϕ_λ is still a damping potential. In particular, $\phi_\lambda(0) = \inf_{\mathcal{H}} \phi_\lambda = 0$. According to the subdifferential inequality for convex functions, this implies that, for all $u \in \mathcal{H}$,

$$\langle \nabla\phi_\lambda(u), u \rangle \geq 0. \tag{8}$$

We will also use the following inequality [53, Proposition 2.6]: for any $\lambda > 0$ and any $u \in \mathcal{H}$,

$$\|\nabla\phi_\lambda(u)\| \leq \|(\partial\phi)^0(u)\|. \tag{9}$$

So, for each $\lambda > 0$, we consider the approximate evolution equation

$$\ddot{x}_\lambda(t) + \nabla\phi_\lambda(\dot{x}_\lambda(t)) + \nabla f(x_\lambda(t)) = 0, \quad t \in [0, +\infty[. \tag{10}$$

We will first prove the existence and uniqueness of a global classical solution x_λ of (10) satisfying $x_\lambda(0) = x_0$ and $\dot{x}_\lambda(0) = x_1$. Then, we will prove that the filtered sequence (x_λ)

converges uniformly as $\lambda \to 0$ over bounded time intervals to a solution of (ADIGE-V). In view of the Hamiltonian formulation of (10), it is equivalent to consider the first-order (in time) system

$$\begin{cases} \dot{x}_\lambda(t) - u_\lambda(t) = 0, \\ \dot{u}_\lambda(t) + \nabla\phi_\lambda(u_\lambda(t)) + \nabla f(x_\lambda(t)) = 0, \end{cases} \tag{11}$$

with Cauchy data $x_\lambda(0) = x_0$, $u_\lambda(0) = x_1$. Set

$$Z_\lambda(t) = (x_\lambda(t), u_\lambda(t)) \in \mathcal{H} \times \mathcal{H}.$$

The system (11) can be written equivalently as

$$\dot{Z}_\lambda(t) + F_\lambda(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1),$$

where $F_\lambda : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$, $(x,u) \mapsto F_\lambda(x,u)$, is defined by

$$F_\lambda(x,u) = (0, \nabla\phi_\lambda(u)) + (-u, \nabla f(x)).$$

Hence F_λ splits as $F_\lambda(x,u) = \nabla\Phi_\lambda(x,u) + G(x,u)$, where

$$\Phi(x,u) = \phi(u), \quad \Phi_\lambda(x,u) = \phi_\lambda(u), \quad G(x,u) = (-u, \nabla f(x)).$$

Therefore, it is equivalent to consider the first-order differential inclusion with Cauchy data

$$\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1). \tag{12}$$

By the Lipschitz continuity of $\nabla\Phi_\lambda$, and the fact that G is Lipschitz continuous on bounded sets, the sum operator $\nabla\Phi_\lambda + G$ which governs (12) is Lipschitz continuous on bounded sets. As a consequence, the existence of a local solution to (12) follows from the classical Cauchy–Lipschitz theorem. To pass from a local solution to a global solution, we use a standard energy argument, and the following a priori estimate on solutions of (10). After taking the scalar product of (10) with $\dot{x}_\lambda$, and using (8), we find that the global energy

$$\mathcal{E}_\lambda(t) := f(x_\lambda(t)) - \inf_{\mathcal{H}} f + \tfrac{1}{2}\|\dot{x}_\lambda(t)\|^2 \tag{13}$$

is a decreasing function of t. In view of the Cauchy data, and since f is bounded below, this implies that, on any bounded time interval, the filtered sequences (x_λ) and $(\dot{x}_\lambda)$ are bounded. In view of the property (9) of Yosida approximation, and the property (iii) of the damping potential ϕ, this implies that

$$\|\nabla\phi_\lambda(x_\lambda(t))\| \leq \|(\partial\phi)^0(x_\lambda(t))\|$$

is also uniformly bounded for $\lambda > 0$ and t bounded. By the constitutive equation (10), this in turn implies that the filtered sequence $(\ddot{x}_\lambda)$ is also bounded. This implies that if a maximal solution is defined on a finite time interval $[0, T[$, then the limits of $x_\lambda(t)$ and $\dot{x}_\lambda(t)$ as $t \to T$ exist. Then, we can apply the local existence result, which gives a solution defined on a larger interval, thus contradicting the maximality of T.

To prove the uniform convergence of the filtered sequence (Z_λ) on bounded time intervals, we proceed in a similar way to [53, proof of Theorem 3.1] (see also Adly–Attouch [4] in the context of damped inertial dynamics). Take $T > 0$, and $\lambda, \mu > 0$. Consider the corresponding solutions of (12) on $[0, T]$,

$$\begin{aligned}
\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1),\\
\dot{Z}_\mu(t) + \nabla\Phi_\mu(Z_\mu(t)) + G(Z_\mu(t)) = 0, \quad Z_\mu(0) = (x_0, x_1).
\end{aligned}$$

Subtracting the two equations above, and taking the scalar product with $Z_\lambda(t) - Z_\mu(t)$, we obtain

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 + \langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \\
+ \langle G(Z_\lambda(t)) - G(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle = 0. \qquad (14)
\end{aligned}$$

We now use the following ingredients:

(a) By the properties of Yosida approximation [53, Theorem 3.1], we have

$$\begin{aligned}
&\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \\
&\qquad \geq -\frac{\lambda}{4}\|\nabla\Phi_\mu(Z_\mu(t))\|^2 - \frac{\mu}{4}\|\nabla\Phi_\lambda(Z_\lambda(t))\|^2.
\end{aligned}$$

Since the filtered sequences (x_λ) and $(\dot{x}_\lambda)$ are uniformly bounded on $[0, T]$, there exists a constant C_T such that, for all $0 \leq t \leq T$,

$$\|Z_\lambda(t)\| \leq C_T.$$

From (9), and the fact that ϕ is a damping potential (Definition 1 (ii)), we deduce that

$$\|\nabla\Phi_\lambda(Z_\lambda(t))\| \leq \sup_{\|\xi\|\leq C_T} \|(\partial\phi)^0(\xi)\| = M_T < +\infty.$$

Therefore

$$\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \geq -\tfrac{1}{4}M_T(\lambda + \mu).$$

(b) On account of the local Lipschitz assumption on ∇f, the mapping $G : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is Lipschitz continuous on bounded sets. Using again the fact that the sequence (Z_λ) is uniformly bounded on $[0, T]$, we deduce that there exists a constant L_T such that, for all $t \in [0, T]$ and all $\lambda, \mu > 0$,

$$\|G(Z_\lambda(t)) - G(Z_\mu(t))\| \leq L_T\|Z_\lambda(t) - Z_\mu(t)\|.$$

Combining the above results, and using the Cauchy–Schwarz inequality, we deduce from (14) that

$$\frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{1}{4}M_T(\lambda + \mu) + L_T\|Z_\lambda(t) - Z_\mu(t)\|^2.$$

We now integrate this differential inequality. Elementary calculus using $Z_\lambda(0) - Z_\mu(0) = 0$ gives

$$\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{M_T}{4L_T}(\lambda + \mu)(e^{2L_T t} - 1).$$

Therefore, (Z_λ) is a Cauchy sequence for uniform convergence on $[0, T]$, and hence it converges uniformly. This means that x_λ and $\dot{x}_\lambda$ converge uniformly to x and $\dot{x}$ respectively. To pass to the limit in (10), let us write it as

$$\nabla \phi_\lambda(\dot{x}_\lambda(t)) = \xi_\lambda(t) \tag{15}$$

where $\xi_\lambda(t) := -\ddot{x}_\lambda(t) - \nabla f(x_\lambda(t))$. We now rely on the variational convergence properties of Yosida approximation. Since (ϕ_λ) converges increasingly to ϕ as $\lambda \downarrow 0$, the sequence of integral functionals

$$\Psi^\lambda(\xi) := \int_0^T \phi_\lambda(\xi(t))\, dt$$

converges increasingly to $\Psi(\xi) = \int_0^T \phi(\xi(t))\, dt$. Therefore (Ψ^λ) Mosco-converges to Ψ in $L^2(0, T; \mathcal{H})$. According to the theorem which makes the link between the Mosco convergence of a sequence of convex lower semicontinuous functions and the graph convergence of their subdifferentials (see Attouch [15, Theorem 3.66]), we have

$$\partial \Psi^\lambda \to \partial \Psi$$

in the strong-$L^2(0, T; \mathcal{H}) \times$ weak-$L^2(0, T; \mathcal{H})$ topology. According to (15) we have

$$\xi_\lambda = \nabla \Psi^\lambda(\dot{x}_\lambda).$$

Since $\dot{x}_\lambda \to \dot{x}$ strongly in $L^2(0, T; \mathcal{H})$ and ξ_λ converges weakly in $L^2(0, T; \mathcal{H})$ to ξ given by

$$\xi(t) = -\ddot{x}(t) - \nabla f(x(t)), \tag{16}$$

we deduce that $\xi \in \partial \Psi(\dot{x})$, that is,

$$\xi(t) \in \partial \phi(\dot{x}(t)).$$

From (16), we finally conclude that x is a solution of (ADIGE-V).

The uniqueness of the solution of the Cauchy problem is obtained exactly in the same way as in the case of the global Lipschitz assumption. ■

Remark 1. The above existence and uniqueness result uses as essential ingredient the fact that the potential function f to be minimized is a differentiable function whose gradient is locally Lipschitz continuous. The introduction of constraints into f via the indicator function would lead to solutions involving shocks when reaching the boundary of the constraint. In this case, existence can still be obtained in finite dimensions, but uniqueness may not be satisfied (see Attouch–Cabot–Redont [23]).

4. Closed-loop velocity control: preliminary convergence results

Let $x : [0, +\infty[\to \mathcal{H}$ be a solution of (ADIGE-V).

4.1. Energy estimates

Define the global energy at time t as follows:

$$\mathcal{E}(t) := f(x(t)) - \inf_{\mathcal{H}} f + \tfrac{1}{2}\|\dot{x}(t)\|^2. \tag{17}$$

Take the scalar product of (ADIGE-V) with $\dot{x}(t)$. By the chain rule, we get

$$\frac{d}{dt}\mathcal{E}(t) + \langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle = 0. \tag{18}$$

The convex subdifferential inequality and $\phi(0) = 0$ gives, for all $u \in \mathcal{H}$,

$$\langle \partial\phi(u), u\rangle \geq \phi(u). \tag{19}$$

Combining the above two inequalites, we get

$$\frac{d}{dt}\mathcal{E}(t) + \phi(\dot{x}(t)) \leq 0. \tag{20}$$

Since ϕ is nonnegative, this implies that the global energy is nonincreasing. Since f is bounded below, this implies that the velocity $\dot{x}(t)$ is bounded over $[0, +\infty[$. Precisely,

$$\sup_{t\geq 0} \|\dot{x}(t)\| \leq R_1 := \sqrt{2\mathcal{E}(0)}. \tag{21}$$

To go further, suppose that the trajectory $x(\cdot)$ is bounded (this is so for example if f is coercive), and set

$$\sup_{t\geq 0} \|x(t)\| \leq R_2. \tag{22}$$

Let us now establish a bound on the acceleration. For this, we rely on the approximate dynamics

$$\ddot{x}_\lambda(t) + \nabla\phi_\lambda(\dot{x}_\lambda(t)) + \nabla f(x_\lambda(t)) = 0, \quad t \in [0, +\infty[. \tag{23}$$

A similar estimate as above gives $\sup_{t\geq 0} \|\dot{x}_\lambda(t)\| \leq R_1 := \sqrt{2\mathcal{E}(0)}$. According to property (iii) of the damping potential, we obtain

$$\|\nabla\phi_\lambda(\dot{x}_\lambda(t))\| \leq \sup_{\|u\|\leq R_1} \|(\partial\phi)^0(u)\| = M_1 < +\infty.$$

By the local Lipschitz continuity of ∇f,

$$\|\nabla f(x_\lambda(t))\| \leq \sup_{\|x\|\leq R_2} \|\nabla f(x)\| = M_2 < +\infty.$$

Combining the above two inequalities with (23), we see that for all $\lambda > 0$ and all $t \geq 0$,

$$\|\ddot{x}_\lambda(t)\| \leq M_1 + M_2.$$

Since $\ddot{x}_\lambda(t)$ converges weakly to $\ddot{x}(t)$ as $\lambda \to 0$ (see the proof of Theorem 3), we obtain

$$\sup_{t\geq 0} \|\ddot{x}(t)\| < +\infty. \tag{24}$$

Moreover, by integrating (20), we immediately obtain $\int_0^{+\infty} \phi(\dot{x}(t))\,dt < +\infty$. Let us summarize the above results, and complete them, in the following proposition.

Proposition 1. *Let* $x : [0,+\infty[\to \mathcal{H}$ *be a solution of* (ADIGE-V). *Then the global energy* $\mathcal{E}(t) = f(x(t)) - \inf_{\mathcal{H}} f + \frac{1}{2}\|\dot{x}(t)\|^2$ *is nonincreasing, and*

$$\sup_{t\geq 0} \|\dot{x}(t)\| < +\infty, \qquad \int_0^{+\infty} \phi(\dot{x}(t))\,dt < +\infty.$$

Suppose moreover that x *is bounded. Then*

$$\sup_{t\geq 0} \|\ddot{x}(t)\| < +\infty. \tag{25}$$

If additionally there exist $p \geq 1$ *and* $r > 0$ *such that* $\phi(u) \geq r\|u\|^p$ *for all* $u \in \mathcal{H}$*, then*

$$\lim_{t\to+\infty} \|\dot{x}(t)\| = 0. \tag{26}$$

Proof. We just need to prove the last point. From $\int_0^{+\infty} \phi(\dot{x}(t))\,dt < +\infty$ and $\phi(u) \geq r\|u\|^p$, we get $\int_0^{+\infty} \|\dot{x}(t)\|^p\,dt < +\infty$. This estimate, combined with $\sup_{t\geq 0} \|\ddot{x}(t)\| < +\infty$, classically implies that $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$. ∎

Let us complete the above result by examining the convergence of the acceleration to zero. To get this result, we need additional assumptions on the data f and ϕ.

Proposition 2. *Let* $x : [0,+\infty[\, \to \mathcal{H}$ *be a bounded solution of* (ADIGE-V). *Suppose that* f *is a* $\mathcal{C}^2$ *function, and* ϕ *is a* $\mathcal{C}^2$ *function which satisfies*

(i) (local strong convexity) *there exist constants* $\gamma > 0$ *and* $\rho > 0$ *such that for all* u *in* $\mathcal{H}$ *with* $\|u\| \leq \rho$,
$$\langle \nabla^2\phi(u)\xi, \xi\rangle \geq \gamma\|\xi\|^2 \quad \textit{for all } \xi \in \mathcal{H};$$

(ii) (global growth) *there exist* $p \geq 1$ *and* $r > 0$ *such that* $\phi(u) \geq r\|u\|^p$ *for all* $u \in \mathcal{H}$.

Then

$$\lim_{t\to+\infty} \|\ddot{x}(t)\| = 0. \tag{27}$$

Proof. Let us differentiate (ADIGE-V), and set $w(t) := \ddot{x}(t)$. We obtain

$$\dot{w}(t) + \nabla^2\phi(\dot{x}(t))w(t) = -\nabla^2 f(x(t))\dot{x}(t).$$

Take the scalar product of the above equation with $w(t)$. We get

$$\frac{1}{2}\frac{d}{dt}\|w(t)\|^2 + \langle \nabla^2\phi(\dot{x}(t))w(t), w(t)\rangle = -\langle \nabla^2 f(x(t))\dot{x}(t), w(t)\rangle.$$

By Proposition 1, we have $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$. From the local strong convexity assumption (i), and the Cauchy–Schwarz inequality, we deduce that for t sufficiently large, say $t \geq t_1$,

$$\frac{1}{2}\frac{d}{dt}\|w(t)\|^2 + \gamma\|w(t)\|^2 \leq \|\nabla^2 f(x(t))\dot{x}(t)\|\,\|w(t)\|.$$

Since $x(\cdot)$ is bounded and ∇f is Lipschitz continuous on bounded sets, we deduce that for some $C > 0$,

$$\frac{1}{2}\frac{d}{dt}\|w(t)\|^2 + \gamma\|w(t)\|^2 \leq C\|\dot{x}(t)\|\,\|w(t)\| \quad \text{for all } t \geq t_1.$$

After multiplication by $e^{2\gamma t}$, and integration from t_1 to t, we get

$$\tfrac{1}{2}(e^{\gamma t}\|w(t)\|)^2 \leq \tfrac{1}{2}(e^{\gamma t_1}\|w(t_1)\|)^2 + C\int_{t_1}^{t} e^{\gamma\tau}\|\dot{x}(\tau)\|(e^{\gamma\tau}\|w(\tau)\|)\,d\tau.$$

According to the Gronwall inequality (see [53, Lemma A.5]) we obtain

$$e^{\gamma t}\|w(t)\| \leq e^{\gamma t_1}\|w(t_1)\| + C\int_{t_1}^{t} e^{\gamma\tau}\|\dot{x}(\tau)\|\,d\tau.$$

Therefore

$$\|\ddot{x}(t)\| \leq \|\ddot{x}(t_1)\|e^{-\gamma(t-t_1)} + Ce^{-\gamma t}\int_{t_1}^{t} e^{\gamma\tau}\|\dot{x}(\tau)\|\,d\tau.$$

Since $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$, we have

$$\lim_{t\to+\infty} e^{-\gamma t}\int_{t_1}^{t} e^{\gamma\tau}\|\dot{x}(\tau)\|\,d\tau = 0.$$

Therefore, by passing to the limit in the above inequality we get $\lim_{t\to+\infty}\|\ddot{x}(t)\| = 0$. ■

Corollary 1. *Under the assumption of Proposition 2, suppose that the trajectory $x(\cdot)$ is relatively compact. Then for any sequence $x(t_n) \to x_\infty$ with $t_n \to +\infty$ we have $\nabla f(x_\infty) = 0$. Set $S = \{x \in \mathcal{H} : \nabla f(x) = 0\}$. Then*

$$\lim_{t\to+\infty} d(x(t), S) = 0.$$

Remark 2. (a) Without any geometric assumption on the function f, the trajectories of (ADIGE-V) may fail to converge. In [28] Attouch–Goudou–Redont exhibit a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is $\mathcal{C}^1$, coercive, whose gradient is Lipschitz continuous on bounded sets, and such that the (HBF) system admits an orbit $t \mapsto x(t)$ which does not converge as $t \to +\infty$. The above result shows that in such a situation, the attractor is the set $S = \{x \in \mathcal{H} : \nabla f(x) = 0\}$.

(b) It is necessary to assume that ϕ is a smooth function in order to get the conclusion of Corollary 1. In fact, in the case of dry friction, that is, $\phi(u) = r\|u\|$, there is convergence of the orbits to points satisfying $\|\nabla f(x_\infty)\| \leq r$, which are not in general critical points of f.

4.2. Model example

Consider the case $\phi(u) = \frac{r}{p}\|u\|^p$ with $p > 1$, in which case the dynamic (ADIGE-V) reads

$$\ddot{x}(t) + r\|\dot{x}(t)\|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0. \tag{28}$$

Therefore, for $p > 2$, the viscous damping coefficient $\gamma(\cdot)$ that enters equation (28), and which is equal to

$$\gamma(t) := r\|\dot{x}(t)\|^{p-2}, \tag{29}$$

tends to zero as $t \to +\infty$. So, we are in the setting of inertial dynamics with vanishing damping coefficient. Consequently, in the associated inertial gradient algorithms, the extrapolation coefficient tends to 1, and we can expect fast asymptotic convergence. To summarize, in the case of (28), and for f coercive, we have shown that, for $p > 2$,

$$\lim_{t\to+\infty} \gamma(t) = 0, \quad \gamma(\cdot) \in L^{\frac{p}{p-2}}(0, +\infty). \tag{30}$$

Is this information sufficient to derive the convergence rate of the values, and obtain similar convergence properties to those for the $(\mathrm{AVD})_\alpha$ system?

To give a first answer to this question, we rely on the results of Cabot–Engler–Gaddat [55], Attouch–Cabot [19] and Attouch–Chbani–Riahi [27] which concern the asymptotic stabilization of inertial gradient dynamics with general time-dependent viscosity coefficient $\gamma(t)$. In the case of a vanishing damping coefficient, the key property which ensures the asymptotic minimization property is that

$$\int_0^{+\infty} \gamma(t)\, dt = +\infty.$$

This means that the coefficient $\gamma(t)$ can go to zero as $t \to +\infty$, but not too fast in order to dissipate the energy enough. On the positive side, $\gamma(t) = \alpha/t$ does satisfy (30) for any $p > 0$, which does not exclude the Nesterov case. On the negative side, we can easily find $\gamma(t)$ such that

$$\lim_{t\to+\infty} \gamma(t) = 0, \quad \gamma(\cdot) \in L^{\frac{p}{p-2}}(0, +\infty) \quad \text{and} \quad \int_0^{+\infty} \gamma(t)\, dt < +\infty. \tag{31}$$

So, without any other hypothesis, we cannot rely on this information alone. At this point, the idea is to introduce additional information, assuming a geometric property of the function f being minimized. In the next two sections, we first consider the case where f is a strongly convex function, and then the case of the functions f satisfying the Kurdyka–Łojasiewicz property.

5. The strongly convex case: exponential convergence rate

We will study the asymptotic behavior of the system (ADIGE-V) when f is a strongly convex function. Recall that $f : \mathcal{H} \to \mathbb{R}$ is said to be μ-strongly convex (with $\mu > 0$) if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex. Then, we will consider the particular case where f is strongly convex and quadratic. Finally, we will give numerical illustrations in dimension 1.

5.1. General strongly convex function f

Theorem 4. *Let $f : \mathcal{H} \to \mathbb{R}$ be a differentiable function which is μ-strongly convex for some $\mu > 0$, and whose gradient is Lipschitz continuous on bounded sets. Let $\overline{x}$ be the unique minimizer of f. Let $\phi : \mathcal{H} \to \mathbb{R}_+$ be a damping potential (see Definition 1) which is differentiable, and whose gradient is Lipschitz continuous on bounded subsets of $\mathcal{H}$. Suppose that ϕ satisfies the following growth conditions:*

(i) (local) *there exist constants $\alpha > 0$ and $\rho > 0$ such that for all u in $\mathcal{H}$ with $\|u\| \le \rho$,*

$$\langle \nabla \phi(u), u\rangle \ge \alpha \|u\|^2;$$

(ii) (global) *there exist $p \ge 1$ and $c > 0$ such that $\phi(u) \ge c\|u\|^p$ for all u in $\mathcal{H}$.*

Then, for any solution $x : [0, +\infty[\to \mathcal{H}$ of (ADIGE-V)*, we have exponential convergence rate to zero as $t \to +\infty$ for $f(x(t)) - f(\overline{x})$, $\|x(t) - \overline{x}\|$ and the velocity $\|\dot{x}(t)\|$.*

Proof. We will use the following inequalities which are attached to the strong convexity of f:

$$f(\overline{x}) - f(x(t)) \ge \langle \nabla f(x(t)), \overline{x} - x(t)\rangle + \frac{\mu}{2}\|x(t) - \overline{x}\|^2, \tag{32}$$

$$f(x(t)) - f(\overline{x}) \ge \frac{\mu}{2}\|x(t) - \overline{x}\|^2. \tag{33}$$

Let us consider the global energy (introduced in (17), in the preliminary estimates)

$$\mathcal{E}(t) := \tfrac{1}{2}\|\dot{x}(t)\|^2 + f(x(t)) - f(\overline{x}).$$

By Proposition 1, $\dot{x}(t)$ is bounded on $\mathbb{R}_+$. Moreover, $\mathcal{E}(\cdot)$ is nonincreasing, and hence bounded from above. By definition of $\mathcal{E}(t)$, this implies that $f(x(t))$ is bounded from above. Since f is strongly convex, it is coercive, which implies that $x(\cdot)$ is bounded. Since $x(\cdot)$ and $\dot{x}(\cdot)$ are bounded, and the vector fields ∇f and $\nabla\phi$ are locally Lipschitz continuous, we deduce from the constitutive equation $\ddot{x}(t) = -\nabla\phi(\dot{x}(t)) - \nabla f(x(t))$ that $\ddot{x}(\cdot)$ is also bounded. According to the preliminary estimates established in Proposition 1, we have $\int_0^{+\infty} \phi(\dot{x}(t))\, dt < +\infty$. Combining this property with the global growth assumption (ii) on ϕ, we deduce that there exists $p \ge 1$ such that

$$\int_0^{+\infty} \|\dot{x}(t)\|^p \, dt < +\infty.$$

Since $\ddot{x}(\cdot)$ is bounded, this implies that $\dot{x}(t) \to 0$ as $t \to +\infty$. So, for t sufficiently large, say $t \ge t_1$,

$$\|\dot{x}(t)\| \le \rho.$$

The time derivative of $\mathcal{E}(\cdot)$, together with the constitutive equation (ADIGE-V), gives, for $t \ge t_1$,

$$\begin{aligned}\dot{\mathcal{E}}(t) &= \langle \dot{x}(t), -\nabla\phi(\dot{x}(t)) - \nabla f(x(t))\rangle + \langle \dot{x}(t), \nabla f(x(t))\rangle \\ &= -\langle \dot{x}(t), \nabla\phi(\dot{x}(t))\rangle \le -\alpha\|\dot{x}(t)\|^2,\end{aligned} \tag{34}$$

where the last inequality comes from the growth condition (i) on ϕ, and from $\|\dot{x}(t)\| \leq \rho$ for $t \geq t_1$.

Since $\dot{x}(\cdot)$ is bounded, let L be the Lipschitz constant of $\nabla\phi$ on a ball that contains the velocity vector $\dot{x}(t)$ for all $t \geq 0$. Since $\nabla\phi(0) = 0$ we have, for all $t \geq 0$,

$$\|\nabla\phi(\dot{x}(t))\| \leq L\|\dot{x}(t)\|. \tag{35}$$

Using successively (ADIGE-V), (35) and (32), we obtain

$$\begin{aligned}
\frac{d}{dt}(\langle x(t)-\overline{x}, \dot{x}(t)\rangle) &= \|\dot{x}(t)\|^2 + \langle x(t)-\overline{x}, -\nabla\phi(\dot{x}(t)) - \nabla f(x(t))\rangle \\
&\leq \|\dot{x}(t)\|^2 + L\|x(t)-\overline{x}\|\|\dot{x}(t)\| - \langle x(t)-\overline{x}, \nabla f(x(t))\rangle \\
&\leq \|\dot{x}(t)\|^2 + \frac{L^2}{2\mu}\|\dot{x}(t)\|^2 + \frac{\mu}{2}\|x(t)-\overline{x}\|^2 + \langle \overline{x}-x(t), \nabla f(x(t))\rangle \\
&\leq \left(1+\frac{L^2}{2\mu}\right)\|\dot{x}(t)\|^2 + f(\overline{x}) - f(x(t)).
\end{aligned} \tag{36}$$

Take now $\epsilon > 0$ (to be specified below), and define

$$h_\epsilon(t) := \mathcal{E}(t) + \epsilon\langle x(t)-\overline{x}, \dot{x}(t)\rangle.$$

The time derivative of h_ϵ, together with (34) and (36), gives, for $t \geq t_1$,

$$\dot{h}_\epsilon(t) \leq -\left(\alpha - \epsilon\left(1+\frac{L^2}{2\mu}\right)\right)\|\dot{x}(t)\|^2 - \epsilon(f(x(t)) - f(\overline{x})).$$

Choose $\epsilon > 0$ such that $\alpha - \epsilon(1+\frac{L^2}{2\mu}) > 0$. Set $C_1 := \min\{\alpha - \epsilon(1+\frac{L^2}{2\mu}), \epsilon\}$. We deduce that

$$\dot{h}_\epsilon(t) \leq -C_1\big(\|\dot{x}(t)\|^2 + f(x(t)) - f(\overline{x})\big). \tag{37}$$

Further, from (33) and the Cauchy–Schwarz inequality we easily obtain

$$\begin{aligned}
h_\epsilon(t) &\leq \frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t)) - f(\overline{x}) + \frac{\epsilon}{2}\|x(t)-\overline{x}\|^2 + \frac{\epsilon}{2}\|\dot{x}(t)\|^2 \\
&\leq \left(\frac{1}{2}+\frac{\epsilon}{2}\right)\|\dot{x}(t)\|^2 + \left(1+\frac{\epsilon}{\mu}\right)(f(x(t)) - f(\overline{x})) \\
&\leq \left(1+\epsilon\left(\frac{1}{2}+\frac{1}{\mu}\right)\right)\big(\|\dot{x}(t)\|^2 + f(x(t)) - f(\overline{x})\big).
\end{aligned}$$

Combining this inequality with (37), we obtain $\dot{h}_\epsilon(t) + C_2 h_\epsilon(t) \leq 0$ with $C_2 := \frac{C_1}{1+\epsilon(\frac{1}{2}+\frac{1}{\mu})} > 0$. Then, the Gronwall inequality classically implies

$$h_\epsilon(t) \leq h_\epsilon(0)e^{-C_2 t}. \tag{38}$$

Finally, from (33) and the Cauchy–Schwarz inequality we have

$$\begin{aligned}
h_\epsilon(t) &\geq \frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t)) - f(\overline{x}) - \frac{\epsilon}{2}\|x(t)-\overline{x}\|^2 - \frac{\epsilon}{2}\|\dot{x}(t)\|^2 \\
&\geq \left(\frac{1}{2}-\frac{\epsilon}{2}\right)\|\dot{x}(t)\|^2 + \left(1-\frac{\epsilon}{\mu}\right)(f(x(t)) - f(\overline{x})).
\end{aligned}$$

Therefore, by taking ϵ small enough, we obtain the existence of $C_3 > 0$ such that

$$h_\epsilon(t) \geq C_3\big(\|\dot{x}(t)\|^2 + f(x(t)) - f(\overline{x})\big).$$

Combining this inequality with (38) and (33), we obtain an exponential convergence rate to zero for $f(x(t)) - f(\overline{x})$, $\|x(t) - \overline{x}\|$ and the velocity $\|\dot{x}(t)\|$. ■

Remark 3. In Section 7.4, as a consequence of the Kurdyka–Łojasiewicz theory, we will extend the above results to the case where we only assume a quadratic growth assumption

$$f(x) - \inf_{\mathcal{H}} f \geq c\,\mathrm{dist}(x, \mathrm{argmin}\, f)^2.$$

Remark 4. In Section 6, we will give indications concerning the case of a general convex function f, whose solution set argmin f is nonempty. Let us recall that, in the case of (HBF), which corresponds to $\phi(u) = r\|u\|^2$, each trajectory converges weakly and its limit belongs to argmin f. Apart from this important case, convergence of trajectories depends both on the geometric properties of the function f to be minimized and on those of the damping potential ϕ. In Section 6 we will give an example in dimension 1, with trajectories which do not converge.

5.2. *Case of f convex quadratic positive definite*

Let us make precise the previous results in the case $f(x) = \frac{1}{2}\langle Ax, x\rangle$, where $A : \mathcal{H} \to \mathcal{H}$ is a linear continuous positive definite self-adjoint operator. Then $\nabla f(x) = Ax$, and (ADIGE-V) can be written as

$$\ddot{x}(t) + \partial\phi(\dot{x}(t)) + A(x(t)) \ni 0. \tag{39}$$

Let us prove the following ergodic convergence result, valid for a general damping potential ϕ.

Theorem 5. *Let $x : [0, +\infty[\to \mathcal{H}$ be a solution of (39), where ϕ is a damping potential, and $A : \mathcal{H} \to \mathcal{H}$ is a linear continuous positive definite self-adjoint operator. Then we have the following ergodic convergence result for the weak topology: as $t \to +\infty$,*

$$\frac{1}{t}\int_0^t x(\tau)\,d\tau \rightharpoonup x_\infty,$$

where the limit x_∞ satisfies $0 \in \partial\phi(0) + Ax_\infty$.

When ϕ is differentiable at the origin, we have $Ax_\infty = 0$, that is, $x_\infty = 0$.

When $\phi(x) = r\|x\|$, we have $\|Ax_\infty\| \leq r$.

Proof. The Hamiltonian formulation of (39) gives the equivalent first-order differential inclusion in the product space $\mathcal{H} \times \mathcal{H}$:

$$0 \in \dot{z}(t) + \partial\Phi(z(t)) + F(z(t)), \tag{40}$$

where $z(t) = (x(t), \dot{x}(t)) \in \mathcal{H} \times \mathcal{H}$, and

- $\Phi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is the convex continuous function defined by $\Phi(x, u) = \phi(u)$,
- $F : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is defined by $F(x, u) = (-u, Ax)$.

The trick is to renorm the product space $\mathcal{H} \times \mathcal{H}$ as follows. The mapping $(x, y) \mapsto \langle Ax, y\rangle$ defines a scalar product on $\mathcal{H}$, which is equivalent to the initial one. Accordingly, let us equip $\mathcal{H} \times \mathcal{H}$ with the scalar product

$$\langle\langle (x_1 u_1), (x_2, u_2)\rangle\rangle := \langle Ax_1, x_2\rangle + \langle u_1, u_2\rangle.$$

L et us observe that, with respect to this new scalar product:

- $F : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is a linear continuous skew-symmetric operator. Since A is self-adjoint,

$$\langle\langle F(x, u), (x, u)\rangle\rangle = \langle\langle (-u, Ax), (x, u)\rangle\rangle = -\langle Au, x\rangle + \langle Ax, u\rangle = 0.$$

- The subdifferential of Φ is unchanged, that is, $\partial\Phi(x, u) = (0, \partial\phi(u))$.

Therefore, the differential inclusion (40) is governed by the sum of two maximally monotone operators; one of them is the subdifferential of a convex continuous function, the other is a monotone skew-symmetric operator. By the classical Rockafellar theorem [43, Corollary 24.4], their sum is still maximally monotone. Consequently, we can apply the theory of semigroups generated by maximally monotone operators, and conclude that $z(t)$ converges weakly and in an ergodic way towards a zero $z_\infty = (x_\infty, u_\infty)$ of $\partial\Phi + F$. This means

$$(0, \partial\phi(u_\infty)) + (-u_\infty, Ax_\infty) = (0, 0).$$

Equivalently, $u_\infty = 0$ and $\partial\phi(0) + Ax_\infty \ni 0$. ■

In the case of the wave equation, this type of argument was developed by Haraux [68, Lecture 12, Theorem 45]. A recent account of these questions can be found in Haraux–Jendoubi [71] and Alabau-Boussouira–Privat–Trélat [6].

5.3. *Numerical illustrations*

Finding explicit solutions in closed form of nonlinear oscillators has direct applications in various fields. In the one-dimensional case, the corresponding second-order differential equation $\ddot{x}(t) + d(x(t), \dot{x}(t))\dot{x}(t) + g(x(t)) = 0$ is known as the Levinson–Smith equation. It reduces to the Liénard equation when d depends only on x. One can consult [63, 65] for recent reports on the subject and the description of some of the different techniques developed to resolve these questions. In our setting, we will provide some insight on this question by combining energetic and topological arguments.

5.3.1. A numerical one-dimensional example. Consider the case $\mathcal{H} = \mathbb{R}$, $f(x) = \frac{1}{2}|x|^2$, and $\phi(u) = \frac{1}{p}|u|^p$ with $p > 1$. Then (ADIGE-V) reads

$$\ddot{x}(t) + |\dot{x}(t)|^{p-2}\dot{x}(t) + x(t) = 0. \tag{41}$$

It is a linear oscillator with nonlinear damping. According to the previous facts, we have the following results.

• For $p = 2$, according to the strong convexity of the potential function $f(x) = \frac{1}{2}|x|^2$ and Theorem 4, we have convergence at an exponential rate of $x(t)$ and $\dot{x}(t)$ toward 0. Indeed, $p = 2$ is the only value of p for which the hypotheses of Theorem 4 are satisfied. For $p > 2$ the local hypothesis (i) is not satisfied, and for $p < 2$ the gradient of ϕ fails to be Lipschitz continuous on bounded sets containing the origin.

• For $p > 1$, let us first show that $\lim_{t\to+\infty} \dot{x}(t) = 0$. This results from Proposition 1, and the fact that the trajectory is bounded. This last property results from the fact that the global energy $\mathcal{E}(t) = \frac{1}{2}|\dot{x}(t)|^2 + \frac{1}{2}|x(t)|^2$ is nonincreasing, and hence convergent (and bounded from above).

Let us show that $x(t)$ tends to zero. Since $\lim_{t\to+\infty} \dot{x}(t) = 0$, and $\lim_{t\to+\infty} \mathcal{E}(t)$ exists, we see that

$$\lim_{t\to+\infty} |x(t)|^2 = \lim_{t\to+\infty} \mathcal{E}(t) \quad \text{exists.} \tag{42}$$

Since the identity operator clearly satisfies the assumptions of Theorem 5, we have the ergodic convergence $\lim_{t\to+\infty} \frac{1}{t}\int_0^t x(\tau)\, d\tau = 0$. There are two possibilities:

(a) For t sufficiently large, $x(t)$ has a fixed sign. By (42), $\lim_{t\to+\infty} x(t) := x_\infty$ exists. This convergence implies ergodic convergence: $\lim_{t\to+\infty} \frac{1}{t}\int_0^t x(\tau)\, d\tau = x_\infty$. But we know that the ergodic limit is zero, hence $x_\infty = 0$.

(b) The trajectory changes sign an infinite number of times as $t \to +\infty$. This means that there exist sequences s_n and t_n which tend to infinity such that $x(t_n)x(s_n) < 0$. Since the trajectory is continuous, by the mean value theorem there exists $\tau_n \in [s_n, t_n]$ such that $x(\tau_n) = 0$. Hence $x(\tau_n)^2 = 0$ for all $n \in \mathbb{N}$, with $\tau_n \to +\infty$. Since $\lim_{t\to+\infty} |x(t)|^2$ exists, this implies that $\lim_{t\to+\infty} |x(t)|^2 = 0$. Clearly, this implies that $\lim_{t\to+\infty} x(t) = 0$.

So, for $p > 1$, for any solution trajectory of (41), we have

$$\lim_{t\to+\infty} x(t) = 0 \quad \text{and} \quad \lim_{t\to+\infty} \dot{x}(t) = 0. \tag{43}$$

Now let us analyze how the trajectories and their speeds go to zero. As we shall see, the case $p > 2$ corresponds to weak damping, while $p < 2$ corresponds to strong damping.

Case $p > 2$. Since the speed $|\dot{x}(t)|$ tends to zero, we have $\gamma(t) := |\dot{x}(t)|^{p-2} \to 0$ as $t \to +\infty$. The viscous damping coefficient $\gamma(t)$ becomes asymptotically small. Consequently, the damping effect also becomes weak (that is what we call weak damping). As p increases, the damping effect tends to decrease, the trajectory tends to oscillate more and more, and the rate of convergence deteriorates.

This is illustrated in Figure 3, where we can see the evolution of the trajectory $x(t)$ (blue line) and its derivative $\dot{x}(t)$ (red line) of the dynamical system (41) with starting point $(x(0), \dot{x}(0)) = (3, 1)$ and for different values of $p \geq 2$. The trajectory and the velocity tend to zero, but the oscillations become stronger as p increases, and the convergence

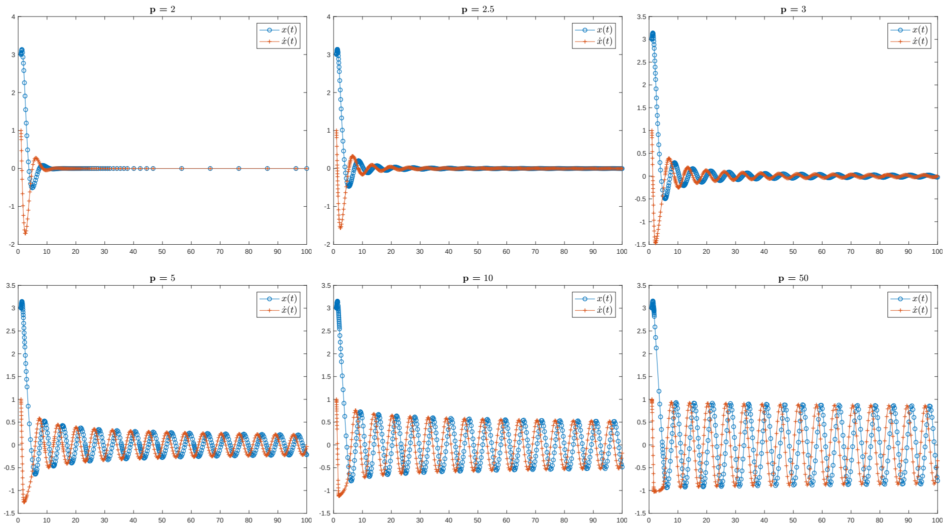

Fig. 3. The evolution of the trajectories $x(t)$ (blue line) and $\dot{x}(t)$ (red line) of the dynamical system (41) for different values of $p \geq 2$.

to zero becomes very slow. For large p, the oscillatory aspect conforms to the ergodic convergence of the trajectory to 0 (indeed, in dimension 1 the trajectory tends to zero, but we can expect that in higher dimensions there is only ergodic convergence to zero). Note also that for $p > 2$, and p close to 2, the trajectory is close to that corresponding to $p = 2$, and therefore enjoys excellent convergence properties. It would be interesting to study this situation, because it is a natural candidate to obtain convergence rates similar to the accelerated gradient method of Nesterov.

Case $1 < p < 2$. According to (43) we have $\lim_{t\to+\infty} \dot{x}(t) = 0$ and $\lim_{t\to+\infty} x(t) = 0$. Since $\lim_{t\to+\infty} \dot{x}(t) = 0$ and $2 - p > 0$, the viscous damping coefficient satisfies

$$\gamma(t) := \frac{1}{|\dot{x}(t)|^{2-p}} \to +\infty \quad \text{as } t \to +\infty.$$

We are in the setting of a strong damping effect. This situation was analyzed in the following result of [21], concerning the asymptotic behavior of

$$(\mathrm{IGS})_\gamma \qquad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla f(x(t)) = 0.$$

Proposition 3. *Let $f : \mathcal{H} \to \mathbb{R}$ be a function of class $\mathcal{C}^1$ such that ∇f is bounded on bounded subsets of $\mathcal{H}$. Given $r > 0$ and $\theta > 1$, assume that $\gamma(t) = rt^\theta$ for every $t \geq t_0 \geq 0$. Then each bounded solution $x(\cdot)$ of* $(\mathrm{IGS})_\gamma$ *satisfies $\int_{t_0}^{+\infty} \|\dot{x}(t)\|\, dt < +\infty$, and hence converges strongly to some $x^* \in \mathcal{H}$.*

From this result, we can obtain some information about the convergence rate of the velocity to zero. We have two cases: either $\int_{t_0}^{+\infty} \|\dot{x}(t)\|\, dt < +\infty$, or $\int_{t_0}^{+\infty} \|\dot{x}(t)\|\, dt$

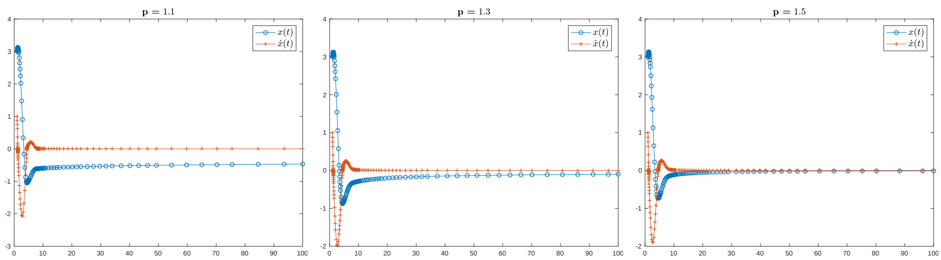

Fig. 4. Evolution of $x(t)$ (blue) and $\dot{x}(t)$ (red) for different values of $1 < p < 2$.

$= +\infty$. In this last case, according to Proposition 3, we cannot have $\gamma(t) = 1/|\dot{x}(t)|^{2-p}$ of order rt^{θ} with $\theta > 1$. This excludes the possibility to have $|\dot{x}(t)|$ of order $1/t^{\theta/(2-p)}$ with $\theta > 1$. So, the best that we can expect is $|\dot{x}(t)| \sim 1/t^{1/(2-p)}$ as $t \to +\infty$. This estimate is in accordance with the exponential decay when $p = 2$, and the finite length property when $p = 1$. We emphasize that the above argument is not a rigorous proof; it just gives an indication of the type of convergence rate we can expect.

In Figure 4 we can see the evolution of the trajectory $t \mapsto x(t)$ (blue line) and of its derivative $t \mapsto \dot{x}(t)$ (red line) of the dynamical system (41) with starting point $(x(0), \dot{x}(0)) = (3, 1)$ for different values of $1 < p < 2$. Because of strong damping, the trajectories exhibit small oscillations, and the velocity converges fast to zero. By contrast, convergence of $x(t)$ to zero highly depends on the parameter p. When p is close to 1, the convergence of the trajectory to zero is poor, but already a slight increase of p improves the convergence.

6. Weak damping: from slow convergence to attractor effect

As already mentioned, even in the case of a strongly convex function f, when the damping effect becomes too weak, the convergence property deteriorates. In the case of the damping $\|\dot{x}(t)\|^{p-2}\dot{x}(t)$, this corresponds to situations where $p > 2$. In this section, we give examples showing that in the case of a general convex function, the situation is even worse, and the trajectory may not converge in the case of weak damping. In this case, one has to replace the convergence notions by the concept of attractor, a central object of the theory of dynamical systems and PDEs; see Hale [67], Haraux [69] for seminal contributions to the subject in the case of gradient systems (i.e. systems for which there exists a Lyapunov function). For optimization purposes, this is a promising research topic, still largely to be explored in the case of a general damping function. In the next section, we take a convex function f with a continuum of minimizers, and examine the lack of convergence when damping becomes too weak. In fact, as already underlined, convergence depends both on the geometric properties of the damping potential and on the potential function f to be minimized. The corresponding geometric aspects concerning f will be examined later.

6.1. An example where convergence fails to hold

The following example is based on Haraux [69, Section 5.1]. Take $\mathcal{H} = \mathbb{R}$, and $f : \mathbb{R} \to \mathbb{R}$ a convex function of class $\mathcal{C}^1$ which achieves its minimal value on the line segment $[a, b]$ with $a < b$. We suppose that f is coercive, i.e. $\lim_{|x|\to+\infty} f(x) = +\infty$. Its graph looks like a bowl with a flat bottom.

Consider the evolution equation with closed-loop damping

$$\ddot{x}(t) + |\dot{x}(t)|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0. \tag{44}$$

Let us discuss, depending to the value of p, the convergence properties of the trajectories of this system. We will need the following elementary lemma; we recall the proof for completeness.

Lemma 1 ([69, Lemma 5.1.3]). *Let $v \in \mathcal{C}^2(\mathbb{R}_+)$ satisfy, for some $c > 0$,*

$$\dot{v}(0) > 0, \quad \ddot{v}(t) \geq -c\dot{v}(t)^2 \quad \textit{for all } t \geq 0.$$

Then v is increasing, and $\lim_{t\to+\infty} v(t) = +\infty$.

Proof. As long as $\dot{v}(t) > 0$, by integration of the differential inequality $\ddot{v}(t) + c\dot{v}(t)^2 \geq 0$ we obtain

$$\dot{v}(t) \geq \frac{1}{ct + \frac{1}{\dot{v}(0)}}.$$

This immediately implies that $\dot{v}(t) > 0$ for all $t \geq 0$. By integrating, we obtain

$$v(t) \geq v(0) + \int_0^t \frac{1}{c\tau + \frac{1}{\dot{v}(0)}}\, d\tau,$$

which implies $\lim_{t\to+\infty} v(t) = +\infty$. ■

Proposition 4. *Suppose that $p \geq 3$. Then any trajectory of (44) which is not constant passes infinitely many times through the points a and b.*

Proof. According to Proposition 1, the trajectory $x(\cdot)$ is bounded and satisfies

$$\lim_{t\to+\infty} \|\dot{x}(t)\| = 0 \quad \text{and} \quad \sup_{t\geq 0} \|\ddot{x}(t)\| < +\infty. \tag{45}$$

Let us argue by contradiction, and assume that there exists some $t_1 > 0$ such that $x(t) \geq a$ for all $t \geq t_1$. We can distinguish two cases:

• **First case:** $\dot{x}(t) \geq 0$ for all $t \geq t_1$. Then $t \mapsto x(t)$ is increasing and bounded, hence converges to some $x_\infty \in \mathbb{R}$. From the constitutive equation (44), $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$, and the continuity of ∇f, we deduce that $\lim_{t\to+\infty} \ddot{x}(t) = -\nabla f(x_\infty)$. Using again $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$, we deduce that $\nabla f(x_\infty) = 0$. Since $x \mapsto \nabla f(x)$ is an increasing function, and $\nabla f(x(t_1)) \geq 0$, we obtain

$$\nabla f(x(t)) = 0 \quad \text{for all } t \geq t_1.$$

Returning to the constitutive equation (44), we get

$$\ddot{x}(t) + |\dot{x}(t)|^{p-2}\dot{x}(t) = 0 \quad \text{for all } t \geq t_1.$$

Since $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$, and $p \geq 3$, for t sufficiently large, say $t \geq t_2 \geq t_1$, we have $|\dot{x}(t)|^{p-1} \leq |\dot{x}(t)|^2$. Therefore, for all $t \geq t_2$,

$$\ddot{x}(t) + |\dot{x}(t)|^2 \geq 0.$$

Since $x(\cdot)$ is not constant, there exists some $t_3 \geq t_2$ such that $\dot{x}(t_3) > 0$. According to Lemma 1, we have $\lim_{t\to+\infty} x(t) = +\infty$, contradicting the convergence of $x(t)$.

• **Second case:** there exists $t_2 \geq t_1$ such that $\dot{x}(t_2) < 0$. From the constitutive equation (44) and $x(t) \geq a$ we get, for all $t \geq t_2$,

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

This implies, for t large enough,

$$\ddot{x}(t) \leq |\dot{x}(t)|^2.$$

Let us apply Lemma 1 to $-x(\cdot)$. Since $-\dot{x}(t_2) > 0$, we obtain $\lim_{t\to+\infty} x(t) = -\infty$, a contradiction.

A similar argument gives the same kind of result for b, namely, for every $t_1 > 0$ there exists $t > t_1$ such that $x(t) > b$. Therefore, an infinite number of times the trajectory takes the values a and b, so it oscillates indefinitely between a and b. ■

By contrast, if the damping effect is sufficiently strong, there is convergence. In our situation, this corresponds to the case $2 \leq p < 3$, as shown in the following proposition.

Proposition 5. *Suppose that $2 \leq p < 3$. Then any trajectory of (44) converges, and its limit belongs to $[a,b]$.*

Proof. When $p = 2$ the convergence follows from Alvarez's theorem for (HBF). So suppose $2 < p < 3$. We sketch the main lines of the proof; the details can be found in Haraux–Jendoubi [71, Theorem 9.2.1], which deals with a slightly more general situation. Let x be a solution of (44), and denote by $\omega(x)$ its limit set, that is, the set of its limit points as $t \to +\infty$ (limits of sequences $x(t_n)$ for $t_n \to +\infty$). By a classical argument, this set is a connected subset of $\{\nabla f = 0\}$, that is, $\omega(x) \subset [a,b]$. If $\omega(x)$ is reduced to a singleton, the proof is finished. Let us therefore examine the complementary case

$$\omega(x) = [c,d] \subset [a,b] \quad \text{with } c < d,$$

and show that this leads to a contradiction. Set $l := \frac{1}{2}(c+d)$. Let us prove that $\lim_{t\to+\infty} x(t) = l$, which gives $\omega(x) = \{l\}$, contrary to $\omega(x) = [c,d]$, $c \neq d$. First, since l is in the interior of $\omega(x)$, by the intermediate value property there exists a sequence (t_n)

with $t_n \to +\infty$ such that $x(t_n) = l$. By continuity of x, and since l is in the interior of $[c, d]$, for each $n \in \mathbb{N}$ there exists $\delta_n > 0$ such that

$$x(t) \in [c, d] \quad \text{for all } t \in [t_n, t_n + \delta_n].$$

Let us prove that for n large enough we can take $\delta_n = +\infty$. Set

$$\theta_n = \inf\{t > t_n : x(t) \notin [c, d]\},$$

and assume $\theta_n < +\infty$. So for all $t \in [t_n, \theta_n]$ we have $\nabla f(x(t)) = 0$, and (44) reduces to

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

After multiplying by $\dot{x}(t)$ we get, for all $t \in [t_n, \theta_n]$,

$$\frac{d}{dt}|\dot{x}(t)|^2 + 2|\dot{x}(t)|^p = 0.$$

After integration from t_n to $t \in [t_n, \theta_n]$ we get, for all $t \in [t_n, \theta_n]$,

$$|\dot{x}(t)| = \left(|\dot{x}(t_n)|^{-p+2} + (p-2)(t - t_n)\right)^{\frac{-1}{p-2}}.$$

After further integration we get, for all $t \in [t_n, \theta_n]$,

$$\begin{aligned}
|x(t) - l| = |x(t) - x(t_n)| &\leq \int_{t_n}^{t} |\dot{x}(s)|\, ds \\
&= \frac{1}{p-3}\left(|\dot{x}(t_n)|^{-p+2} + (p-2)(t - t_n)\right)^{\frac{p-3}{p-2}} + \frac{1}{3-p}|\dot{x}(t_n)|^{3-p} \\
&\leq \frac{1}{3-p}|\dot{x}(t_n)|^{3-p},
\end{aligned}$$

where, for the last inequality, we use the hypothesis $2 < p < 3$. Since $\dot{x}(t_n)$ converges to zero as $t_n \to +\infty$, for n large enough we have $\theta_n = +\infty$. This means that for n large enough,

$$x(t) \in [c, d] \text{ and } |x(t) - l| \leq \frac{1}{3-p}|\dot{x}(t_n)|^{3-p} \quad \forall t \in [t_n, +\infty),$$

which implies that $x(t)$ converges to l as $t \to +\infty$. ∎

Figure 5 illustrates the attractor effect when damping becomes too weak. Take $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{2}(x+1)^2$ for $x \leq -1$, $f(x) = 0$ for $|x| < 1$, and $f(x) = \frac{1}{2}(x-1)^2$ for $x \geq 1$.

For $p = 3$ there is a radical change in the trajectory behavior. For $p \geq 3$, they do not converge asymptotically, and exhibit very oscillating behavior by steadily passing through the points -1 and $+1$. For $2 < p < 3$, there is numerical evidence that the trajectories converge, which confirms the conclusion of Proposition 5.

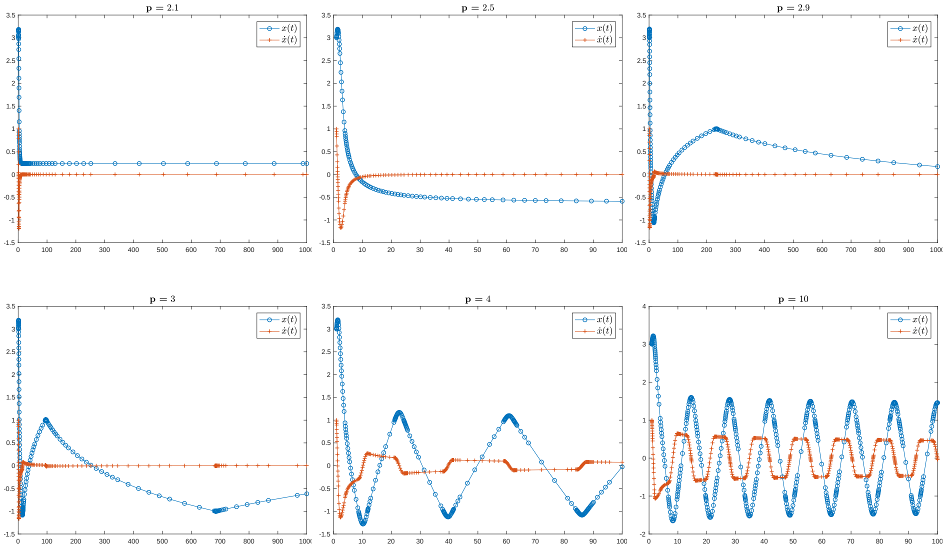

Fig. 5. Evolution of the trajectories $x(t)$ (blue) and $\dot{x}(t)$ (red) of the dynamical system (41) for $p \geq 2$.

6.2. *An explicit one-dimensional example*

Take $\mathcal{H} = \mathbb{R}$ and $f(x) = c|x|^\gamma$, where c and γ are positive parameters. Let us look for solutions of

$$\ddot{x}(t) + r|\dot{x}(t)|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0, \tag{46}$$

when $p > 1$. More precisely, we look for nonnegative solutions of the form $x(t) = 1/t^\theta$ with $\theta > 0$. This means that the trajectory is not oscillating, it is completely damped. We proceed by identification, and determine the values of the parameters c, γ, r, p and θ which provide such solutions. On the one hand,

$$\ddot{x}(t) + r|\dot{x}(t)|^{p-2}\dot{x}(t) = \frac{\theta(\theta+1)}{t^{\theta+2}} - \frac{\theta^{p-1}r}{t^{(\theta+1)(p-1)}}.$$

On the other hand, $\nabla f(x) = c\gamma|x|^{\gamma-2}x$, which gives

$$\nabla f(x(t)) = \frac{c\gamma}{t^{\theta(\gamma-1)}}.$$

Thus, $x(t) = 1/t^\theta$ is a solution of (46) if and only if

$$\frac{\theta(\theta+1)}{t^{\theta+2}} - \frac{\theta^{p-1}r}{t^{(\theta+1)(p-1)}} + \frac{c\gamma}{t^{\theta(\gamma-1)}} = 0.$$

This is equivalent to solving the following system:

(i) $\theta + 2 = \theta(\gamma - 1)$;

(ii) $\theta + 2 = (\theta + 1)(p - 1)$;

(iii) $c\gamma = \theta^{p-1}r - \theta(\theta+1)$;

(iv) $\theta > 0$, $c > 0$.

Solving (i) and (ii) for γ gives $2 < p < 3$, $\gamma > 2$, and the following values:

$$\theta = \frac{2}{\gamma-2}, \quad p = \frac{3\gamma-2}{\gamma}.$$

Condition (iii) gives $c = \frac{\theta}{\gamma}(\theta^{p-2}r - (\theta+1))$ and the positivity condition (iv) gives $r > \frac{\theta+1}{\theta^{p-2}}$. We have $\min f = 0$ and

$$f(x(t)) = c\frac{1}{t^{2\gamma/(\gamma-2)}} = \frac{c}{t^{2/(p-2)}}.$$

To summarize, we have shown that when taking $2 < p < 3$ and $f(x) = c|x|^{\frac{2}{3-p}}$ there exists a solution of

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

of the form $x(t) = 1/t^{3-p/(p-2)}$, for which

$$f(x(t)) - \min f = \frac{c}{t^{2/(p-2)}}.$$

As expected, the speed of convergence of $f(x(t))$ to 0 depends on p. Therefore, without other geometric assumptions on f, for $2 < p < 3$ we cannot expect a convergence rate better than $\mathcal{O}(1/t^{2/(p-2)})$. When $p \to 3$ from below, the function $f(x) = c|x|^{2/(3-p)}$ becomes very flat around its minimum (the origin) and the convergence of $x(t) = 1/t^{(3-p)/(p-2)}$ to the origin becomes very slow.

7. Damping via closed-loop velocity control, quasi-gradient and (KŁ)

In this section, $\mathcal{H} = \mathbb{R}^N$ is a finite-dimensional Euclidean space. This will allow us to use the Kurdyka–Łojasiewicz property, briefly (KŁ). Unless otherwise indicated, no convexity assumption is made on the function f being minimized, which will be assumed to satisfy (KŁ). To obtain the convergence of orbits, the need for a geometric assumption on the function f to be minimized has long been recognized. As for the steepest descent, without additional geometric assumptions on the potential function f, the bounded orbits of the heavy ball with friction dynamic (HBF) may not converge. Let us recall the result from [28] where a function $f : \mathbb{R}^2 \to \mathbb{R}$ is shown which is $\mathcal{C}^1$, coercive, its gradient is Lipschitz continuous on bounded sets, and the (HBF) system admits an orbit $t \mapsto x(t)$ which does not converge as $t \to +\infty$. This example is an inertial version of the famous Palis–De Melo counterexample for the continuous steepest descent [84]. In this section, we examine an important situation where the convergence property is satisfied, namely when f is assumed to satisfy (KŁ), a geometric notion which is presented below.

7.1. *Some basic facts concerning (KŁ)*

A function $G : \mathbb{R}^N \to \mathbb{R}$ satisfies the (KŁ) property if its values can be reparametrized in the neighborhood of each of its critical points so that the resulting function becomes sharp in the sense that there exists a continuous, concave, increasing function θ such that for all u in a slice of G,

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

The function θ captures the geometry of G around its critical point, it is called a *desingularizing function*; see [16,17,45] for further results. Tame functions satisfy property (KŁ). Tameness refers to an ubiquitous geometric property of functions and sets encountered in most finite-dimensional optimization problems. Sets or functions are called tame when they can be described by a finite number of basic formulas/inequalities/Boolean operations involving standard functions such as polynomial, exponential, or max functions. Classical examples of tame objects are piecewise linear objects (with finitely many pieces) and semi-algebraic objects. The general notion covering these situations is the concept of o-minimal structure; see van den Dries [91]. Tameness models nonsmoothness via the so-called stratification property of tame sets/functions. It was this property that motivated the vocable of tame topology, "la topologie modérée" according to Grothendieck. All these aspects have been well documented in a series of recent papers devoted to nonconvex nonsmooth optimization; see Ioffe [74], Castera–Bolte–Févotte–Pauwels [58] for an application to deep learning, and references therein. We refer to [16] for illustrations, and examples within a general optimization setting.

This property is particularly interesting in our context, because we work with an *autonomous dynamical system*, in which case (KŁ) theory applies to quasi-gradient systems. This contrasts with the accelerated gradient method of Nesterov which is based on a nonautonomous dynamical system, and for which we have no convergence theory based on the (KŁ) property. Under this property, we will obtain convergence results with convergence rates linked to the geometry of the data functions f and ϕ, via the desingularizing function.

7.2. *Quasi-gradient systems*

Let us first recall the main lines of the quasi-gradient approach to inertial gradient systems as developed by Bégout–Bolte–Jendoubi [44]. The geometric interpretation is simple: a vector field F is called quasi-gradient for a function E if it has the same singular point as E and if the angle between the field F and the gradient ∇E remains acute and bounded away from $\pi/2$. A precise definition is given below. Of course, such systems behave very similarly to gradient systems. We refer to Bárta–Chill–Fašangová [41, 42, 61], Chergui [60], Huang [73] and the references therein for further geometrical insights into this topic.

Definition 3. Let Γ be a nonempty closed subset of $\mathbb{R}^N$, and let $F : \mathbb{R}^N \to \mathbb{R}^N$ be a locally Lipschitz continuous mapping. We say that the first-order system

$$\dot{z}(t) + F(z(t)) = 0 \tag{47}$$

has a *quasi-gradient structure* for E on Γ if there exist a differentiable function $E : \mathbb{R}^N \to \mathbb{R}$ and $\alpha > 0$ such that the following two conditions are satisfied:

(angle condition) $\quad \langle \nabla E(z), F(z)\rangle \geq \alpha \|\nabla E(z)\| \, \|F(z)\| \quad$ for all $z \in \Gamma$;
(rest point equivalence) $\quad \operatorname{crit} E \cap \Gamma = F^{-1}(0) \cap \Gamma$.

Based on this notion, we have the following convergence properties of bounded trajectories of (47). The following result is a localized version and straight adaptation of [44, Theorem 3.2].

Theorem 6. *Let $F : \mathbb{R}^N \to \mathbb{R}^N$ be a locally Lipschitz continuous mapping. Let $z : [0, +\infty[\to \mathbb{R}^N$ be a bounded solution of (47). Take $R \geq \sup_{t\geq 0} \|z(t)\|$. Assume that F defines a quasi-gradient vector field for E_R on $\bar{B}(0, R)$, where $E_R : \mathbb{R}^N \to \mathbb{R}$ is a differentiable function. Assume further that the function E_R is* (KŁ)*. Then*

(i) $z(t) \to z_\infty$ *as* $t \to +\infty$, *where* $z_\infty \in F^{-1}(0)$*;*

(ii) $\dot{z} \in L^1(0, +\infty; \mathbb{R}^N)$ *and* $\dot{z}(t) \to 0$ *as* $t \to +\infty$*;*

(iii) $\|z(t) - z_\infty\| \leq \frac{1}{\alpha_R}\theta(E_R(z(t)) - E_R(z_\infty))$,

where θ is the desingularizing function for E_R at z_∞, and α_R enters the angle condition of Definition 3.

7.3. Convergence of systems with closed-loop velocity control under (KŁ)

Let us apply the above approach to the inertial system with closed-loop damping

$$\ddot{x}(t) + \nabla\phi(\dot{x}(t)) + \nabla f(x(t)) = 0, \tag{48}$$

by writing it as a first-order system, via its Hamiltonian formulation. We will assume that $\nabla\phi$ is locally Lipschitz continuous. Indeed, we can reduce to this situation by using a regularization procedure based on the Moreau envelope.

Theorem 7. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a $\mathcal{C}^2$ function whose gradient is Lipschitz continuous on bounded sets, and such that $\inf_{\mathbb{R}^N} f > -\infty$. Let $E_\lambda : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be defined by*

$$E_\lambda(x, u) := \tfrac{1}{2}\|u\|^2 + f(x) + \lambda\langle \nabla f(x), u\rangle \quad \text{for all } (x, u) \in \mathbb{R}^N \times \mathbb{R}^N.$$

Suppose that the function E_λ satisfies the (KŁ) *property. Let $\phi : \mathbb{R}^N \to \mathbb{R}_+$ be a damping potential* (*see Definition 1*) *which is differentiable and satisfies the following growth conditions:*

(i) (local) *there exist constants $\gamma, \delta, \epsilon > 0$ such that, for all u in $\mathbb{R}^N$ with $\|u\| \leq \epsilon$,*

$$\phi(u) \geq \gamma\|u\|^2 \quad \textit{and} \quad \|\nabla\phi(u)\| \leq \delta\|u\|;$$

(ii) (global) *there exist $p \geq 1$ and $c > 0$ such that $\phi(u) \geq c\|u\|^p$ for all u in $\mathbb{R}^N$.*

Let $x : [0, +\infty[\to \mathbb{R}^N$ *be a bounded solution of*

$$\ddot{x}(t) + \nabla\phi(\dot{x}(t)) + \nabla f(x(t)) = 0.$$

Then

(i) $x(t) \to x_\infty$ *as* $t \to +\infty$*, where* $x_\infty \in \mathrm{crit}\, f$*;*
(ii) $\dot{x} \in L^1(0, +\infty; \mathbb{R}^N)$ *and* $\dot{x}(t) \to 0$ *as* $t \to +\infty$*;*
(iii) *for* λ *sufficiently small, and* t *sufficiently large,*

$$\|x(t) - x_\infty\| \leq \frac{1}{\alpha}\theta\big(E_\lambda(x(t), u(t)) - E_\lambda(x_\infty, 0)\big),$$

where θ *is the desingularizing function for* E_λ *at* $(x_\infty, 0)$*, and* α *enters the corresponding angle condition.*

Proof. From the preliminary estimates established in Proposition 1, we have

$$\int_0^{+\infty} \phi(\dot{x}(t))\, dt < +\infty \quad \text{and} \quad \sup_{t\geq 0} \|\dot{x}(t)\| < +\infty.$$

Combining the first property above with the global growth assumption on ϕ, we deduce that there exists $p \geq 1$ such that

$$\int_0^{+\infty} \|\dot{x}(t)\|^p\, dt < +\infty.$$

From the constitutive equation, we have

$$\ddot{x}(t) = -\nabla\phi(\dot{x}(t)) - \nabla f(x(t)).$$

Since $x(\cdot)$ and $\dot{x}(\cdot)$ are bounded, and ∇f is locally Lipschitz continuous, we deduce that $\ddot{x}(\cdot)$ is also bounded. Classically, these properties imply that $\dot{x}(t) \to 0$ as $t \to +\infty$. Take $R \geq \sup_{t\geq 0} \|x(t)\|$. Then, for t sufficiently large, the trajectory $t \mapsto (x(t), \dot{x}(t))$ in the phase space $\mathbb{R}^N \times \mathbb{R}^N$ lies in the closed set $\Gamma = \bar{B}(0, R) \times \bar{B}(0, \epsilon)$.

The Hamiltonian formulation of (48) gives the first-order differential system

$$\dot{z}(t) + F(z(t)) = 0, \tag{49}$$

where $z(t) = (x(t), \dot{x}(t)) \in \mathbb{R}^N \times \mathbb{R}^N$, and $F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ is defined by

$$F(x, u) = (-u, \nabla\phi(u) + \nabla f(x)). \tag{50}$$

Following [44], take $E_\lambda : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ defined by

$$E_\lambda(x, u) := \tfrac{1}{2}\|u\|^2 + f(x) + \lambda\langle\nabla f(x), u\rangle, \tag{51}$$

where the parameter $\lambda > 0$ will be adjusted to have the quasi-gradient property. We have

$$\nabla E_\lambda(x, u) = \big(\nabla f(x) + \lambda\nabla^2 f(x)u,\ u + \lambda\nabla f(x)\big).$$

Let us analyze the angle condition with $\Gamma = \bar{B}(0, R) \times \bar{B}(0, \epsilon)$. According to the above formulas for F and ∇E_λ, we have

$$\begin{aligned}
&\langle \nabla E_\lambda(x,u), F(x,u)\rangle \\
&\qquad = \langle (\nabla f(x) + \lambda\nabla^2 f(x)u, u + \lambda\nabla f(x)), (-u, \nabla\phi(u) + \nabla f(x))\rangle \\
&\qquad = -\langle \nabla f(x) + \lambda\nabla^2 f(x)u, u\rangle + \langle u + \lambda\nabla f(x), \nabla\phi(u) + \nabla f(x)\rangle.
\end{aligned}$$

After development and simplification, we get

$$\begin{aligned}
&\langle \nabla E_\lambda(x,u), F(x,u)\rangle \\
&\qquad = -\lambda\langle \nabla^2 f(x)u,\, u\rangle + \langle u,\, \nabla\phi(u)\rangle + \lambda\langle \nabla f(x),\, \nabla\phi(u)\rangle + \lambda\|\nabla f(x)\|^2.
\end{aligned}$$

By the local Lipschitz assumption on ∇f, let

$$M := \sup_{\|x\|\le R} \|\nabla^2 f(x)\| < +\infty.$$

Since ϕ is a damping potential, we have

$$\langle u, \nabla\phi(u)\rangle \ge \phi(u).$$

Combining the above results, we obtain

$$\begin{aligned}
&\langle \nabla E_\lambda(x,u), F(x,u)\rangle \\
&\qquad \ge -\lambda M\|u\|^2 + \phi(u) + \lambda\langle \nabla f(x),\, \nabla\phi(u)\rangle + \lambda\|\nabla f(x)\|^2 \\
&\qquad \ge -\lambda M\|u\|^2 + \phi(u) - \frac{\lambda}{2}\|\nabla f(x)\|^2 - \frac{\lambda}{2}\|\nabla\phi(u)\|^2 + \lambda\|\nabla f(x)\|^2 \\
&\qquad \ge -\lambda M\|u\|^2 + \phi(u) - \frac{\lambda}{2}\|\nabla\phi(u)\|^2 + \frac{\lambda}{2}\|\nabla f(x)\|^2. \qquad (52)
\end{aligned}$$

At this point, we use the local growth assumption on ϕ: for all u in $\mathbb{R}^N$ with $\|u\| \le \epsilon$,

$$\phi(u) \ge \gamma\|u\|^2 \quad \text{and} \quad \|\nabla\phi(u)\| \le \delta\|u\|. \qquad (53)$$

By combining (52) with (53), we obtain

$$\langle \nabla E_\lambda(x,u), F(x,u)\rangle \ge \left(\gamma - \lambda M - \frac{\lambda}{2}\delta^2\right)\|u\|^2 + \frac{\lambda}{2}\|\nabla f(x)\|^2. \qquad (54)$$

Take λ small enough to satisfy

$$\gamma > \lambda(M + \delta^2/2).$$

Then

$$\langle \nabla E_\lambda(x,u), F(x,u)\rangle \ge \alpha_0(\|u\|^2 + \|\nabla f(x)\|^2) \qquad (55)$$

with $\alpha_0 := \min\{\gamma - \lambda(M + \delta^2/2), \lambda/2\}$.

On the other hand,

$$\|\nabla E_\lambda(x,u)\| \le \sqrt{2}(1+\lambda \max\{1,M\})(\|u\|^2 + \|\nabla f(x)\|^2)^{1/2}$$
$$\|F(x,u)\| \le \sqrt{2}(1+\delta)(\|u\|^2 + \|\nabla f(x)\|^2)^{1/2}.$$

Therefore

$$\|\nabla E_\lambda(x,u)\|\,\|F(x,u)\| \le 2(1+\lambda \max\{1,M\})(1+\delta)(\|u\|^2 + \|\nabla f(x)\|^2). \quad (56)$$

As a consequence, the angle condition

$$\langle \nabla E(z), F(z)\rangle \ge \alpha \|\nabla E(z)\|\,\|F(z)\|$$

is satisfied on Γ with

$$\alpha = \frac{\min\{\gamma - \lambda(M+\delta^2/2), \lambda/2\}}{2(1+\lambda\max\{1,M\})(1+\delta)}.$$

Finally, the rest point equivalence is a consequence of the inequality (55). Then, apply the abstract Theorem 6 to obtain claims (i)–(iii). ■

Remark 5. (i) The above result allows us to consider nonlinear damping. The main restrictive assumption is that the damping potential is assumed to be nearly quadratic close to the origin. It is not necessarily quadratic close to the origin but it has to satisfy: for all u in $\mathbb{R}^N$ with $\|u\| \le \epsilon$,

$$\phi(u) \ge \gamma\|u\|^2 \quad \text{and} \quad \|\nabla\phi(u)\| \le \delta\|u\|.$$

(ii) According to [44, Proposition 3.11], a desingularizing function of f (see [44, Definition 2.1]) is desingularizing for E_λ too, for all $\lambda \in [0, \lambda_1]$.

(iii) In Sections 9.3 and 10.6 we will develop a similar analysis for related dynamical systems which involve Hessian-driven damping.

(iv) Following [44, Theorems 4.1 and 3.7], a key condition which yields convergence rates for trajectories of a quasi-gradient system with the (KŁ) property is

$$\|\nabla E_\lambda(x,u)\| \le b\|F(x,u)\| \quad \text{for } (x,u) \in \Gamma \quad (57)$$

with $b > 0$. Let us check this condition in the setting of Theorem 7. We have seen there that

$$\|\nabla E_\lambda(x,u)\|^2 \le C_1(\|u\|^2 + \|\nabla f(x)\|^2)$$

for some $C_1 > 0$. Further, from the Cauchy–Schwarz inequality and the properties of ϕ we derive, for all $\sigma > 1$,

$$\begin{aligned}
\|F(x,u)\|^2 &= \|u\|^2 + \|\nabla\phi(u) + \nabla f(x)\|^2 \\
&\ge \|u\|^2 + \|\nabla\phi(u)\|^2 + \|\nabla f(x)\|^2 - 2\|\nabla\phi(x)\|\,\|\nabla f(x)\| \\
&\ge \|u\|^2 + \|\nabla\phi(u)\|^2 + \|\nabla f(x)\|^2 - \sigma\|\nabla\phi(u)\|^2 - \frac{1}{\sigma}\|\nabla f(x)\|^2 \\
&\ge \left(1 - \frac{1}{\sigma}\right)\|\nabla f(x)\|^2 + (1-(\sigma-1)\delta^2)\|u\|^2.
\end{aligned}$$

Hence, we can choose $\sigma > 1$ such that

$$\|F(x,u)\|^2 \geq C_3(\|u\|^2 + \|\nabla f(x)\|^2)$$

for some $C_3 > 0$. Condition (57) is now met with $b = \sqrt{C_1/C_3}$.

As in [44, Section 5], explicit convergence rates can be derived from [44, Theorems 4.1 and 3.7], based on [44, (3.19) and Remark 3.4(c)].

7.4. *Application: f with polynomial growth*

This concerns the question raised at the end of Section 5.1. Additionally to the hypotheses of Theorem 7, assume that f is convex, $\operatorname{argmin} f \neq \emptyset$ and for each $x^* \in \operatorname{argmin} f$, there exists $\eta > 0$ such that

$$f(x) - \inf_{\mathcal{H}} f \geq c \operatorname{dist}(x, \operatorname{argmin} f)^r \quad \forall x \in B(x^*, \eta),$$

for some $r \geq 1$ and $c > 0$.

According to [44, proof of Corollary 5.5], f satisfies the Łojasiewicz inequality with desingularizing function (see [44, Definition 2.1]) of the form $\varphi(s) = c's^{1/r}$ with $c' > 0$. By [44, Proposition 3.11], this is a desingularizing function of E_λ too, for all $\lambda \in [0, \lambda_1]$ (with E_λ defined in Theorem 7). In Remark 5 (iv) we have shown that (57) holds. Relying now on [44, Theorem 3.7 and Remark 3.4(c)], we derive sublinear rates for $\|x(t) - x_\infty\|$ in case $r < 2$ and an exponential rate in case $r = 2$.

7.5. *Application: fixed damping matrix*

We will recover and improve the results of Alvarez [8, Theorem 2.6], which concern the case of f convex, and the damped inertial equation

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

where $A : \mathcal{H} \to \mathcal{H}$ is a positive definite self-adjoint linear operator, which is possibly anisotropic (see also [47]). While the proof of [8, Theorem 2.6] works in general Hilbert spaces, we have to restrict ourselves to finite-dimensional spaces, but we can drop the convexity assumption on f. The following result is a direct consequence of Theorem 7 applied to $\phi : \mathbb{R}^N \to \mathbb{R}_+$, $\phi(x) = \frac{1}{2}\langle Ax, x\rangle$, where $A : \mathbb{R}^N \to \mathbb{R}^N$ is a positive definite self-adjoint linear operator. We note that in this setting, ϕ is a damping potential, it is convex continuous and attains its minimum at the origin. Moreover, the local and global growth conditions are met. Indeed, for all $u \in \mathbb{R}^N$, we have $\phi(u) \geq \frac{1}{2}\lambda_{\min}\|u\|^2$ and $\|\nabla\phi(u)\| \leq \lambda_{\max}\|u\|$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and the largest positive eigenvalues of A respectively.

Theorem 8. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a $\mathcal{C}^2$ function whose gradient is Lipschitz continuous on bounded sets, and such that $\inf_{\mathbb{R}^N} f > -\infty$. Let $A : \mathbb{R}^N \to \mathbb{R}^N$ be a positive definite* *self-adjoint linear operator. Suppose that the function E_λ is* (KŁ) (*which is true if f satisfies* (KŁ)) *where*

$$E_\lambda(x,u) := \tfrac{1}{2}\|u\|^2 + f(x) + \lambda\langle\nabla f(x), u\rangle.$$

Let $x : [0,+\infty[\to \mathbb{R}^N$ be a bounded solution of

$$\ddot{x}(t) + A(\dot{x}(t)) + \nabla f(x(t)) = 0.$$

Then

(i) $x(t) \to x_\infty$ *as* $t \to +\infty$, *where* $x_\infty \in \operatorname{crit} f$;

(ii) $\dot{x} \in L^1(0,+\infty;\mathbb{R}^N)$ *and* $\dot{x}(t) \to 0$ *as* $t \to +\infty$;

(iii) $f(x(t)) \to f(x_\infty) \in f(\operatorname{crit} f)$ *as* $t \to +\infty$.

Indeed, we can complete this result with the convergence rates which are linked to the desingularization function provided by the (KŁ) property of f.

8. Algorithmic results: an inertial type algorithm

The following convergence result is a discrete algorithmic version of Theorem 7. To stay close to the continuous dynamic we use a semi-implicit discretization: implicit with respect to the damping potential ϕ, and explicit with respect to the function f to minimize. This will make it possible to make minimal assumptions on the damping potential ϕ, and thus cover various situations. As above, the underlying structure of the proof is the quasi-gradient property. We give a direct proof, which is a bit simpler in this case. Consider the following temporal discretization of (ADIGE-V) with step size $h > 0$:

$$\frac{1}{h^2}(x_{n+2} - 2x_{n+1} + x_n) + \nabla\phi\left(\frac{1}{h}(x_{n+2} - x_{n+1})\right) + \nabla f(x_{n+1}) = 0.$$

Equivalently,

$$\frac{x_{n+2} - x_{n+1}}{h} - \frac{x_{n+1} - x_n}{h} + h\nabla\phi\left(\frac{x_{n+2} - x_{n+1}}{h}\right) + h\nabla f(x_{n+1}) = 0. \tag{58}$$

This gives the proximal-gradient algorithm

$$x_{n+2} = x_{n+1} + h\,\operatorname{prox}_{h\phi}\left(\frac{1}{h}(x_{n+1} - x_n) - h\nabla f(x_{n+1})\right). \tag{59}$$

Recall that, for any $x \in \mathcal{H} = \mathbb{R}^N$ and any $\lambda > 0$,

$$\operatorname{prox}_{\lambda\phi}(x) := \operatorname*{argmin}_{\xi\in\mathcal{H}}\left\{\lambda\phi(\xi) + \tfrac{1}{2}\|x - \xi\|^2\right\}.$$

Let us start by establishing a decrease property for the sequence $(W_n)_{n\in\mathbb{N}}$ of global energies

$$W_n := \tfrac{1}{2}\|u_n\|^2 + f(x_{n+1}),$$

where $u_n = \frac{1}{h}(x_{n+1} - x_n)$ is the discrete velocity.

Lemma 2. *Suppose that $f : \mathbb{R}^N \to \mathbb{R}$ is a differentiable function whose gradient is L-Lipschitz continuous on a ball containing the iterates $(x_n)_{n\in\mathbb{N}}$, and such that $\inf_{\mathbb{R}^N} f > -\infty$. Suppose that $\phi(u) \geq \gamma\|u\|^2$ for all $u \in \mathbb{R}^N$. Then, for all $n \in \mathbb{N}$,*

$$W_{n+1} - W_n + h\big(\gamma - \tfrac{1}{2}Lh\big)\|u_{n+1}\|^2 \leq 0.$$

As a consequence, under the assumption $\gamma > \frac{1}{2}Lh$, we have

$$\sum_{n\in\mathbb{N}} \|u_n\|^2 < +\infty \quad \text{and} \quad u_n \to 0 \text{ as } n \to +\infty.$$

Proof. Write the algorithm as

$$u_{n+1} - u_n + h\nabla\phi(u_{n+1}) + h\nabla f(x_{n+1}) = 0.$$

By taking the scalar product with u_{n+1}, we obtain

$$\langle u_{n+1} - u_n, u_{n+1}\rangle + h\langle\nabla\phi(u_{n+1}), u_{n+1}\rangle + \langle\nabla f(x_{n+1}), x_{n+2} - x_{n+1}\rangle = 0. \tag{60}$$

We have

$$\begin{aligned}
&\langle u_{n+1} - u_n, u_{n+1}\rangle \geq \tfrac{1}{2}\|u_{n+1}\|^2 - \tfrac{1}{2}\|u_n\|^2,\\
&\langle\nabla\phi(u_{n+1}), u_{n+1}\rangle \geq \phi(u_{n+1}) \geq \gamma\|u_{n+1}\|^2,\\
&f(x_{n+2}) - f(x_{n+1}) \leq \langle\nabla f(x_{n+1}), x_{n+2} - x_{n+1}\rangle + \tfrac{1}{2}Lh^2\|u_{n+1}\|^2,
\end{aligned}$$

where the last inequality follows from the descent gradient lemma. By combining the above inequalities with (60), we obtain

$$\tfrac{1}{2}\|u_{n+1}\|^2 - \tfrac{1}{2}\|u_n\|^2 + h\gamma\|u_{n+1}\|^2 + f(x_{n+2}) - f(x_{n+1}) - \tfrac{1}{2}Lh^2\|u_{n+1}\|^2 \leq 0. \tag{61}$$

Equivalently,

$$W_{n+1} - W_n + h\big(\gamma - \tfrac{1}{2}Lh\big)\|u_{n+1}\|^2 \leq 0.$$

By summing the above inequalites, and since f is bounded below, we get

$$h\big(\gamma - \tfrac{1}{2}Lh\big)\sum_{n\geq 1}\|u_n\|^2 \leq W_0 - \inf f.$$

Since $\gamma - \frac{1}{2}Lh > 0$, we get $\sum_{n\in\mathbb{N}}\|u_n\|^2 < +\infty$, and hence $u_n \to 0$ as $n \to +\infty$. ∎

Theorem 9. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a $\mathcal{C}^2$ function whose gradient is Lipschitz continuous on bounded sets, and such that $\inf_{\mathbb{R}^N} f > -\infty$. Let $\phi : \mathbb{R}^N \to \mathbb{R}_+$ be a damping potential (see Definition 1) which is differentiable. Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence generated by the algorithm*

$$x_{n+2} = x_{n+1} + h\,\mathrm{prox}_{h\phi}\left(\frac{1}{h}(x_{n+1} - x_n) - h\nabla f(x_{n+1})\right). \tag{62}$$

We make the following assumptions on the data f, ϕ, and h:

- (Assumption on f) *Suppose that the function H satisfies the* (KŁ) *property, where $H : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is defined for all $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$ by*

$$H(x, y) := f(x) + \frac{1}{2h^2}\|x - y\|^2.$$

- (Assumption on ϕ) *Suppose that ϕ satisfies the following growth conditions: there exist $\gamma, \varepsilon, \delta > 0$ such that $\phi(u) \geq \gamma\|u\|^2$ for all u in $\mathbb{R}^N$, and $\|\nabla\phi(u)\| \leq \delta\|u\|$ for all u with $\|u\| \leq \varepsilon$.*

- (Assumption on h) *Suppose that the step size h is so small that*

$$0 < h < 2\gamma/L,$$

where L is the Lipschitz constant of ∇f on the ball centered at the origin and with radius $R = \sup_{n\in\mathbb{N}} \|x_n\|$.

Then

(i) $x_n \to x_\infty$ *as* $n \to +\infty$, *where* $x_\infty \in \operatorname{crit} f$;

(ii) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\| < +\infty$.

Proof. By assumption, the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded. Lemma 2 shows that $(u_n)_{n\in\mathbb{N}}$ tends to zero, where $u_n = \frac{1}{h}(x_{n+1} - x_n)$. In addition,

$$W_{n+1} - W_n + h\big(\gamma - \tfrac{1}{2}Lh\big)\|u_{n+1}\|^2 \leq 0$$

for all $n \in \mathbb{N}$, where

$$W_n := \tfrac{1}{2}\|u_n\|^2 + f(x_{n+1}).$$

Equivalently, by setting

$$H(x, y) = f(x) + \frac{1}{2h^2}\|x - y\|^2,$$

we have, for all $n \in \mathbb{N}$,

$$H(x_{n+2}, x_{n+1}) + C\|x_{n+2} - x_{n+1}\|^2 \leq H(x_{n+1}, x_n), \tag{63}$$

where $C = \frac{1}{h}(\gamma - \frac{1}{2}Lh) > 0$. The rest of the proof is classical in the framework of (KŁ) theory; we refer the reader to [49, 82] for similar techniques relying on the above decrease property. Relation (63) implies that

$$\lim_{n\to+\infty} H(x_{n+1}, x_n) \in \mathbb{R} \tag{64}$$

exists. Further, let us denote by $\omega((x_n)_{n\in\mathbb{N}})$ the set of cluster points of the sequence $(x_n)_{n\in\mathbb{N}}$, and recall that $\operatorname{crit} f = \{x \in \mathbb{R}^N : \nabla f(x) = 0\}$ is the set of critical points of f.

We easily derive

$$\operatorname{crit} H = \{(x, x) \in \mathbb{R}^N \times \mathbb{R}^N : x \in \operatorname{crit} f\}$$

and notice that $\omega((x_n)_{n\in\mathbb{N}}) \subseteq \operatorname{crit} f$, thus $\omega((x_{n+1}, x_n)_{n\in\mathbb{N}}) \subseteq \operatorname{crit} H$. From (64) one can easily conclude that H is constant on $\omega((x_{n+1}, x_n))$. Indeed, for $x^* \in \omega((x_n)_{n\in\mathbb{N}})$, we see from the above (and the definition of H) that

$$\lim_{n\to+\infty} H(x_{n+1}, x_n) = f(x^*) = H(x^*, x^*). \tag{65}$$

Assume now that H satisfies the (KŁ) property with desingularizing function θ. We consider two cases.

I. There exists $\overline{n} \geq 0$ such that $H(x_{\overline{n}+1}, x_{\overline{n}}) = H(x^*, x^*)$. From the decreasing property (63) we find that $(x_n)_{n\geq\overline{n}}$ is a constant sequence and the conclusion follows.

II. For all $n \geq 0$ we have $H(x_{n+1}, x_n) > H(x^*, x^*)$. Since θ is concave and $\theta' > 0$, we deduce from (63) that there exists $n' \geq 0$ such that for all $n \geq n'$,

$$\begin{aligned}
\Delta_{n,n+1} &:= \theta(H(x_{n+1}, x_n) - H(x^*, x^*)) - \theta(H(x_{n+2}, x_{n+1}) - H(x^*, x^*)) \\
&\geq \theta'(H(x_{n+1}, x_n) - H(x^*, x^*)) \cdot (H(x_{n+1}, x_n) - H(x_{n+2}, x_{n+1})) \\
&\geq \theta'(H(x_{n+1}, x_n) - H(x^*, x^*)) \cdot C \cdot \|x_{n+2} - x_{n+1}\|^2 \\
&\geq \frac{C\|x_{n+2} - x_{n+1}\|^2}{\|\nabla H(x_{n+1}, x_n)\|},
\end{aligned} \tag{66}$$

where the last inequality follows from the uniformized (KŁ) property [46, Lemma 6] applied to the nonempty compact and connected set $\Omega = \omega((x_{n+1}, x_n)_{n\in\mathbb{N}})$ (by [46, Remark 5] the connectedness is generic for sequences satisfying $\lim_{n\to+\infty}(x_{n+1} - x_n) = 0$).

Further, since $\nabla H(x, y) = (\nabla f(x) + \frac{1}{h^2}(x - y), \frac{1}{h^2}(y - x))$, we deduce from (58), the fact that $\lim_{n\to+\infty}(x_{n+1} - x_n) = 0$ and the properties of ϕ that there exists $C_2 > 0$ such that

$$\|\nabla H(x_{n+1}, x_n)\| \leq C_2(\|x_{n+2} - x_{n+1}\| + \|x_{n+1} - x_n\|) \quad \forall n \in \mathbb{N}.$$

Hence there exist $C_3 > 0$ and $n'' \in \mathbb{N}$ such that for all $n \geq n''$,

$$\frac{a_{n+1}^2}{a_{n+1} + a_n} \leq C_3\Delta_{n,n+1},$$

where $a_n := \|x_{n+1} - x_n\|$. From this we see that for all $n \geq n''$,

$$a_{n+1} = \sqrt{C_3\Delta_{n,n+1}(a_{n+1} + a_n)} \leq \frac{a_{n+1} + a_n}{4} + C_3\Delta_{n,n+1},$$

which implies

$$a_{n+1} \leq \tfrac{1}{3}a_n + \tfrac{4}{3}C_3\Delta_{n,n+1}.$$

Summing the last inequality over n we obtain $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\| < +\infty$. This classically implies that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}^N$, hence it converges to a critical point of f. ■

Remark 6. (i) If $f : \mathbb{R}^N \to \mathbb{R}$ is a $\mathcal{C}^2$ coercive function whose gradient is Lipschitz continuous on $\mathbb{R}^N$, then the boundedness of the sequence $(x_n)_{n\in\mathbb{N}}$ follows from (63). The function H is a (KŁ) function if f is, for instance, semialgebraic; we refer to [77] for other results related to the preservation of the (KŁ) property under addition.

(ii) For a general damping function ϕ we obtain in the limit

$$\nabla f(x_\infty) + \partial\phi(0) \ni 0.$$

When ϕ is differentiable at the origin, it attains its minimum at this point, and hence $\nabla\phi(0) = 0$. So, we get $\nabla f(x_\infty) = 0$, i.e. x_∞ is a critical point of f, which is the situation considered above. In the case of dry friction, for example $\phi(u) = r\|u\|$, we get $\|\nabla f(x_\infty)\| \le r$, which gives an approximate critical point [2,3,5].

(iii) The above result has been given as an illustration of our results, showing that the continuous dynamic approach gives a valuable guideline to develop corresponding algorithmic results. In the particular case $\phi(u) = \|u\|^2$ one can also consult [66,82]. The explicit discretization gives rise to inertial gradient algorithms, an interesting subject to explore in this general setting.

9. Closed-loop velocity control with Hessian-driven damping

9.1. Hessian-driven damping

We now tackle questions similar to the previous sections, concerning the combination of closed-loop velocity control with Hessian-driven damping. The following system combines closed-loop velocity control with Hessian-driven damping:

$$\ddot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0. \tag{67}$$

This autonomous system will be our main object of study in this section.

• The case $\phi(u) = \frac{\gamma}{2}\|u\|^2$ of a fixed viscous coefficient was first considered by Alvarez–Attouch–Bolte–Redont [11]. In this case, (67) can be equivalently written as a first-order system in time and space (different from the Hamiltonian formulation), which allows one to extend this system naturally to the case of a nonsmooth function f. This property has been exploited by Attouch–Maingé–Redont [31] to model nonelastic shocks in unilateral mechanics. To accelerate this system, several recent studies considered the case where the viscous damping is vanishing, that is,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0; \tag{68}$$

see [24,35,52,58,76,79,89] and Section 2.3 for the properties of this system.

• The case $\phi(u) = \frac{\gamma}{2}\|u\|^2 + r\|u\|$ which combines viscous friction with dry friction and Hessian-driven damping has been considered by Adly–Attouch [3,4].

• By taking $\phi(u) = \frac{r}{p}\|u\|^p$, we get

$$\ddot{x}(t) + r\|\dot{x}(t)\|^{p-2}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0, \tag{69}$$

for which we will address similar issues. In addition to the fast minimization property, one can expect fast convergence of gradients to zero.

9.2. *Existence and uniqueness results*

Let us consider the differential inclusion

$$\text{(ADIGE-VH)} \qquad \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0, \tag{70}$$

which involves a damping potential ϕ, and geometric damping driven by the Hessian of f. The suffix V refers to velocity and H to Hessian. They both enter the damping terms. This allows one to cover different situations, in particular system (69) corresponds to $\phi(u) = \frac{r}{p}\|u\|^p$ for $p > 1$. To prove existence and uniqueness results for the associated Cauchy problem, we make additional assumptions. We assume that f is convex, and that the Hessian mapping $\mathcal{H} \ni x \mapsto \nabla^2 f(x) \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ is Lipschitz continuous on bounded sets, where $\mathcal{L}(\mathcal{H}, \mathcal{H})$ is equipped with the norm operator. Note that this property implies that ∇f is Lipschitz continuous on bounded subsets of $\mathcal{H}$ (apply the mean value theorem in the vectorial case). However, in the following statement, we formulate the two hypotheses for the sake of clarity.

Theorem 10. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function which is twice continuously differentiable, and $\inf_{\mathcal{H}} f > -\infty$. Suppose that*

(i) *∇f is Lipschitz continuous on bounded subsets of $\mathcal{H}$;*

(ii) *$\nabla^2 f$ is Lipschitz continuous on bounded subsets of $\mathcal{H}$.*

Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous damping function. Then, for any Cauchy data $(x_0, x_1) \in \mathcal{H} \times \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\to \mathcal{H}$ of (ADIGE-VH) *satisfying $x(0) = x_0$ and $\dot{x}(0) = x_1$.*

Proof. To make the reading of the proof easier, we divide the proof into several steps.

Step 1: *A priori estimate.* Let us establish a priori energy estimates of solutions of (70). After taking the scalar product of (70) with $\dot{x}(t)$, we get

$$\frac{d}{dt}\mathcal{E}(t) + \langle\partial\phi(\dot{x}(t)), \dot{x}(t)\rangle + \beta\langle\nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\rangle = 0,$$

where $\mathcal{E}(t) := f(x(t)) - \inf_{\mathcal{H}} f + \frac{1}{2}\|\dot{x}(t)\|^2$ is the global energy. Since ϕ is a damping potential, the subdifferential inequality for convex functions, combined with $\phi(0) = 0$, gives

$$\langle\partial\phi(\dot{x}(t)), \dot{x}(t)\rangle \geq \phi(\dot{x}(t)).$$

Since f is convex, $\nabla^2 f$ is positive semidefinite, which gives

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

Collecting the above results, we obtain the following decay property of the energy:

$$\frac{d}{dt}\mathcal{E}(t) + \phi(\dot{x}(t)) \leq 0. \tag{71}$$

Therefore, the energy is nonincreasing, which implies that, as long as the trajectory is defined,

$$\|\dot{x}(t)\|^2 \leq 2\mathcal{E}(0). \tag{72}$$

Step 2: *Hamiltonian formulation of* (70). According to the Hamiltonian formulation of (70), it is equivalent to solve the first-order system

$$\begin{cases} \dot{x}(t) - u(t) = 0, \\ \dot{u}(t) + \partial\phi(u(t)) + \nabla f(x(t)) + \beta\nabla^2 f(x(t))u(t) \ni 0, \end{cases}$$

with the Cauchy data $x(0) = x_0$, $u(0) = x_1$. Set $Z(t) = (x(t), u(t)) \in \mathcal{H} \times \mathcal{H}$. The above system can be written equivalently as

$$\dot{Z}(t) + F(Z(t)) \ni 0, \quad Z(0) = (x_0, x_1),$$

where $F : \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}$, $(x, u) \mapsto F(x, u)$, is defined by

$$F(x, u) = (0, \partial\phi(u)) + (-u, \nabla f(x) + \beta\nabla^2 f(x)u).$$

Hence F splits as $F(x, u) = \partial\Phi(x, u) + G(x, u)$, where

$$\Phi(x, u) = \phi(u) \quad \text{and} \quad G(x, u) = (-u,\ \nabla f(x) + \beta\nabla^2 f(x)u). \tag{73}$$

Therefore, it is equivalent to solve the following first-order differential inclusion with Cauchy data:

$$\dot{Z}(t) + \partial\Phi(Z(t)) + G(Z(t)) \ni 0, \quad Z(0) = (x_0, x_1). \tag{74}$$

Let us prove that the mapping $(x, u) \mapsto G(x, u)$ is Lipschitz continuous on bounded subsets of $\mathcal{H} \times \mathcal{H}$. For any $(x, u) \in \mathcal{H} \times \mathcal{H}$, set $G(x, u) = (-u, K(x, u))$ where

$$K(x, u) := \nabla f(x) + \beta\nabla^2 f(x)u.$$

Let L_R be the Lipschitz constant of ∇f and $\nabla^2 f$ on the ball centered at the origin and with radius R, and set $M_R = \sup_{\|x\|\leq R} \|\nabla^2 f(x)\|$. Take $(x_i, u_i) \in \mathcal{H} \times \mathcal{H}$, $i = 1, 2$, with $\|(x_i, u_i)\| \leq R$. We have

$$K(x_2, u_2) - K(x_1, u_1) = \nabla f(x_2) - \nabla f(x_1) + \beta(\nabla^2 f(x_2)u_2 - \nabla^2 f(x_1)u_1).$$

By the triangle inequality, and the local Lipschitz continuity of ∇f and $\nabla^2 f$,

$$\begin{aligned} \|K(x_2, u_2) - K(x_1, u_1)\| &\leq \|\nabla f(x_2) - \nabla f(x_1)\| + \beta\|\nabla^2 f(x_2)u_2 - \nabla^2 f(x_1)u_2\| \\ &\quad + \beta\|\nabla^2 f(x_1)u_2 - \nabla^2 f(x_1)u_1\| \\ &\leq L_R\|x_2 - x_1\| + \beta L_R\|x_2 - x_1\|\,\|u_2\| + \beta M_R\|u_2 - u_1\| \\ &\leq L_R(1 + R\beta)\|x_2 - x_1\| + \beta M_R\|u_2 - u_1\|. \end{aligned}$$

Therefore,

$$\|G(x_2,u_2)-G(x_1,u_1)\| \leq L_R(1+R\beta)\|x_2-x_1\|+(1+\beta M_R)\|u_2-u_1\|, \quad (75)$$

so $(x,u)\mapsto G(x,u)$ is Lipschitz continuous on bounded subsets of $\mathcal{H}\times\mathcal{H}$.

Step 3: *Approximate dynamics.* We proceed as in Theorem 3 (which corresponds to the case $\beta=0$), and consider the approximate dynamics

$$\ddot{x}_\lambda(t)+\nabla\phi_\lambda(\dot{x}_\lambda(t))+\beta\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)+\nabla f(x_\lambda(t))=0, \quad t\in[0,+\infty[, \quad (76)$$

which uses the Moreau–Yosida approximations (ϕ_λ) of ϕ. We will prove that the filtered sequence (x_λ) converges uniformly as $\lambda\to 0$ over bounded time intervals to a solution of (70). The Hamiltonian formulation of (76) gives the first-order (in time) system

$$\begin{cases}\dot{x}_\lambda(t)-u_\lambda(t)=0,\\ \dot{u}_\lambda(t)+\nabla\phi_\lambda(u_\lambda(t))+\nabla f(x_\lambda(t))+\beta\nabla^2 f(x_\lambda(t))u_\lambda(t)=0,\end{cases}$$

with the Cauchy data $x_\lambda(0)=x_0$, $u_\lambda(0)=x_1$. Set $Z_\lambda(t)=(x_\lambda(t),u_\lambda(t))\in\mathcal{H}\times\mathcal{H}$. The above system can be written equivalently as

$$\dot{Z}_\lambda(t)+F_\lambda(Z_\lambda(t))\ni 0, \quad Z_\lambda(t_0)=(x_0,x_1),$$

where $F_\lambda:\mathcal{H}\times\mathcal{H}\to\mathcal{H}\times\mathcal{H}$, $(x,u)\mapsto F_\lambda(x,u)$, is defined by

$$F_\lambda(x,u)=(0,\nabla\phi_\lambda(u))+(-u,\nabla f(x)+\beta\nabla^2 f(x)u).$$

Hence $F_\lambda(x,u)=\nabla\Phi_\lambda(x,u)+G(x,u)$ where Φ and G have been defined in (73). Therefore, the approximate equation is equivalent to the first-order differential system with Cauchy data

$$\dot{Z}_\lambda(t)+\nabla\Phi_\lambda(Z_\lambda(t))+G(Z_\lambda(t))=0, \quad Z_\lambda(0)=(x_0,x_1). \quad (77)$$

Fix $\lambda>0$. According to the Lipschitz continuity of $\nabla\Phi_\lambda$, and the fact that G is Lipschitz continuous on bounded sets, the operator $\nabla\Phi_\lambda+G$ which governs (77) is Lipschitz continuous on bounded sets. As a consequence, the existence of a local solution to (77) follows from the classical Cauchy–Lipschitz theorem. To pass from a local solution to a global solution, we use the a priori estimate obtained in Step 1 of the proof. Note that this estimate is valid for any damping potential, in particular for ϕ_λ. In view of the Cauchy data, and since f is bounded below, this implies that, on any bounded time interval, the functions (x_λ) and $(\dot{x}_\lambda)$ are bounded. By the property (9) of Yosida approximation, and the property (iii) of the damping potential ϕ, this implies that

$$\|\nabla\phi_\lambda(x_\lambda(t))\|\leq\|(\partial\phi)^0(x_\lambda(t))\|$$

is also bounded uniformly for t bounded. Moreover, in view of the local boundedness assumption made on the gradient and the Hessian of f, we infer that $\nabla f(x_\lambda(t))$ and $\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)$ are also bounded. By the constitutive equation (76), this in turn implies that $(\ddot{x}_\lambda)$ is also bounded. This implies that if a maximal solution is defined on a finite time interval $[0,T[$, then the limits of $x_\lambda(t)$ and $\dot{x}_\lambda(t)$ as $t \to T$ exist. Hence, passing from a local to a global solution is a classical argument. So for any $\lambda > 0$ we have a unique global solution of (76) with satisfies the Cauchy data $x_\lambda(0) = x_0$, $\dot{x}_\lambda(0) = x_1$.

Step 4: *Passing to the limit as* $\lambda \to 0$. Take $T > 0$ and $\lambda, \mu > 0$. Consider the corresponding solutions on $[0,T]$,

$$\begin{aligned} &\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1),\\ &\dot{Z}_\mu(t) + \nabla\Phi_\mu(Z_\mu(t)) + G(Z_\mu(t)) = 0, \quad Z_\mu(0) = (x_0, x_1). \end{aligned}$$

Subtracting the two equations, and taking the scalar product with $Z_\lambda(t) - Z_\mu(t)$, we get

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 &+ \langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \\ &+ \langle G(Z_\lambda(t)) - G(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle = 0. \end{aligned} \tag{78}$$

We now use the following ingredients:

(i) By the properties of Yosida approximation [53, Theorem 3.1], we have

$$\begin{aligned} &\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \\ &\qquad \geq -\frac{\lambda}{4}\|\nabla\Phi_\mu(Z_\mu(t))\|^2 - \frac{\mu}{4}\|\nabla\Phi_\lambda(Z_\lambda(t))\|^2. \end{aligned}$$

In view of the energy estimates, the sequence (Z_λ) is uniformly bounded on $[0,T]$:

$$\|Z_\lambda(t)\| \leq C_T.$$

From these properties we immediately infer that

$$\|\nabla\Phi_\lambda(Z_\lambda(t))\| \leq \sup_{\|\xi\|\leq C_T} \|(\partial\phi)^0(\xi)\| = M_T < +\infty,$$

because our assumption on ϕ gives that $(\partial\phi)^0$ is bounded on bounded sets. Therefore

$$\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \geq -\tfrac{1}{4}M_T(\lambda + \mu).$$

(ii) Since $G : \mathcal{H}\times\mathcal{H} \to \mathcal{H}\times\mathcal{H}$ is Lipschitz continuous on bounded sets, and (Z_λ) is uniformly bounded on $[0,T]$, there exists a constant L_T such that

$$\|G(Z_\lambda(t)) - G(Z_\mu(t))\| \leq L_T\|Z_\lambda(t) - Z_\mu(t)\|.$$

Combining the above results, and using the Cauchy–Schwarz inequality, we deduce from (78) that

$$\frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{1}{4}M_T(\lambda+\mu) + L_T\|Z_\lambda(t) - Z_\mu(t)\|^2.$$

Integrating this differential inequality and taking into account that $Z_\lambda(0) - Z_\mu(0) = 0$, elementary calculus gives

$$\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{M_T}{4L_T}(\lambda + \mu)(e^{2L_T(t-t_0)} - 1).$$

Therefore, (Z_λ) is a Cauchy sequence for uniform convergence on $[0, T]$, and hence it converges uniformly. This means that x_λ and $\dot{x}_\lambda$ converge uniformly on $[0, T]$ to x and $\dot{x}$ respectively. That x is solution of (70) is proved in a similar way to Theorem 3. Just rely on the classical chain rule $\frac{d}{dt}(\nabla f(x_\lambda(t))) = \nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)$ to pass to the limit in the Hessian term. ∎

9.3. Convergence based on the quasi-gradient approach

Our objective is to address, from the perspective of quasi-gradient systems, the system (ADIGE-VH)

$$\ddot{x}(t) + \nabla\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0, \tag{79}$$

as in Section 7.3. We assume that $\mathcal{H} = \mathbb{R}^N$ is a finite-dimensional Hilbert space, and that the hypotheses of Theorems 7 and 10 hold. We follow the steps of the proof of Theorem 7. By using the estimates in Step 1 of the proof of Theorem 10, we easily derive the first part of the proof of Theorem 7, namely that the trajectory $t \mapsto (x(t), \dot{x}(t))$ in the phase space $\mathbb{R}^N \times \mathbb{R}^N$ lies in the closed bounded set $\Gamma = \bar{B}(0, R) \times \bar{B}(0, \epsilon)$. According to Step 2 in the proof of Theorem 10, the Hamiltonian formulation of (79) gives the first-order differential system

$$\dot{z}(t) + F(z(t)) = 0, \tag{80}$$

where $z(t) = (x(t), \dot{x}(t)) \in \mathbb{R}^N \times \mathbb{R}^N$, and $F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ is defined by

$$F(x, u) = (-u, \nabla\phi(u) + \nabla f(x) + \beta\nabla^2 f(x)u).$$

Let us focus on the key point which is the angle condition (E_λ is defined as in Theorem 7). We have

$$\begin{aligned}
&\langle \nabla E_\lambda(x, u), F(x, u)\rangle \\
&\quad = \langle (\nabla f(x) + \lambda\nabla^2 f(x)u,\ u + \lambda\nabla f(x)), (-u, \nabla\phi(u) + \nabla f(x) + \beta\nabla^2 f(x)u)\rangle.
\end{aligned}$$

After simplification, we get

$$\begin{aligned}
\langle \nabla E_\lambda(x, u), F(x, u)\rangle& \\
= -\lambda\langle \nabla^2 f(x)u,\ u\rangle &+ \langle u,\ \nabla\phi(u)\rangle + \lambda\langle \nabla f(x),\ \nabla\phi(u)\rangle + \lambda\|\nabla f(x)\|^2 \\
&+ \beta\langle u + \lambda\nabla f(x), \nabla^2 f(x)u\rangle \\
\geq -\lambda\langle \nabla^2 f(x)u,\ u\rangle &+ \langle u,\ \nabla\phi(u)\rangle + \lambda\langle \nabla f(x),\ \nabla\phi(u)\rangle + \lambda\|\nabla f(x)\|^2 \\
&+ \lambda\beta\langle \nabla f(x), \nabla^2 f(x)u\rangle,
\end{aligned}$$

where we have used the fact that $\nabla^2 f(x)$ is positive semidefinite. The only difference with respect to the next step in the proof of Theorem 7 is that we need to estimate the extra term $\lambda\beta\langle\nabla f(x), \nabla^2 f(x)u\rangle$. We do this by writing $\lambda\beta\langle\nabla f(x), \nabla^2 f(x)u\rangle \geq -\frac{\lambda}{4}\|\nabla f(x)\|^2 - \lambda\beta^2 M^2\|u\|^2$, and get

$$\langle\nabla E_\lambda(x,u), F(x,u)\rangle \geq \left(\gamma - \lambda M - \frac{\lambda}{2}\delta^2 - \lambda\beta^2 M^2\right)\|u\|^2 + \frac{\lambda}{4}\|\nabla f(x)\|^2. \tag{81}$$

Take λ small enough to satisfy $\gamma > \lambda(M + \delta^2/2 + \beta^2 M^2)$. Then

$$\langle\nabla E_\lambda(x,u), F(x,u)\rangle \geq \alpha_0(\|u\|^2 + \|\nabla f(x)\|^2), \tag{82}$$

with $\alpha_0 := \min\{\gamma - \lambda(M + \delta^2/2 + \beta^2 M^2), \lambda/4\}$. On the other hand, as in Theorem 7,

$$\|\nabla E_\lambda(x,u)\| \leq C_1(\|u\|^2 + \|\nabla f(x)\|^2)^{1/2},$$
$$\|F(x,u)\| \leq C_2(\|u\|^2 + \|\nabla f(x)\|^2)^{1/2},$$

where $C_2 = \sqrt{4 + 3\delta^2 + 3\beta^2 M^2}$. Therefore

$$\|\nabla E_\lambda(x,u)\|\,\|F(x,u)\| \leq C_1 C_2(\|u\|^2 + \|\nabla f(x)\|^2). \tag{83}$$

Therefore, for $\alpha := \frac{\alpha_0}{C_1 C_2}$, the angle condition $\langle\nabla E(z), F(z)\rangle \geq \alpha\|\nabla E(z)\|\,\|F(z)\|$ is satisfied on Γ. Let us summarize the above results.

Theorem 11. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function which is twice continuously differentiable, and such that $\inf_{\mathcal{H}} f > -\infty$. Suppose that*

(i) *∇f is Lipschitz continuous on bounded subsets of $\mathcal{H}$;*

(ii) *$\nabla^2 f$ is Lipschitz continuous on bounded subsets of $\mathcal{H}$.*

Let $E_\lambda : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be defined by

$$E_\lambda(x,u) := \tfrac{1}{2}\|u\|^2 + f(x) + \lambda\langle\nabla f(x), u\rangle \quad \textit{for } (x,u) \in \mathbb{R}^N \times \mathbb{R}^N.$$

Suppose that E_λ satisfies the (KŁ) *property. Let $\phi : \mathbb{R}^N \to \mathbb{R}_+$ be a damping potential which is differentiable, and which satisfies the following growth conditions:*

(i) (local) *there exist constants $\gamma, \delta, \epsilon > 0$ such that, for all u in $\mathbb{R}^N$ with $\|u\| \leq \epsilon$,*

$$\phi(u) \geq \gamma\|u\|^2 \quad \textit{and} \quad \|\nabla\phi(u)\| \leq \delta\|u\|.$$

(ii) (global) *there exist $p \geq 1$ and $c > 0$ such that $\phi(u) \geq c\|u\|^p$ for all u in $\mathbb{R}^N$.*

Let $x : [0, +\infty[\to \mathbb{R}^N$ be a bounded solution of

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

Then

(i) *$x(t) \to x_\infty$ as $t \to +\infty$, where $x_\infty \in$* crit *f;*

(ii) $\dot{x} \in L^1(0, +\infty; \mathbb{R}^N)$ *and* $\dot{x}(t) \to 0$ *as* $t \to +\infty$*;*

(iii) *for* λ *sufficiently small and* t *sufficiently large,*

$$\|x(t) - x_\infty\| \leq \frac{1}{\alpha}\theta(E_\lambda(x(t), u(t)) - E_\lambda(x_\infty, 0))$$

where θ *is the desingularizing function for* E_λ *at* $(x_\infty, 0)$*, and* α *enters the angle condition.*

9.4. Numerical illustrations

We revisit the numerical examples of Section 5.3 where we introduce additional Hessian-drive damping.

So, we take $\mathcal{H} = \mathbb{R}$, $f(x) = \frac{1}{2}|x|^2$, and $\phi(u) = \frac{1}{p}|u|^p$ with $p > 1$. Then equation (ADIGE-VH) reads

$$\ddot{x}(t) + |\dot{x}(t)|^{p-2}\dot{x}(t) + \beta\dot{x}(t) + x(t) = 0. \tag{84}$$

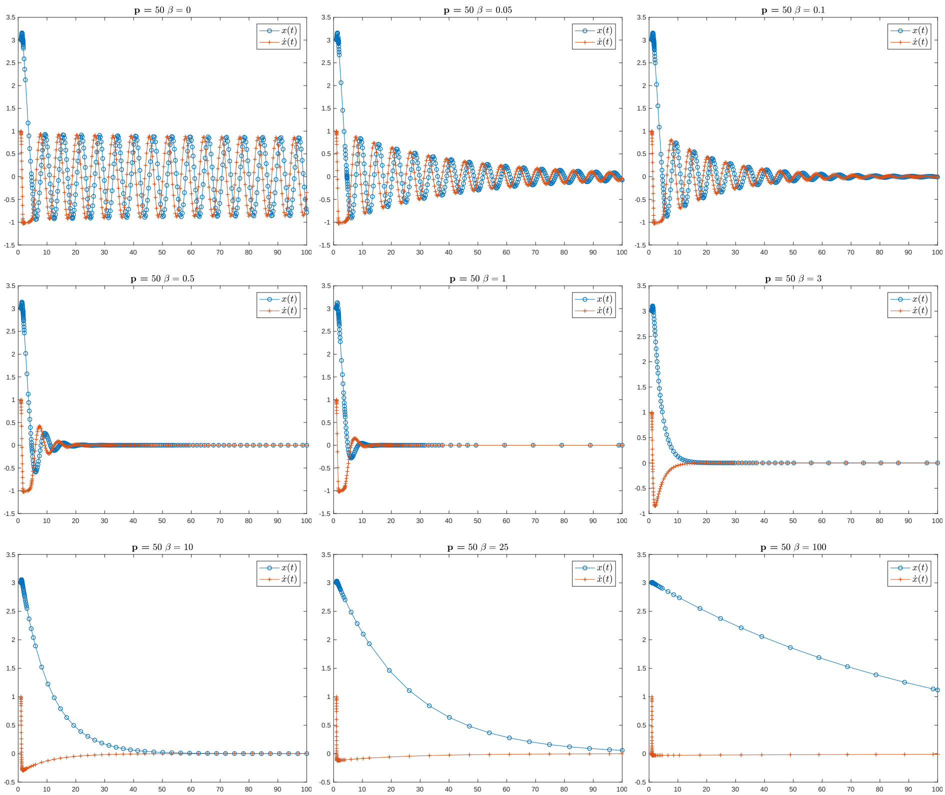

Fig. 6. Evolution of the trajectories $x(t)$ (blue) and $\dot{x}(t)$ (red) of (84) for different values of β.

For $\beta > 0$, we are in the framework of Theorem 4 with $\phi(u) = \frac{\beta}{2}|u|^2 + \frac{p}{2}|u|^p$. So we have convergence at an exponential rate of $x(t)$ and $\dot{x}(t)$ to zero. This makes a big contrast with the case $\beta = 0$, for which we have convergence to zero, but with many oscillations in the case of weak damping (p large). Note that even for very small $\beta > 0$, we have a rapid stabilization of the trajectory towards the origin. On the other hand, taking large β is not beneficial: we can observe in Figure 6 that the convergence deteriorates in this case. Indeed, since the damping attached to $|\dot{x}(t)|^{p-2}\dot{x}(t)$ is negligible for large p compared to the damping attached to $\beta\dot{x}(t)$, the "optimal" value of β is close to the optimal value for (HBF). So, by Theorem 2, it is close to $2\sqrt{\mu}$ where μ is the coefficient of strong convexity of f (see Theorem 2). In our situation, this gives $\beta \sim 2$.

9.5. Link with the regularized Newton method

Let us specify the link between our study and Newton's method for solving $Ax \ni 0$, where A is a general maximally monotone operator (for convex minimization take $A = \partial f$). To overcome the ill-posedness of the continuous Newton method, the following first-order evolution system was studied by Attouch–Svaiter [37]:

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

The system can be considered as a continuous version of the Levenberg–Marquardt system, which acts as a regularization of the Newton method. Under a fairly general assumption on the regularization parameter $\gamma(t)$, this system is well-posed and generates trajectories that converge weakly to equilibria. Parallel results have been obtained for the associated proximal algorithms obtained by implicit temporal discretization [1,32,36]. Formally, when A is differentiable, this system reads $\gamma(t)\dot{x}(t) + \beta\frac{d}{dt}(A(x(t))) + A(x(t)) = 0$. When $A = \nabla f$ we obtain

$$\gamma(t)\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0. \tag{85}$$

The system (ADIGE-VH) considered in the previous section can be seen as an inertial version of (85). Most interestingly, Attouch–Redont–Svaiter developed in [36] a closed-loop version of the above results. They showed the convergence of the trajectories generated by the closed-loop control system when $0 < p < 1$, where A is a general maximally monotone operator:

$$\begin{cases} v(t) \in A(x(t)), \\ \|v(t)\|^p\dot{x}(t) + \dot{v}(t) + v(t) = 0, \\ x(0) = x_0, \quad v(0) \in A(x_0), \quad v_0 \neq 0. \end{cases}.$$

For optimization problems, this naturally suggests considering autonomous inertial systems where the damping coefficient yields closed-loop control of the gradient of f.

A first answer to this question has been obtained by Lin–Jordan [79] who considered the autonomous system

$$\ddot{x}(t)+\gamma(t)\dot{x}(t)+\beta(t)\nabla^2 f(x(t))\dot{x}(t)+b(t)\nabla f(x(t))=0, \tag{86}$$

where γ, β and b are defined by the following formulas:.

$$\begin{cases} |\lambda(t)|^p\|\nabla f(x(t))\|^{p-1}=\theta, \\ a(t)=\frac{1}{4}(\int_0^t\sqrt{\lambda(s)}\,ds+c)^2, \\ \gamma(t)=2\frac{\dot{a}(t)}{a(t)}-\frac{\ddot{a}(t)}{\dot{a}(t)}, \\ \beta(t)=\left(\frac{\dot{a}(t)}{a(t)}\right)^2, \\ b(t)=\frac{\dot{a}(t)(\dot{a}(t)+\ddot{a}(t))}{a(t)}. \end{cases} \tag{87}$$

As a specific feature, the damping coefficients are expressed with the help of $\lambda(t)$ which is equal to a power of the inverse of the norm of the gradient of f. The authors give some interesting nontrivial convergence rates of values. Owing to the presence of the Hessian-driven damping term, they show fast convergence to zero of gradient norms.

10. Closed-loop damping involving the velocity and the gradient

Let us consider the following system, where the damping term $\partial\phi(\dot{x}(t)+\beta\nabla f(x(t))$ involves both the velocity vector and the gradient of the potential function f:

$$\text{(ADIGE-VGH)}\quad \ddot{x}(t)+\partial\phi\big(\dot{x}(t)+\beta\nabla f(x(t))\big)+\beta\nabla^2 f(x(t))\dot{x}(t)+\nabla f(x(t))\ni 0. \tag{88}$$

The parameter $\beta\geq 0$ is attached to the geometric damping induced by the Hessian. As previously, ϕ is a damping potential function. The suffixes V,G,H make respectively reference to the Velocity, the Gradient of f, and the Hessian of f, which enter the damping terms of the above dynamic. This model makes it possible to encompass several situations.

- When $\beta=0$, we recover the closed loop controled system

$$\ddot{x}(t)+\partial\phi(\dot{x}(t))+\nabla f(x(t))=0, \tag{89}$$

studied in Sections 3 to 7. So studying (88) can be viewed as an extension of our previous study. Still, we will see that taking $\beta>0$ yields several favorable properties.

- When $\phi(u)=\frac{\gamma}{2}\|u\|^2$, we obtain the system

$$\ddot{x}(t)+\gamma\dot{x}(t)+\beta\nabla^2 f(x(t))\dot{x}(t)+(1+\gamma\beta)\nabla f(x(t))=0, \tag{90}$$

studied in Section 9, and which was introduced by Alvarez–Attouch–Bolte–Redont in [11].

10.1. Existence and uniqueness results

A key property for studying (88) is the following equivalent formulation, different from the Hamiltonian formulation, and whose proof is immediate. Just introduce the new variable $u(t) := \dot{x}(t) + \beta\nabla f(x(t))$.

Proposition 6. *The following are equivalent:*

(i) $\ddot{x}(t) + \partial\phi(\dot{x}(t) + \beta\nabla f(x(t))) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0$.

(ii) $\begin{cases} \dot{x}(t) + \beta\nabla f(x(t)) - u(t) = 0, \\ \dot{u}(t) + \partial\phi(u(t)) + \nabla f(x(t)) \ni 0. \end{cases}$

A major interest of the formulation (ii) is that it is a first-order system in time and space (without occurrence of the Hessian). As such, it requires fewer regularity assumptions on f than in Theorem 10.

Theorem 12. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex $\mathcal{C}^2$ function such that $\inf_{\mathcal{H}} f > -\infty$. Suppose that ∇f is Lipschitz continuous on bounded subsets of $\mathcal{H}$. Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous damping function. Then, for any Cauchy data $(x_0, x_1) \in \mathcal{H} \times \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\to \mathcal{H}$ of* (ADIGE-VGH) *satisfying $x(0) = x_0$ and $\dot{x}(0) = x_1$.*

Proof. The structure of the proof being similar to Theorem 10, we only develop the original aspects.

Step 1: *A priori estimate.* Note that (88) can be equivalently written as

$$\frac{d}{dt}\big(\dot{x}(t) + \beta\nabla f(x(t))\big) + \partial\phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big) + \nabla f(x(t)) \ni 0. \tag{91}$$

After taking the scalar product of (91) with $\dot{x}(t) + \beta\nabla f(x(t))$, we get

$$\frac{1}{2}\frac{d}{dt}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + \big\langle\partial\phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big), \dot{x}(t) + \beta\nabla f(x(t))\big\rangle \\ + \langle\nabla f(x(t)), \dot{x}(t) + \beta\nabla f(x(t))\rangle = 0. \tag{92}$$

Since ϕ is a damping potential, the subdifferential inequality for convex functions gives

$$\big\langle\partial\phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big), \dot{x}(t) + \beta\nabla f(x(t))\big\rangle \geq \phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big).$$

Collecting the above results, we obtain

$$\frac{d}{dt}\big(\tfrac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + f(x(t)) - \inf_{\mathcal{H}} f\big) \\ + \phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big) + \beta\|\nabla f(x(t))\|^2 \leq 0. \tag{93}$$

Therefore, the energy-like function

$$t \mapsto \tfrac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + f(x(t)) - \inf_{\mathcal{H}} f \quad \text{is nonincreasing.} \tag{94}$$

This implies that, as long as the trajectory is defined,

$$\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 \leq C := \|x_1 + \beta\nabla f(x_0)\|^2 + 2(f(x_0) - \inf_{\mathcal{H}} f). \tag{95}$$

From this, we will obtain a bound on the trajectory. We have

$$\dot{x}(t) + \beta\nabla f(x(t)) = k(t)$$

with $\|k(t)\| \leq \sqrt{C}$. Take the scalar product of the above equation with $x(t) - x_0$:

$$\frac{1}{2}\frac{d}{dt}\|x(t) - x_0\|^2 + \beta\langle\nabla f(x(t)) - \nabla f(x_0), x(t) - x_0\rangle + \beta\langle\nabla f(x_0), x(t) - x_0\rangle$$
$$= \langle k(t), x(t) - x_0\rangle.$$

By the convexity of f, and hence the monotonicity of ∇f, and by the Cauchy–Schwarz inequality,

$$\frac{1}{2}\frac{d}{dt}\|x(t) - x_0\|^2 \leq (\|k(t)\| + \beta\|\nabla f(x_0)\|)\|x(t) - x_0\|. \tag{96}$$

According to the Gronwall inequality, and $\|k(t)\| \leq \sqrt{C}$, we obtain

$$\|x(t) - x_0\| \leq t\big(\|x_1 + \beta\nabla f(x_0)\| + \sqrt{2(f(x_0) - \inf_{\mathcal{H}} f)} + \beta\|\nabla f(x_0)\|\big). \tag{97}$$

Step 2: *Frst-order formulation of* (88). According to Proposition 6, it is equivalent to solve the first-order system

$$\begin{cases} \dot{x}(t) + \beta\nabla f(x(t)) - u(t) = 0, \\ \dot{u}(t) + \partial\phi(u(t)) + \nabla f(x(t)) \ni 0, \end{cases}$$

with the Cauchy data $x(0) = x_0, u(0) = x_1$. Set $Z(t) = (x(t), u(t)) \in \mathcal{H} \times \mathcal{H}$. The above system can be written equivalently as

$$\dot{Z}(t) + F(Z(t)) \ni 0, \quad Z(0) = (x_0, x_1),$$

where $F : \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}$, $(x, u) \mapsto F(x, u)$, is defined by

$$F(x, u) = (0, \partial\phi(u)) + (\beta\nabla f(x) - u, \nabla f(x)).$$

Hence F splits as $F(x, u) = \partial\Phi(x, u) + G(x, u)$, where

$$\Phi(x, u) = \phi(u) \quad \text{and} \quad G(x, u) = (\beta\nabla f(x) - u, \nabla f(x)). \tag{98}$$

Therefore, it is equivalent to solve the following first-order differential inclusion with Cauchy data:

$$\dot{Z}(t) + \partial\Phi(Z(t)) + G(Z(t)) \ni 0, \quad Z(0) = (x_0, x_1). \tag{99}$$

By the local Lipschitz assumption on the gradient of f, we immediately see that $(x, u) \mapsto G(x, u)$ is Lipschitz continuous on bounded subsets of $\mathcal{H} \times \mathcal{H}$.

Step 3: *Approximate dynamics.* We consider the approximate dynamics

$$\ddot{x}_\lambda(t) + \nabla\phi_\lambda\big(\dot{x}_\lambda(t) + \beta\nabla f(x_\lambda(t))\big) \\ + \beta\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t) + \nabla f(x_\lambda(t)) = 0, \quad t \in [0, +\infty[, \tag{100}$$

which uses the Moreau–Yosida approximations (ϕ_λ) of ϕ. We will prove that the filtered sequence (x_λ) converges uniformly as $\lambda \to 0$ over bounded time intervals to a solution of (88). The first-order formulation of (100) gives the system

$$\begin{cases} \dot{x}_\lambda(t) + \beta\nabla f(x_\lambda(t)) - u_\lambda(t) = 0, \\ \dot{u}_\lambda(t) + \nabla\phi_\lambda(u_\lambda(t)) + \nabla f(x_\lambda(t)) = 0, \end{cases}$$

with the Cauchy data $x_\lambda(0) = x_0$, $u_\lambda(0) = x_1$. Set $Z_\lambda(t) = (x_\lambda(t), u_\lambda(t)) \in \mathcal{H} \times \mathcal{H}$. The above system can be written equivalently as

$$\dot{Z}_\lambda(t) + F_\lambda(Z_\lambda(t)) \ni 0, \quad Z_\lambda(t_0) = (x_0, x_1),$$

where $F_\lambda : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$, $(x, u) \mapsto F_\lambda(x, u)$, is defined by

$$F_\lambda(x, u) = (0, \nabla\phi_\lambda(u)) + (\beta\nabla f(x) - u, \nabla f(x)).$$

Hence $F_\lambda(x, u) = \nabla\Phi_\lambda(x, u) + G(x, u)$ where Φ and G have been defined in (98). Therefore, the approximate equation is equivalent to the first-order differential system with Cauchy data

$$\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1). \tag{101}$$

Fix $\lambda > 0$ fixed. By the Lipschitz continuity of $\nabla\Phi_\lambda$, and the fact that G is Lipschitz continuous on bounded sets, the sum operator $\nabla\Phi_\lambda + G$ which governs (101) is Lipschitz continuous on bounded sets. As a consequence, the existence of a local solution to (101) follows from the Cauchy–Lipschitz theorem. To pass from a local solution to a global solution, we use the a priori estimates (95) and (97) obtained in Step 1 of the proof. Note that these estimates are valid for any damping potential, in particular for ϕ_λ. Suppose that a maximal solution is defined on a finite time interval $[0, T[$. From (97) we first find that $x_\lambda(t)$ remains bounded on $[0, T[$. Then, using (95) and the fact that the gradient of f is Lipschitz continuous on the bounded sets, we infer that $\dot{x}_\lambda(t)$ is also bounded on $[0, T[$. By the property (9) of Yosida approximation, the property (iii) of the damping potential ϕ, and (95), this implies that

$$\big\|\nabla\phi_\lambda\big(\dot{x}_\lambda(t) + \beta\nabla f(x_\lambda(t))\big)\big\| \leq \big\|(\partial\phi)^0\big(\dot{x}_\lambda(t) + \beta\nabla f(x_\lambda(t))\big)\big\|$$

is also bounded on $[0, T[$. Moreover, by the local boundedness assumption on the gradient, and the boundedness of $x_\lambda(t)$ and $\dot{x}_\lambda(t)$, we deduce that $\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)$ is also bounded. The constitutive equation (76) then implies that $(\ddot{x}_\lambda)$ is also bounded. Hence the limits of $x_\lambda(t)$ and $\dot{x}_\lambda(t)$ as $t \to T$ exist. Passing from a local to a global solution is then a classical argument. So, for any $\lambda > 0$ we have a unique global solution of (76) with Cauchy data $x_\lambda(0) = x_0$, $\dot{x}_\lambda(0) = x_1$.

Step 4: *Passing to the limit as* $\lambda \to 0$. Take $T > 0$ and $\lambda, \mu > 0$. Consider the corresponding solutions on $[0, T]$:

$$\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1),$$
$$\dot{Z}_\mu(t) + \nabla\Phi_\mu(Z_\mu(t)) + G(Z_\mu(t)) = 0, \quad Z_\mu(0) = (x_0, x_1).$$

Subtracting these equations, and taking the scalar product with $Z_\lambda(t) - Z_\mu(t)$, we get

$$\frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 + \langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle$$
$$+ \langle G(Z_\lambda(t)) - G(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle = 0. \tag{102}$$

We now use the following ingredients:

(i) By the general properties of Yosida approximation [53, Theorem 3.1], we have

$$\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle$$
$$\geq -\frac{\lambda}{4}\|\nabla\Phi_\mu(Z_\mu(t))\|^2 - \frac{\mu}{4}\|\nabla\Phi_\lambda(Z_\lambda(t))\|^2.$$

According to the energy estimates, the sequence (Z_λ) is uniformly bounded on $[0, T]$; let

$$\|Z_\lambda(t)\| \leq C_T.$$

From these properties we immediately infer that

$$\|\nabla\Phi_\lambda(Z_\lambda(t))\| \leq \sup_{\|\xi\|\leq C_T} \|(\partial\phi)^0(\xi)\| = M_T < +\infty$$

because our assumption on ϕ gives that $(\partial\phi)^0$ is bounded on bounded sets. Therefore

$$\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \geq -\tfrac{1}{4}M_T(\lambda + \mu).$$

(ii) Since $G : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is Lipschitz continuous on bounded sets, and (Z_λ) is uniformly bounded on $[0, T]$, we deduce that there exists a constant L_T such that

$$\|G(Z_\lambda(t)) - G(Z_\mu(t))\| \leq L_T\|Z_\lambda(t) - Z_\mu(t)\|.$$

Combining the above results, and using the Cauchy–Schwarz inequality, we deduce from (78) that

$$\frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{1}{4}M_T(\lambda + \mu) + L_T\|Z_\lambda(t) - Z_\mu(t)\|^2.$$

We now integrate this differential inequality. Elementary calculus using $Z_\lambda(0) - Z_\mu(0) = 0$ gives

$$\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{M_T}{4L_T}(\lambda + \mu)(e^{2L_T(t-t_0)} - 1).$$

Therefore, the filtered sequence (Z_λ) is a Cauchy sequence for uniform convergence on $[0, T]$, and hence it converges uniformly. This means that x_λ and $\dot{x}_\lambda$ converge uniformly on $[0, T]$ to x and $\dot{x}$ respectively. That x is solution of (88) is proved as in Theorem 3. Just rely on the property $\frac{d}{dt}(\nabla f(x_\lambda(t))) = \nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)$ to pass to the limit in the Hessian term. ∎

10.2. Convergence properties

We have the following convergence properties of trajectories of the system (88) with closed-loop damping involving both the velocity and the gradient.

Theorem 13. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex $\mathcal{C}^2$ function such that* $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$. *Suppose that ∇f is Lipschitz continuous on bounded subsets of $\mathcal{H}$. Let $\beta > 0$. Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous damping function. Then, for any solution $x : [0, +\infty[\to \mathcal{H}$ of* (ADIGE-VGH) *we have*

(i) *the energy-like function $t \mapsto \frac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + f(x(t))$ is nonincreasing;*

(ii) $\int_0^{+\infty} \phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big)\, dt < +\infty$;

(iii) $\int_0^{+\infty} \|\nabla f(x(t))\|^2\, dt < +\infty$.

Suppose moreover that there exists $r > 0$ such that $\phi(u) \geq r\|u\|$ for all $u \in \mathcal{H}$. Then

(a) *the trajectory $x(t)$ converges weakly as $t \to +\infty$, and its limit belongs to* $\operatorname{argmin}_{\mathcal{H}} f$;

(b) *$\dot{x}(t)$ and $\nabla f(x(t))$ converge strongly to zero as $t \to +\infty$.*

Proof. Items (i) to (iii) are direct consequences of the estimate (93) established in Step 1 of the proof of Theorem 12.

Let us now make the additional assumption $\phi(u) \geq r\|u\|$. By (ii),

$$\int_0^{+\infty} \|\dot{x}(t) + \beta\nabla f(x(t))\|\, dt \leq \frac{1}{r}\int_0^{+\infty} \phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big)\, dt < +\infty.$$

Therefore, $x(\cdot)$ is a solution of the nonautonomous steepest descent equation

$$\dot{x}(t) + \beta\nabla f(x(t)) = k(t)$$

with $k \in L^1(0, +\infty; \mathcal{H})$. Theorem 3.11 of [53] gives the convergence of the trajectory to a point in $\operatorname{argmin}_{\mathcal{H}} f$. In particular, the trajectory remains bounded. In view of (i), $\dot{x}(t)$ is also bounded. Returning to the constitutive equation (88), we deduce that the acceleration $\ddot{x}(t)$ is also bounded. This implies that $\xi(t) = \dot{x}(t) + \beta\nabla f(x(t))$ satisfies

$$\int_0^{+\infty} \|\xi(t)\|\, dt < +\infty \quad \text{and} \quad \|\dot{\xi}(t)\| \leq M$$

for some $M > 0$. This classically implies that $\xi(t) = \dot{x}(t) + \beta\nabla f(x(t))$ tends to zero as $t \to +\infty$. From (iii), the same argument applied to $\nabla f(x(t))$ shows that $\nabla f(x(t))$ tends to zero as $t \to +\infty$. As the difference of the previous two quantities, $\dot{x}(t)$ tends to zero as $t \to +\infty$. ∎

Indeed, Theorem 3.11 of [53] was proved under the additional assumption that f is inf-compact. Recent progress based on the Opial lemma [83] and the Bruck theorem [54] allows one to extend it to general convex functions f, without making this additional assumption. This is made precise below.

Proposition 7. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function such that* $\operatorname{argmin}_{\mathcal{H}} f \neq \emptyset$*, and let $k \in L^1(0,+\infty;\mathcal{H})$. Suppose that $x : [0,+\infty[\to \mathcal{H}$ is a strong global solution of*

$$\dot{x}(t) + \partial f(x(t)) \ni k(t).$$

Then the trajectory $x(t)$ converges weakly as $t \to +\infty$, and its limit belongs to $\operatorname{argmin}_{\mathcal{H}} f$.

Proof. Take $\epsilon > 0$. Since $k \in L^1(0,+\infty;\mathcal{H})$, there exists $T_\epsilon > 0$ such that $\int_{T_\epsilon}^{+\infty} \|k(t)\|\, dt < \epsilon$. Consider the solution $v : [0,+\infty[\to \mathcal{H}$ of

$$\dot{v}(t) + \nabla f(v(t)) \ni 0, \quad v(0) = x(T_\epsilon).$$

By properties of contraction semigroups we have, for all $t \geq T_\epsilon$,

$$\|x(t) - v(t - T_\epsilon)\| \leq \|x(T_\epsilon) - v(0)\| + \int_{T_\epsilon}^{t} \|k(t)\|\, dt \leq \epsilon. \tag{103}$$

Take $\xi \in \mathcal{H}$. By the Cauchy–Schwarz inequality,

$$|\langle x(t) - v(t - T_\epsilon), \xi\rangle| \leq \epsilon \|\xi\|.$$

By the triangle inequality, we deduce that, for all $t, t' \geq T_\epsilon$,

$$|\langle x(t), \xi\rangle - \langle x(t'), \xi\rangle| \leq |\langle v(t - T_\epsilon) - v(t' - T_\epsilon), \xi\rangle| + 2\epsilon\|\xi\|.$$

According to the Bruck theorem, we know that the weak limit of $v(t)$ exists. Passing to the limsup on the above inequality we get

$$\limsup_{t,t' \to +\infty} |\langle x(t), \xi\rangle - \langle x(t'), \xi\rangle| \leq \limsup_{t,t' \to +\infty} |\langle v(t - T_\epsilon) - v(t' - T_\epsilon), \xi\rangle| + 2\epsilon\|\xi\| \leq 2\epsilon\|\xi\|.$$

This being true for any $\epsilon > 0$, we deduce that the limit of $\langle x(t), \xi\rangle$ exists, which implies that the weak limit of $x(t)$ exists as $t \to +\infty$; let x_∞ be that limit. Passing to the lower limit in (103), by the lower semicontinuity of the norm for the weak topology, we deduce that

$$\left\| x_\infty - \lim_{t \to +\infty} v(t) \right\| \leq \epsilon. \tag{104}$$

Since the weak limit of $v(t)$ belongs to $\operatorname{argmin}_{\mathcal{H}} f$, we deduce that $\operatorname{dist}(x_\infty, \operatorname{argmin}_{\mathcal{H}} f) \leq \epsilon$. This being true for any $\epsilon > 0$, and since $\operatorname{argmin}_{\mathcal{H}} f$ is closed, we conclude that $x_\infty \in \operatorname{argmin}_{\mathcal{H}} f$. ∎

10.3. An approach based on Opial's lemma

Here we will prove the weak convergence of the trajectory x to a minimizer of f, based on the continuous version of the Opial lemma [83]. As in the proof of Theorem 13, items (i) to (iii) hold. Assume $\phi(u) \geq r\|u\|$ for all $u \in \mathcal{H}$. According to item (ii) we obtain

$$\int_0^{+\infty} \|\dot{x}(t) + \beta \nabla f(x(t))\| \, dt < +\infty.$$

Equivalently, we have

$$\dot{x}(t) + \beta \nabla f(x(t)) = k(t)$$

with $k \in L^1(0, +\infty; \mathcal{H})$. Let us prove that x is bounded. Relying on Step 1 of the proof of Theorem 12, notice that (96) holds for a generic $x_0 \in \mathcal{H}$. Taking an arbitrary $z \in \operatorname{argmin} f$, we deduce from (96) that

$$\frac{1}{2}\frac{d}{dt}\|x(t) - z\|^2 \leq \|k(t)\| \cdot \|x(t) - z\|. \tag{105}$$

Integrating we obtain

$$\frac{1}{2}\|x(T) - z\|^2 \leq \frac{1}{2}\|x_0 - z\|^2 + \int_0^T \|k(t)\| \cdot \|x(t) - z\| \, dt \quad \forall T \geq 0. \tag{106}$$

Now apply [53, Lemme A.5, p. 157] to conclude that

$$\|x(T) - z\| \leq \|x_0 - z\| + \int_0^T \|k(t)\| \, dt \quad \forall T \geq 0.$$

Since $k \in L^1(0, +\infty; \mathcal{H})$ we see that x is bounded. Now we can repeat the arguments in the proof of Theorem 13 to conclude that $\lim_{t\to\infty} \dot{x}(t) = \lim_{t\to\infty} \nabla f(x(t)) = 0$, so we omit the proof. Let us pass forward and see how the Opial lemma [83] can be applied.

Since x is bounded and $k \in L^1(0, +\infty; \mathcal{H})$, we see from (105) that $\lim_{t\to\infty} \|x(t) - z\|$ exists, hence the first condition in the Opial lemma is fulfilled. To check the second condition in the Opial lemma is standard. Take $\overline{x} \in \mathcal{H}$ and $t_n \to +\infty$ such that $x(t_n)$ converges weakly to $\overline{x}$ as $n \to +\infty$. The convexity of f yields, for all $x \in \mathcal{H}$ and all $n \in \mathbb{N}$,

$$f(x) \geq f(x(t_n)) + \langle \nabla f(x(t_n)), x - x(t_n) \rangle.$$

Fixing x and taking the limit as $n \to +\infty$, and relying on the strong convergence of $\nabla f(x(t))$ to 0 and the boundedness of x, we derive

$$f(x) \geq \liminf_{n\to+\infty} f(x(t_n)) \geq f(\overline{x}),$$

where the last inequality follows from the weak lower semicontinuity of the convex function f. Since the last inequality holds for an arbitrary x, we obtain $\overline{x} \in \operatorname{argmin} f$. Therefore, the second condition in the Opial lemma is fulfilled as well.

10.4. A finite stabilization property

As already mentioned, (ADIGE-VGH) can be equivalently written as

$$\dot{u}(t) + \partial\phi(u(t)) \ni -\nabla f(x(t))$$

where $u(t) = \dot{x}(t) + \beta\nabla f(x(t))$. After taking the scalar product of the above equation with $u(t)$, we get

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|^2 + \langle \partial\phi(u(t)), u(t)\rangle = -\langle \nabla f(x(t)), u(t)\rangle.$$

When $\phi(u) \geq r\|u\|$, by the Cauchy–Schwarz inequality we get

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|^2 + r\|u(t)\| \leq \|\nabla f(x(t))\|\,\|u(t)\|.$$

Since $\nabla f(x(t))$ converges strongly to zero as $t \to +\infty$ (that is the last point of Theorem 13), for t large enough we get $\|\nabla f(x(t))\| \leq \frac{1}{2}r$, and hence

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|^2 + \frac{1}{2}r\|u(t)\| \leq 0.$$

This yields $u(t) \equiv 0$ after a finite time. Let us summarize the above results in the following proposition.

Proposition 8. *Under the hypothesis of Theorem 13, and when $\phi(u) \geq r\|u\|$ for some $r > 0$, after a finite time we have*

$$\dot{x}(t) + \beta\nabla f(x(t)) \equiv 0,$$

i.e. the trajectory follows the steepest descent dynamic.

10.5. The case of f strongly convex: exponential convergence rate

Theorem 14. *Let $f : \mathcal{H} \to \mathbb{R}$ be a γ-strongly convex $\mathcal{C}^2$ function (for some $\gamma > 0$) whose gradient is Lipschitz continuous on bounded sets. Let $\overline{x}$ be the unique minimizer of f. Let $\phi : \mathcal{H} \to \mathbb{R}_+$ be a damping potential which is differentiable, and whose gradient is Lipschitz continuous on bounded subsets of $\mathcal{H}$. Suppose that ϕ satisfies the following growth conditions:*

(i) (local) *there exist constants $\alpha, \epsilon > 0$ such that, for all u in $\mathcal{H}$ with $\|u\| \leq \epsilon$,*

$$\langle \nabla\phi(u), u\rangle \geq \alpha\|u\|^2.$$

(ii) (global) *there exist $p \geq 1$ and $r > 0$ such that $\phi(u) \geq r\|u\|^p$ for all u in $\mathcal{H}$.*

Let $\beta > 0$. Let $x : [0, +\infty[\to \mathcal{H}$ be a solution of (ADIGE-VGH)

$$\ddot{x}(t) + \nabla\phi(\dot{x}(t) + \beta\nabla f(x(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0. \tag{107}$$

Then we have exponential convergence rate to zero as $t \to +\infty$ for $f(x(t)) - f(\overline{x})$, $\|x(t) - \overline{x}\|$ and $\|\dot{x}(t) + \beta\nabla f(x(t))\|$. As a consequence, we also have exponential convergence rate to zero for $\|\dot{x}(t)\|$ and $\|\nabla f(x(t))\|$.

Proof. Since f is strongly convex, f is a coercive function. By the decrease property of the global energy (see (94) and Theorem 13 (i)), $f(x(t))$ is bounded from above, and hence the trajectory x is bounded. Theorem 13 (ii) and the global growth assumption on ϕ imply that, for some $p \geq 1$,

$$\int_0^{+\infty} \|\dot{x}(t) + \beta\nabla f(x(t))\|^p \, dt < +\infty.$$

By a similar argument to the proof of Theorem 13 (where we argued with $p = 1$) we deduce that $\lim_{t\to+\infty} \|\dot{x}(t) + \beta\nabla f(x(t))\| = 0$. Therefore, for t sufficiently large,

$$\|\dot{x}(t) + \beta\nabla f(x(t))\| \leq \epsilon.$$

From (92) and the local property (i) we derive

$$\begin{aligned} &\frac{d}{dt}\Big(\tfrac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + f(x(t)) - f(\overline{x})\Big) \\ &\qquad + \alpha\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + \beta\|\nabla f(x(t))\|^2 \leq 0. \end{aligned} \tag{108}$$

Since $\dot{x}(\cdot) + \beta\nabla f(x(\cdot))$ is bounded, let $L > 0$ be the Lipschitz constant of $\nabla\phi$ on a ball that contains the vectors $\dot{x}(t) + \beta\nabla f(x(t))$ for all $t \geq 0$. Since $\nabla\phi(0) = 0$ we have, for all $t \geq 0$,

$$\|\nabla\phi(\dot{x}(t)) + \beta\nabla f(x(t))\| \leq L\|\dot{x}(t) + \beta\nabla f(x(t))\|. \tag{109}$$

Using successively (107), (109) and (32), we obtain

$$\begin{aligned} &\frac{d}{dt}\langle x(t) - \overline{x}, \dot{x}(t) + \beta\nabla f(x(t))\rangle \\ &\quad = \|\dot{x}(t)\|^2 + \beta\frac{d}{dt}(f(x(t)) - f(\overline{x})) \\ &\qquad + \langle x(t) - \overline{x}, -\nabla\phi(\dot{x}(t) + \beta\nabla f(x(t))) - \nabla f(x(t))\rangle \\ &\quad \leq \|\dot{x}(t)\|^2 + \beta\frac{d}{dt}(f(x(t)) - f(\overline{x})) + \frac{L^2}{2\gamma}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 \\ &\qquad + \frac{\gamma}{2}\|x(t) - \overline{x}\|^2 + \langle \overline{x} - x(t), \nabla f(x(t))\rangle \\ &\quad \leq \|\dot{x}(t)\|^2 + \beta\frac{d}{dt}(f(x(t)) - f(\overline{x})) + \frac{L^2}{2\gamma}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 \\ &\qquad + f(\overline{x}) - f(x(t)). \end{aligned} \tag{110}$$

Take now $\varepsilon > 0$ (to be specified below), and define

$$\begin{aligned} h_{\varepsilon,\beta}(t) := {}& \tfrac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + (1 - \beta\varepsilon)\big(f(x(t)) - f(\overline{x})\big) \\ & + \varepsilon\langle x(t) - \overline{x}, \dot{x}(t) + \beta\nabla f(x(t))\rangle. \end{aligned}$$

Multiplying (110) by ε and adding the result to (108), we derive

$$\dot{h}_{\varepsilon,\beta}(t) \le -\alpha\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 - \beta\|\nabla f(x(t))\|^2 + \varepsilon\|\dot{x}(t)\|^2 + \frac{\varepsilon L^2}{2\gamma}\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 - \varepsilon(f(x(t)) - f(\overline{x})).$$

We use the inequality

$$\varepsilon\|\dot{x}(t)\|^2 \le 2\varepsilon\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 + 2\varepsilon\beta^2\|\nabla f(x(t))\|^2 \tag{111}$$

to obtain

$$\dot{h}_{\varepsilon,\beta}(t) \le -\left(\alpha - 2\varepsilon - \frac{\varepsilon L^2}{2\gamma}\right)\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 - (\beta - 2\varepsilon\beta^2)\|\nabla f(x(t))\|^2 - \varepsilon(f(x(t)) - f(\overline{x})). \tag{112}$$

Choose $\varepsilon > 0$ small enough such that $C_1 := \min\{\alpha - 2\varepsilon - \frac{\varepsilon L^2}{2\gamma}, \beta - 2\varepsilon\beta^2, \varepsilon\} > 0$. We obtain

$$\dot{h}_{\varepsilon,\beta}(t) \le -C_1\big(\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 + \|\nabla f(x(t))\|^2 + f(x(t)) - f(\overline{x})\big). \tag{113}$$

Further, we have

$$h_{\varepsilon,\beta}(t) = \tfrac{1}{2}\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 + \varepsilon\beta\big(\langle x(t)-\overline{x}, \nabla f(x(t))\rangle + f(\overline{x}) - f(x(t))\big) + f(x(t)) - f(\overline{x}) + \varepsilon\langle x(t)-\overline{x}, \dot{x}(t)\rangle. \tag{114}$$

Since f is strongly convex, we have (see for example [80, Theorem 2.1.10])

$$\langle x(t)-\overline{x}, \nabla f(x(t))\rangle + f(\overline{x}) - f(x(t)) \le \frac{1}{2\gamma}\|\nabla f(x(t))\|^2. \tag{115}$$

Moreover, from (33) and (111) we get

$$\begin{aligned} f(x(t)) - f(\overline{x}) + \varepsilon\langle x(t)-\overline{x}, \dot{x}(t)\rangle &\le f(x(t)) - f(\overline{x}) + \frac{\varepsilon}{2}\|x(t)-\overline{x}\|^2 + \frac{\varepsilon}{2}\|\dot{x}(t)\|^2 \\ &\le \left(1+\frac{\varepsilon}{\gamma}\right)\big(f(x(t)) - f(\overline{x})\big) + \varepsilon\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 + \varepsilon\beta^2\|\nabla f(x(t))\|^2. \end{aligned}$$

From this, (115) and (114) we get

$$\begin{aligned} h_{\varepsilon,\beta}(t) &\le \left(\frac{1}{2}+\varepsilon\right)\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 + \left(\frac{\varepsilon\beta}{2\gamma}+\varepsilon\beta^2\right)\|\nabla f(x(t))\|^2 \\ &\quad + \left(1+\frac{\varepsilon}{\gamma}\right)\big(f(x(t)) - f(\overline{x})\big) \\ &\le C_2(\|\dot{x}(t)+\beta\nabla f(x(t))\|^2 + \|\nabla f(x(t))\|^2 + f(x(t)) - f(\overline{x})), \end{aligned}$$

where $C_2 := \max\{\frac{1}{2}+\varepsilon, \frac{\varepsilon\beta}{2\gamma}+\varepsilon\beta^2, 1+\frac{\varepsilon}{\gamma}\} > 0$. Combining this inequality with (113) yields

$$\dot{h}_{\varepsilon,\beta}(t) + C_3 h_{\varepsilon,\beta}(t) \le 0,$$

with $C_3 := \frac{C_1}{C_2} > 0$. Then, the Gronwall inequality implies

$$h_{\varepsilon,\beta}(t) \leq h_{\varepsilon,\beta}(0)e^{-C_3 t}. \tag{116}$$

Finally, from (33) and the Cauchy–Schwarz inequality we have

$$\begin{aligned} h_{\varepsilon,\beta}(t) &\geq \frac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + (1-\beta\varepsilon)\big(f(x(t)) - f(\overline{x})\big) \\ &\quad - \frac{\varepsilon}{2}\|x(t) - \overline{x}\|^2 - \frac{\varepsilon}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 \\ &\geq \frac{1-\varepsilon}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + \left(1 - \beta\varepsilon - \frac{\varepsilon}{\gamma}\right)\big(f(x(t)) - f(\overline{x})\big). \end{aligned}$$

Therefore, by taking ε small enough, we obtain

$$h_{\varepsilon,\beta}(t) \geq C_4\big(\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + f(x(t)) - f(\overline{x})\big) \tag{117}$$

with $C_4 := \min\{\frac{1-\varepsilon}{2}, 1 - \beta\varepsilon - \frac{\varepsilon}{\gamma}\} > 0$. Combining this inequality with (116) and (33), we obtain an exponential convergence rate to zero for $f(x(t)) - f(\overline{x})$, $\|x(t) - \overline{x}\|$ and $\|\dot{x}(t) + \beta\nabla f(x(t))\|$.

Since ∇f is Lipschitz continuous on bounded sets, and $x(t)$ converges to $\overline{x}$, there exists $L_f > 0$ such that for all $t \geq 0$,

$$\|\nabla f(x(t))\| = \|\nabla f(x(t)) - \nabla f(\overline{x})\| \leq L_f\|x(t) - \overline{x}\|.$$

Hence the exponential convergence rate of $\|x(t) - \overline{x}\|$ to zero yields the same property for $\|\nabla f(x(t))\|$. By combining this last property with the exponential convergence rate of $\|\dot{x}(t) + \beta\nabla f(x(t))\|$ to zero, we finally infer that $\|\dot{x}(t)\|$ converges exponentially to zero when $t \to +\infty$. ∎

Remark 7. Similar rates have been reported in [24, Theorem 4.2] for the heavy ball method with Hessian-driven damping.

Remark 8. It is possible to derive similar exponential rates also for the system (70), but for a restrictive choice of $\beta > 0$. To see this, notice that for $\theta > 0$ we have

$$\begin{aligned} &\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t) + \beta\nabla f(x(t))\|^2 + f(x(t)) - f(\overline{x})\right) \\ &\quad = -\langle\dot{x}(t), \nabla\phi(\dot{x}(t))\rangle - \beta\langle\nabla\phi(\dot{x}(t)), \nabla f(x(t))\rangle - \beta\|\nabla f(x(t))\|^2 \\ &\quad \leq -\alpha\|\dot{x}(t)\|^2 + \frac{\beta\theta}{2}\|\nabla f(x(t))\|^2 + \frac{\beta L^2}{2\theta}\|\dot{x}(t)\|^2 - \beta\|\nabla f(x(t))\|^2 \\ &\quad = -\left(\alpha - \frac{\beta L^2}{2\theta}\right)\|\dot{x}(t)\|^2 - \beta\left(1 - \frac{\theta}{2}\right)\|\nabla f(x(t))\|^2. \end{aligned}$$

10.6. Further convergence results based on the quasi-gradient approach

Let us consider the dynamical system (ADIGE-VGH) in case ϕ is differentiable, $f : \mathbb{R}^N \to \mathbb{R}$ is a $\mathcal{C}^2$ function (possibly nonconvex) whose gradient is Lipschitz continuous on bounded sets, and such that $\inf_{\mathbb{R}^N} f > -\infty$:

$$\ddot{x}(t) + \nabla\phi\big(\dot{x}(t) + \beta\nabla f(x(t))\big) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0. \tag{118}$$

The considerations are similar to those of Section 7.3 and Theorem 7.

According to Step 2 in the proof of Theorem 12, the first-order reformulation is

$$\dot{z}(t) + F(z(t)) = 0, \tag{119}$$

where $z(t) = (x(t), u(t)) \in \mathbb{R}^N \times \mathbb{R}^N$, and $F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ is defined by

$$F(x,u) = (\beta\nabla f(x) - u, \nabla\phi(u) + \nabla f(x)).$$

Let us check the angle condition (E_λ is defined as in Theorem 7). We have

$$\begin{aligned}&\langle \nabla E_\lambda(x,u), F(x,u)\rangle\\ &\quad = \langle (\nabla f(x) + \lambda\nabla^2 f(x)u, u + \lambda\nabla f(x)), (\beta\nabla f(x) - u, \nabla\phi(u) + \nabla f(x))\rangle.\end{aligned}$$

After simplification, we get

$$\begin{aligned}\langle \nabla E_\lambda(x,u), F(x,u)\rangle &= -\lambda\langle \nabla^2 f(x)u, u\rangle + \langle u, \nabla\phi(u)\rangle + \lambda\langle \nabla f(x), \nabla\phi(u)\rangle\\ &\quad + \lambda\|\nabla f(x)\|^2 + \beta\|\nabla f(x)\|^2 + \lambda\beta\langle \nabla f(x), \nabla^2 f(x)u\rangle.\end{aligned}$$

We estimate the last term by writing

$$\lambda\beta\langle \nabla f(x), \nabla^2 f(x)u\rangle \geq -\frac{\lambda}{4}\|\nabla f(x)\|^2 - \lambda\beta^2 M^2\|u\|^2$$

and get (as in the proof of Theorem 7)

$$\langle \nabla E_\lambda(x,u), F(x,u)\rangle \geq \left(\gamma - \lambda M - \frac{\lambda}{2}\delta^2 - \lambda\beta^2 M^2\right)\|u\|^2 + \left(\frac{\lambda}{4} + \beta\right)\|\nabla f(x)\|^2. \tag{120}$$

We also have

$$\|F(x,u)\| \leq C_2(\|u\|^2 + \|\nabla f(x)\|^2)^{1/2},$$

where $C_2 = \sqrt{2(2 + \beta^2 + \delta^2)}$. The rest can be done along the lines of the proof of Theorem 7.

11. Conclusion and perspectives

In this article, from the point of view of optimization, we put forward some classical and new properties concerning the asymptotic convergence of autonomous damped inertial dynamics. From a control point of view, the damping terms of these dynamics can be considered as closed-loop controls of the current data: position, speed, gradient of the objective function, Hessian of the objective function, and combinations of these objects. Let us cite some of the main results and advantages of the autonomous approach compared to the nonautonomous approach, where damping involves parameters given from the start as functions of time.

11.1. Pros

- Autonomous systems are easy to implement. It is not necessary to adjust the damping coefficient as is the case for nonautonomous systems.
- When the function to be minimized is strongly convex, there is convergence at an exponential rate, and this is valid for a large class of damping potentials.
- We were able to exploit the quasi-gradient structure of the autonomous damped dynamics and combine them with Kurdyka–Łojasiewicz theory to obtain convergence rates for a large class of functions f, possibly nonconvex. This is specific to the autonomous case because the theories mentioned above are not developed in the nonautonomous case.
- Hessian-driven damping naturally appears within the framework of autonomous systems. It notably improves the theoretical and numerical behavior of the trajectories, by reducing the oscillatory aspects. Its introduction into the algorithms does not change their numerical complexity (it makes appear the difference of the gradient at two consecutive steps). This makes this geometric damping successful; several recent articles have been devoted to it.
- The closed-loop approach clearly distinguishes between strong and weak damping effects, and the transition between them. It also shows the replacement of the theory of convergence by the notion of attractor when damping becomes too weak.
- We have introduced a new autonomous system where damping involves both the speed and the gradient of f, and which benefits from very good convergence properties. At the beginning of time it takes advantage of the inertial effect, then after a finite time it turns into a steepest descent dynamic, thus avoiding oscillatory aspects. This regime change has some similarities with the restart method, and also the recent work of Poon–Liang [88] on adaptive acceleration.
- The closed-loop approach makes it possible to make the link with different fields, such as PDE and control theory, where the stabilization of oscillating systems is a central issue. Although the simple mathematical framework chosen in this article (single function space $\mathcal{H}$, differentiable objective function f) does not make it possible to deal directly with the associated PDEs, the Lyapunov analysis that we have developed can naturally be extended to this framework.
- We have developed an inertial algorithm which shares good convergence properties with the related continuous dynamics, in the case of the quasi-gradient and Kurdyka–Łojasiewicz approach. Note that the quasi-gradient approach reflects relative errors in the algorithms, and therefore gives a lot of flexibility. It is this approach that has made it possible to deal with many different algorithms in Attouch–Bolte–Svaiter [17] in the nonconvex nonsmooth case. It would be interesting to develop these aspects for splitting algorithms, such as proximal gradient algorithms, regularized Gauss–Seidel algorithms, and PALM (see also [51] for a continuous-times approach to structured optimization problems).

11.2. Cons

(1) To date, we do not know, in the autonomous case, the equivalent of the accelerated gradient method of Nesterov and Su–Boyd–Candès damped inertial dynamic, that is, an adjustment of the damping potential which guarantees the rate of convergence $1/t^2$ for any convex function. This is a current research subject; for recent progress in this direction, see Lin–Jordan [79].

(2) The general approach based on the quasi-gradient and Kurdyka–Łojasiewicz theory (as developed in Section 7) works mainly in finite dimensions. Extension of the (KŁ) theory to spaces of infinite dimension is a current research subject.

11.3. Perspectives

(1) Develop closed-loop versions of the Nesterov accelerated gradient method from a theoretical and numerical point of view. Our analysis allowed us to better define the type of damping potential ϕ capable of doing this, but this remains an open question for study. Indeed, the case $p = 2$ (i.e. quadratic behavior of the damping potential near the origin) is the critical case separating weak damping from strong damping. Taking $p > 2$ close to 2 provides a vanishing viscosity damping coefficient, which is a specific property of the Nesterov accelerated gradient method. Our intuition is that we need to refine the power scale which is not precise enough to provide the correct setting of the vanishing damping term (i.e. going from $p = 2$ to $p > 2$, with p even very close to 2 is too sudden a change).

(2) Extend our study to the case of nonsmooth optimization possibly involving a constraint. This is an important subject, which is closely related to item (6) of this list, because the common device to deal with a constrained optimization problem is to use the gradient-projection method, which falls under fixed point methods.

(3) Develop a control perspective with closed-loop damping for restarting methods. Restarting methods take advantage of the inertial effect to accelerate trajectories, then stop when a given criterion deteriorates. Then one restarts from the current point with zero velocity, and so on. In many ways, the dynamic we developed in Section 10 follows a similar strategy. Our results are valid with general data functions f and ϕ, while the known results concerning restart methods only concern the case where f is strongly convex. It is an important subject to explore.

(4) Obtain a closed-loop version of the Tikhonov regularization method, and make the link with the Haugazeau method. The objective is then, within the framework of convex optimization, to obtain an autonomous dynamic whose trajectories strongly converge to the solution of minimum norm; see Attouch–Cabot–Chbani–Riahi [22], and Boţ–Csetnek–László [52] for some recent results in the open-loop case (the Tikhonov regularization parameter tends to zero in a controlled manner, not too fast) and references therein.

(5) Develop the corresponding algorithmic results. Continuous dynamics provide a valuable guide to introduce and analyze algorithms that enjoy similar convergence properties. In Theorem 9 we have analyzed the convergence of an inertial algorithm with general damping potential ϕ and general (tame) function f. A similar analysis can certainly be developed on the basis of Theorems 11 and 13 which also involve Hessian-driven damping. Natural extensions then consist in studying structured optimization problems and the corresponding proximal-based algorithms.

(6) In recent years, most of the previous themes have been extended (in the open-loop case) to the case of maximally monotone operators: see Alvarez–Attouch [10], Attouch–Maingé [30], Attouch–Peypouquet [34], Attouch–Cabot [21], Attouch–László [29], Boţ–Csetnek [47]. It would be interesting to consider closed-loop versions of these dynamics, as was done by Attouch–Redont–Svaiter [36] for first-order Newton-like evolution systems.

(7) Time rescaling is a powerful tool to accelerate inertial systems; see Attouch–Chbani–Riahi [27], Shi–Du–Jordan–Su [89] and references therein. It leads naturally to nonautonomous dynamics. It would be interesting to study autonomous closed-loop versions. This means first extracting quantities which tend monotonically to $+\infty$.

(8) Stability of dynamics and algorithms with respect to perturbations/errors is an important topic from the numerical point of view [22, 25, 27, 92].

(9) The concepts of control theory and dissipative dynamical systems have proven to be useful and intuitive design guidelines for speeding up stochastic gradient methods, especially for the variance-reduction methods for the finite-sum problem (see [72] and the references therein). It is likely that our approach fits these questions well.

Funding. H. Attouch was supported by COST Action, CA16228.
E. R. Csetnek was supported by the Austrian Science Fund, project P 29809-N32.

References

[1] Abbas, B., Attouch, H., Svaiter, B. F.: Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **161**, 331–360 (2014) Zbl 1339.47080 MR 3193795

[2] Adly, S., Attouch, H.: First-order inertial algorithms involving dry friction damping. Math. Program. (online, 2021); DOI:10.1007/s10107-020-01613-y

[3] Adly, S., Attouch, H.: Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping. SIAM J. Optim. **30**, 2134–2162 (2020) Zbl 1452.65103 MR 4133483

[4] Adly, S., Attouch, H.: Finite-time stabilization of continuous inertial dynamics combining dry friction with Hessian-driven damping. J. Convex Anal. **28**, 281–310 (2021) Zbl 1470.37118 MR 4245869

[5] Adly, S., Attouch, H., Cabot, A.: Finite time stabilization of nonlinear oscillators subject to dry friction. In: Nonsmooth Mechanics and Analysis, Adv. Mech. Math. 12, Springer, New York, 289–304 (2006) MR 2205459

[6] Alabau-Boussouira, F., Privat, Y., Trélat, E.: Nonlinear damped partial differential equations and their uniform discretizations. J. Funct. Anal. **273**, 352–403 (2017) Zbl 1364.37155 MR 3646303

[7] Alecsa, C. D., László, S. C., Pinţa, T.: An extension of the second order dynamical system that models Nesterov's convex gradient method. Appl. Math. Optim. **84**, 1687–1716 (2021) MR 4304886

[8] Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control Optim. **38**, 1102–1119 (2000) Zbl 0954.34053 MR 1760062

[9] Alvarez, F., Attouch, H.: Convergence and asymptotic stabilization for some damped hyperbolic equations with non-isolated equilibria. ESAIM Control Optim. Calc. Var. **6**, 539–552 (2001) Zbl 1004.34045 MR 1849415

[10] Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set-Valued Anal. **9**, 3–11 (2001) Zbl 0991.65056 MR 1845931

[11] Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. J. Math. Pures Appl. (9) **81**, 747–779 (2002) Zbl 1036.34072 MR 1930878

[12] Amann, H., Diaz, J. I.: A note on the dynamics of an oscillator in the presence of strong friction. Nonlinear Anal. **55**, 209–216 (2003) Zbl 1035.34023 MR 2007469

[13] Apidopoulos, V., Aujol, J.-F., Dossal, C.: Convergence rate of inertial forward-backward algorithm beyond Nesterov's rule. Math. Program. **180**, 137–156 (2020) Zbl 1439.90055 MR 4062834

[14] Apidopoulos, V., Aujol, J.-F., Dossal, C., Rondepierre, A.: Convergence rates of an inertial gradient descent algorithm under growth and flatness conditions. Math. Program. **187**, 151–193 (2021) Zbl 1465.90062 MR 4246301

[15] Attouch, H.: Variational Convergence for Functions and Operators. Applicable Math. Ser., Pitman (Advanced Publishing Program), Boston, MA (1984) Zbl 0561.49012 MR 773850

[16] Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. Math. Oper. Res. **35**, 438–457 (2010) Zbl 1214.65036 MR 2674728

[17] Attouch, H., Bolte, J., Svaiter, B. F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. Math. Program. **137**, 91–129 (2013) Zbl 1260.49048 MR 3010421

[18] Attouch, H., Buttazzo, G., Michaille, G.: Variational analysis in Sobolev and BV spaces. 2nd ed., MOS-SIAM Ser. Optimization 17, SIAM, Philadelphia, PA, and Mathematical Optimization Society, Philadelphia, PA (2014) Zbl 1311.49001 MR 3288271

[19] Attouch, H., Cabot, A.: Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. J. Differential Equations **263**, 5412–5458 (2017) Zbl 1405.37092 MR 3688419

[20] Attouch, H., Cabot, A.: Convergence rates of inertial forward-backward algorithms. SIAM J. Optim. **28**, 849–874 (2018) Zbl 1387.49047 MR 3775147

[21] Attouch, H., Cabot, A.: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Math. Program. **184**, 243–287 (2020) Zbl 07263694 MR 4163546

[22] Attouch, H., Cabot, A., Chbani, Z., Riahi, H.: Inertial forward-backward algorithms with perturbations: application to Tikhonov regularization. J. Optim. Theory Appl. **179**, 1–36 (2018) Zbl 1417.90114 MR 3846930

[23] Attouch, H., Cabot, A., Redont, P.: The dynamics of elastic shocks via epigraphical regularization of a differential inclusion. Barrier and penalty approximations. Adv. Math. Sci. Appl. **12**, 273–306 (2002) Zbl 1038.49029 MR 1909449

[24] Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping; Math. Program. (online, 2020); DOI:10.1007/s10107-020-01591-1

[25] Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**, 123–175 (2018) Zbl 1395.34068 MR 3767743

[26] Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. ESAIM Control Optim. Calc. Var. **25**, art. 2, 34 pp. (2019) Zbl 1437.49045 MR 3943364

[27] Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via time scaling of damped inertial gradient dynamics. Pure Appl. Funct. Anal. **6**, 1081–1117 (2021) Zbl 07105235 MR 4352915

[28] Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method. I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. Commun. Contemp. Math. **2**, 1–34 (2000) Zbl 0983.37016 MR 1753136

[29] Attouch, H., László, S. C.: Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. SIAM J. Optim. **30**, 3252–3283 (2020) Zbl 07284424 MR 4182908

[30] Attouch, H., Maingé, P.-É.: Asymptotic behavior of second-order dissipative evolution equations combining potential with non-potential effects. ESAIM Control Optim. Calc. Var. **17**, 836–857 (2011) Zbl 1230.34051 MR 2826982

[31] Attouch, H., Maingé, P.-E., Redont, P.: A second-order differential system with Hessian-driven damping; application to non-elastic shock laws. Differential Equations Appl. **4**, 27–65 (2012) Zbl 1248.34091 MR 2952628

[32] Attouch, H., Marques Alves, M., Svaiter, B. F.: A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $\mathcal{O}(1/n^2)$. J. Convex Anal. **23**, 139–180 (2016) Zbl 1348.47053 MR 3522133

[33] Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**, 1824–1834 (2016) Zbl 1346.49048 MR 3543173

[34] Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. **174**, 391–432 (2019) Zbl 1412.37083 MR 3935085

[35] Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping. J. Differential Equations **261**, 5734–5783 (2016) Zbl 1375.49028 MR 3548269

[36] Attouch, H., Redont, P., Svaiter, B. F.: Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **157**, 624–650 (2013) Zbl 1290.90061 MR 3047022

[37] Attouch, H., Svaiter, B. F.: A continuous dynamical Newton-like approach to solving monotone inclusions. SIAM J. Control Optim. **49**, 574–598 (2011) Zbl 1229.34097 MR 2784701

[38] Aubin, J.-P., Ekeland, I.: Applied Nonlinear Analysis. Pure Appl. Math. (New York), Wiley, New York (1984) Zbl 0641.47066 MR 749753

[39] Aujol, J.-F., Dossal, Ch., Rondepierre, A.: Convergence rates of the Heavy-Ball method for quasi-strongly convex optimization. https://hal.archives-ouvertes.fr/hal-02545245 (2020)

[40] Baji, B., Cabot, A., Díaz, J. I.: Asymptotics for some nonlinear damped wave equation: finite time convergence versus exponential decay results. Ann. Inst. H. Poincaré Anal. Non Linéaire **24**, 1009–1028 (2007) Zbl 1139.35017 MR 2371117

[41] Bárta, T., Chill, R., Fašangová, E.: Every ordinary differential equation with a strict Lyapunov function is a gradient system. Monatsh. Math. **166**, 57–72 (2012) Zbl 1253.37019 MR 2901252

[42] Bárta, T., Fašangová, E.: Convergence to equilibrium for solutions of an abstract wave equation with general damping function. J. Differential Equations **260**, 2259–2274 (2016) MR 3427666

[43] Bauschke, H. H., Combettes, P. L.: Convex analysis and monotone operator theory in Hilbert spaces. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York (2011) Zbl 1218.47001 MR 2798533

[44] Bégout, P., Bolte, J., Jendoubi, M. A.: On damped second-order gradient systems. J. Differential Equations **259**, 3115–3143 (2015) Zbl 1347.34082 MR 3360667

[45] Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Amer. Math. Soc. **362**, 3319–3363 (2010) Zbl 1202.26026 MR 2592958

[46] Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**, 459–494 (2014) Zbl 1297.90125 MR 3232623

[47] Boţ, R. I., Csetnek, E. R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control Optim. **54**, 1423–1443 (2016) Zbl 1339.34070 MR 3507089

[48] Boţ, R. I., Csetnek, E. R.: A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function. ESAIM Control Optim. Calc. Var. **24**, 463–477 (2018) Zbl 1428.90128 MR 3816401

[49] Boţ, R. I., Csetnek, E. R., László, S. C.: An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. EURO J. Comput. Optim. **4**, 3–25 (2016) Zbl 1338.90311 MR 3500980

[50] Boţ, R. I., Csetnek, E. R., László, S. C.: Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems. J. Evolution Equations **18**, 1291–1318 (2018) Zbl 1402.34069 MR 3859450

[51] Boţ, R. I., Csetnek, E. R., László, S. C.: A primal-dual dynamical approach to structured convex minimization problems. J. Differential Equations **269**, 10717–10757 (2020) Zbl 1452.37095 MR 4150358

[52] Boţ, R. I., Csetnek, E. R., László, S. C.: Tikhonov regularization of a second order dynamical system with Hessian damping. Math. Program. **189**, 151–186 (2020) Zbl 07402657 MR 4306070

[53] Brézis, H.: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland Math. Stud. 5, North-Holland (1973) Zbl 0252.47055 MR 0348562

[54] Bruck, R. E., Jr.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. J. Funct. Anal. **18**, 15–26 (1975) Zbl 0319.47041 MR 377609

[55] Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation. Trans. Amer. Math. Soc. **361**, 5983–6017 (2009) Zbl 1191.34078 MR 2529922

[56] Cabot, A., Frankel, P.: Asymptotics for some semilinear hyperbolic equations with non-autonomous damping. J. Differential Equations **252**, 294–322 (2012) Zbl 1242.34108 MR 2852207

[57] Carles, R., Gallo, C.: Finite time extinction by nonlinear damping for the Schrödinger equation. Comm. Partial Differential Equations **36**, 961–975 (2011) Zbl 1228.35040 MR 2765425

[58] Castera, C., Bolte, J., Févotte, C., Pauwels, E.: An inertial Newton algorithm for deep learning. J. Mach. Learn. Res. **22**, art. 134, 31 pp. (2021) Zbl 07415077 MR 4318490

[59] Chambolle, A., Dossal, C.: On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**, 968–982 (2015) Zbl 1371.65047 MR 3375610

[60] Chergui, L.: Convergence of global and bounded solutions of a second order gradient like system with nonlinear dissipation and analytic nonlinearity. J. Dynam. Differential Equations **20**, 643–652 (2008) Zbl 1167.34017 MR 2429439

[61] Chill, R., Fašangová, E.: Gradient Systems. Lecture Notes of the 13th International Internet Seminar, Matfyzpress, Prague (2010)

[62] Chill, R., Haraux, A., Jendoubi, M. A.: Applications of the Łojasiewicz-Simon gradient inequality to gradient-like evolution equations. Anal. Appl. (Singap.) **7**, 351–372 (2009) Zbl 1192.34068 MR 2572850

[63] Gambino, G., Tanriver, U., Guha, P., Choudhury, A. G., Choudhury, S. R.: Regular and singular pulse and front solutions and possible isochronous behavior in the short-pulse equation: phase-plane, multi-infinite series and variational approaches. Commun. Nonlinear Sci. Numer. Simul. **20**, 375–388 (2015) Zbl 1308.35049 MR 3251500

[64] Gelfand, I. M., Tsetlin, M. L., Principle of nonlocal search in systems of automatic optimization. Dokl. Akad. Nauk SSSR **137** (1961), 295–298 (in Russian) MR 0129118

[65] Ghisi, M., Gobbino, M., Haraux, A.: Universal bounds for a class of second order evolution equations and applications. J. Math. Pures Appl. (9) **142**, 184–203 (2020) Zbl 1448.35045 MR 4149688

[66] Grasselli, M., Pierre, M.: Convergence to equilibrium of solutions of the backward Euler scheme for asymptotically autonomous second-order gradient-like systems. Commun. Pure Appl. Anal. **11**, 2393–2416 (2012) Zbl 1264.65119 MR 2912753

[67] Hale, J. K.: Stability and gradient dynamical systems. Rev. Mat. Complut. **17**, 7–57 (2004) Zbl 1070.37055 MR 2063940

[68] Haraux, A.: Nonlinear Evolution Equations—Global Behavior of Solutions. Lecture Notes in Math. 841, Springer, Berlin (1981) Zbl 0461.35002 MR 610796

[69] Haraux, A.: Systèmes dynamiques dissipatifs et applications. Recherches Math Appl. 17, Masson, Paris (1991) Zbl 0726.58001 MR 1084372

[70] Haraux, A., Jendoubi, M. A.: Convergence of solutions of second-order gradient-like systems with analytic nonlinearities. J. Differential Equations **144**, 313–320 (1998) Zbl 0915.34060 MR 1616968

[71] Haraux, A., Jendoubi, M. A.: The Convergence Problem for Dissipative Autonomous Systems. Springer, Cham, and BCAM Basque Center for Applied Mathematics, Bilbao (2015) Zbl 1345.37081 MR 3380967

[72] Hu, B., Wright, S., Lessard, L., Dissipativity theory for accelerating stochastic variance reduction: a unified analysis of SVRG and Katyusha using semidefinite programs. In: Proc. 35 Int. Conf. on Machine Learning, PMLR 80, 2038–2047 (2012)

[73] Huang, S.-Z.: Gradient Inequalities. Math. Surveys Monogr. 126, Amer. Math. Soc., Providence, RI (2006) Zbl 1132.35002 MR 2226672

[74] Ioffe, A. D.: An invitation to tame optimization. SIAM J. Optim. **19**, 1894–1917 (2008) Zbl 1182.90083 MR 2486055

[75] Iutzeler, F., Hendrickx, J. M.: A generic online acceleration scheme for optimization algorithms via relaxation and inertia. Optim. Methods Softw. **34**, 383–405 (2019) Zbl 1407.65062 MR 3909025

[76] Kim, D.: Accelerated proximal point method for maximally monotone operators. Math. Program. **190**, 57–87 (2021) Zbl 07411824 MR 4322636

[77] Li, G., Pong, T. K.: Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods. Found. Comput. Math. **18**, 1199–1232 (2018) Zbl 1405.90076 MR 3857908

[78] May, R.: Asymptotic for a second-order evolution equation with convex potential and vanishing damping term. Turkish J. Math. **41**, 681–685 (2017) Zbl 1424.34186 MR 3666623

[79] Muehlebach, M., Jordan, M. I.: Optimization with momentum: dynamical, control-theoretic, and symplectic perspectives. J. Mach. Learn. Res. **22**, art. 73, 50 pp. (2021) Zbl 07370590 MR 4253766

[80] Nesterov, Y.: Introductory Lectures on Convex Optimization. Appl. Optimization 87, Kluwer, Boston, MA (2004) Zbl 1086.90045 MR 2142598

[81] Nesterov, Y. E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR **269**, 543–547 (1983) (in Russian) Zbl 0535.90071 MR 701288

[82] Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for nonconvex optimization. SIAM J. Imaging Sci. **7**, 1388–1419 (2014) Zbl 1296.90094 MR 3218822

[83] Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Amer. Math. Soc. **73**, 591–597 (1967) Zbl 0179.19902 MR 211301

[84] Palis, J., Jr., de Melo, W.: Geometric Theory of Dynamical Systems. Springer, New York (1982) Zbl 0491.58001 MR 669541

[85] Poljak, B. T.: Some methods of speeding up the convergence of iterative methods. Ž. Vyčisl. Mat i Mat. Fiz. **4**, 791–803 (1964) (in Russian) Zbl 0147.35301 MR 169403

[86] Polyak, B. T.: Introduction to Optimization. Optimization Software, New York (1987) MR 1099605

[87] Polyak, B. T., Kulakova, N., Danilova, M.: Accelerated gradient methods revisited. In: Workshop "Variational Analysis and Applications", August 28–September 5, 2018, Erice

[88] Poon, C., Liang, J., Geometry of first-order methods and adaptive acceleration. arXiv:2003.03910v1 (2020)

[89] Shi, B., Du, S. S., Jordan, M. I., Su, W. J.: Understanding the acceleration phenomenon via high-resolution differential equations. Math. Program. (online, 2021); DOI:10.1007/s10107-021-01681-8

[90] Su, W., Boyd, S., Candès, E. J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. J. Mach. Learn. Res. **17**, art. 153, 43 pp. (2016) Zbl 1391.90667 MR 3555044

[91] van den Dries, L.: Tame Topology and o-Minimal Structures. London Math. Soc. Lecture Note Ser. 248, Cambridge Univ. Press, Cambridge (1998) Zbl 0953.03045 MR 1633348

[92] Villa, S., Salzo, S., Baldassarre, L., Verri, A.: Accelerated and inexact forward-backward algorithms. SIAM J. Optim. **23**, 1607–1633 (2013) Zbl 1295.90049 MR 3085126