Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping

Samir Adly, Hedy Attouch

Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping

Samir ADLY* and Hedy ATTOUCH†

ABSTRACT. In a Hilbert space $\mathcal{H}$, we introduce a new class of proximal-gradient algorithms with finite convergence properties. These algorithms naturally occur as discrete temporal versions of an inertial differential inclusion which is damped under the joint action of three dampings: a viscous damping, a geometric damping driven by the Hessian and a dry friction damping. The function $f : \mathcal{H} \to \mathbb{R}$ to be minimized is supposed to be differentiable (not necessarily convex), and enters the algorithm via its gradient. The dry friction damping function $\phi : \mathcal{H} \to \mathbb{R}_+$ is convex with a sharp minimum at the origin, (typically $\phi(x) = r\|x\|$ with $r > 0$). It enters the algorithm via its proximal mapping, which acts as a soft threshold operator on the velocities. The geometrical damping driven by the Hessian intervenes in the dynamics in the form $\nabla^2 f(x(t))\dot{x}(t)$. By treating this term as the time derivative of $\nabla f(x(t))$, this gives, in discretized form, first-order algorithms. The Hessian driven damping allows to control and to attenuate the oscillations which occur naturally with the inertial effect. The convergence results tolerate the presence of errors, under the sole assumption of their asymptotic convergence to zero. Then, replacing the potential function f by its Moreau envelope, we extend the results to the case of a nonsmooth convex function f. In this case, the algorithm involves the proximal operators of f and ϕ separately. Several variants of this algorithm are considered, including the case of the Nesterov accelerated gradient method. We then consider the extension in the case of additive composite optimization, thus leading to splitting methods. Numerical experiments are given for Lasso-type problems. The performance profiles, as a comparison tool, highlight the effectiveness of two variants of the Nesterov accelerated method with dry friction and Hessian-driven viscous damping.

Mathematics Subject Classifications: 37N40, 34A60, 34G25, 49K24, 70F40.

Key words and phrases: proximal-gradient algorithms; inertial methods; differential inclusion; dry friction; Hessian-driven damping; finite convergence; Lasso problem.

1 Introduction and preliminary results

Throughout the paper $\mathcal{H}$ is a real Hilbert space, with the scalar product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$, and $f : \mathcal{H} \to \mathbb{R}$ is a $\mathcal{C}^1$ function whose gradient is Lipschitz continuous. When we consider the continuous dynamics on which the algorithms are based, and where the Hessian intervenes, more regularity is needed for f which is then assumed to be a $\mathcal{C}^2$ function. Several extensions of these hypotheses will be discussed later in the paper.

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

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

1.1 Presentation of the algorithm

We will analyze the finite convergence (within a finite number of steps) of several algorithms that can be obtained by temporal discretization of the second-order differential inclusion

$$\text{(IGDH)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,\ t \in [t_0, +\infty[\qquad (1.1)$$

where γ and β are positive damping parameters. (IGDH) stands shortly for Inertial Gradient system with Dry friction and Hessian-driven damping. The dry friction damping function $\phi : \mathcal{H} \to \mathbb{R}_+$ is convex with a sharp minimum at the origin, typically $\phi(x) = r\|x\|$ with $r > 0$. The geometrical damping driven by the Hessian intervenes in the dynamics in the form $\nabla^2 f(x(t))\dot{x}(t)$. By treating this term as the time derivative of $\nabla f(x(t))$, this gives, in discretized form, first-order algorithms. Our main results concern the finite convergence of the Inertial Proximal-gradient Algorithm with Hessian-Damping and Dry friction (IPAHDD)

$$\begin{cases} z_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f(x_k) \\ x_{k+1} = x_k + h\,\mathrm{prox}_{\frac{h}{1+h\gamma}\phi}(z_k), \end{cases}$$

which comes naturally from the temporal discretization of (IGDH). In the above formula, prox_ϕ denotes the proximal mapping associated with the convex function ϕ. Recall that, for any $x \in \mathcal{H}$, for any $\lambda > 0$

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

We will show that, if the viscous damping parameter γ is taken large enough, then for any sequence (x_k) generated by the algorithm (IPAHDD), the following summability property is satisfied

$$\sum_{k=1}^{+\infty} \|x_{k+1} - x_k\| < +\infty.$$

This property expresses that the trajectory has a finite length, and therefore $\lim_{k\to\infty} x_k := x_\infty$ exists for the strong topology of $\mathcal{H}$. The limit point x_∞ satisfies

$$-\nabla f(x_\infty) \in \partial\phi(0).$$

It is is an "approximate" critical point of f. This amounts to solving the optimization problem $\min_{\mathcal{H}} f$ with the Ekeland variational principle, instead of the Fermat rule. Since our goal is to minimize the function f, we will have to choose a function ϕ whose subdifferential set $\partial\phi(0)$ is "relatively small". When $\phi(x) = r\|x\|$, this corresponds to taking r a small positive number. Moreover, we will show that, under the condition

$$-\nabla f(x_\infty) \in \mathrm{int}\,(\partial\phi(0)),$$

there is finite convergence (*i.e.* within a finite number of steps) of the iterates generated by the algorithm (IPAHDD). In short, dry friction acts as a closed-loop stopping rule. The Hessian driven damping allows to control and to attenuate the oscillation effects which occur naturally with the inertial systems, and which are not desirable from the optimization point of view. In many ways, (IPAHDD) can be compared to the restart method.

1.2 Some historical facts

Let's explain the role and the importance of each of the three damping terms that enter the continous dynamic (IGDH).

1.2.1 Viscous friction

B. Polyak initiated the use of inertial dynamics to accelerate the gradient method in optimization. In [32], based on the inertial system with a fixed viscous damping coefficient $\gamma > 0$

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

he introduced the Heavy Ball with Friction method. For a strongly convex function f, and γ judiciously chosen, (HBF) provides convergence at exponential rate of $f(x(t))$ to $\min_{\mathcal{H}} f$. For general convex functions, the asymptotic convergence rate of (HBF) is $\mathcal{O}(\frac{1}{t})$ (in the worst case). This is however not better than the steepest descent. A decisive step to improve (HBF) was taken by Su-Boyd-Candès [36] with the introduction of an Asymptotic Vanishing Damping coefficient $\gamma(t) = \frac{\alpha}{t}$, that is

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

As a specific feature, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes (tends to zero) as time t goes to infinity, hence the terminology. For general convex functions it provides a continuous version of the accelerated gradient method of Nesterov. For $\alpha \geq 3$, each trajectory $x(\cdot)$ of (AVD)_α satisfies the asymptotic convergence rate of the values $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(1/t^2\right)$. The convergence properties of the dynamic (AVD)_α have been the subject of many recent studies, see [6, 8, 9, 10, 12, 13, 15, 18, 19, 31, 36]. The case $\alpha = 3$, which corresponds to Nesterov's historical algorithm, is critical. In the case $\alpha = 3$, the question of the convergence of the trajectories remains an open problem (except in one dimension where convergence holds [13]). For $\alpha > 3$, it has been shown by Attouch-Chbani-Peypouquet-Redont [12] that each trajectory converges weakly to a minimizer. The corresponding algorithmic result has been obtained by Chambolle-Dossal [25]. For $\alpha > 3$, it is shown in [15] and [31] that the asymptotic convergence rate of the values is actually $o(1/t^2)$. The subcritical case $\alpha \leq 3$ has been examined by Apidopoulos-Aujol-Dossal [6] and Attouch-Chbani-Riahi [13], with the convergence rate of the objective values $\mathcal{O}\left(t^{-\frac{2\alpha}{3}}\right)$. These rates are optimal, that is, they can be reached, or approached arbitrarily close.

1.2.2 Dry friction

The first results concerning the finite convergence property under the action of dry friction have been obtained by Adly-Attouch-Cabot [3] for the continuous dynamics

$$\ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0, \ t \in [t_0, +\infty[. \tag{1.2}$$

Assuming that the potential friction function ϕ has a sharp minimum at the origin (dry friction), they showed that, generically with respect to the initial data, the solution trajectories converge in finite time to equilibria. Similar results for the corresponding proximal-based algorithms have been obtained by Baji-Cabot [20] and Adly-Attouch [2].

Let's make precise the tools that will be useful for the mathematical analysis of the set-valued term $\partial\phi(\dot{x}(t))$ in (1.2) which models dry friction. The friction potential function ϕ is supposed to satisfy the Dry Friction property (denoted by (DF))

$$\text{(DF)} \qquad \begin{cases} \phi : \mathcal{H} \to \mathbb{R} \text{ is convex continuous;} \\ \min\limits_{\xi \in \mathbb{R}^n} \phi(\xi) = \phi(0) = 0; \\ 0 \in \operatorname{int}(\partial\phi(0)). \end{cases}$$

The particular case $\phi = r\|\cdot\|$, with $r > 0$, models dry friction (also called Coulomb friction) in mechanics. The key assumption $0 \in \operatorname{int}(\partial\phi(0))$ expresses that ϕ has a sharp minimum at the origin. This is specified in the following elementary lemma, see [1, Lemma 4.1 page 83], where, in item (iv), ϕ^* is the Fenchel conjugate of ϕ.

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

(i) $0 \in \operatorname{int}(\partial\phi(0))$;

(ii) *there exists some $r > 0$ such that $B(0, r) \subset \partial\phi(0)$;*

(iii) *there exists some $r > 0$ such that, for all $\xi \in \mathcal{H}$, $\phi(\xi) \geq r\|\xi\|$.*

(iv) *there exists some $r > 0$ such that, $\|f\| \leq r \Longrightarrow \partial\phi^*(f) \ni 0$.*

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

The following lemma will play a key role in showing the finite convergence property. It gives the soft thresholding property satisfied by the proximal operator associated with a function ϕ having a sharp minimum at the origin. It is an immediate consequence of Lemma 1.1 (iv).

Lemma 1.2 *Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function which satisfies the property $(\mathrm{DF})_r$, i.e. $\partial\phi(0) \supset B(0, r)$. Then, the following implication holds: for $\lambda > 0$, and $x \in \mathcal{H}$*

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

1.2.3 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 [5]. In line with (HBF), it contains a *fixed* positive friction coefficient γ. The introduction of the Hessian-driven damping makes it possible to neutralize the transversal oscillations likely to occur with (HBF), as observed in [5] in the case of the Rosenbrook function. The need to take a geometric damping adapted to f had already been observed by Alvarez [4] 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. But still this damping operator is fixed. For a general convex function, the Hessian-driven damping in $(\mathrm{DIN})_{\gamma,\beta}$ performs a similar operation in a closed-loop adaptive way. The terminology (DIN) stands shortly for Dynamical Inertial Newton. It refers to the natural link between this dynamic and the continuous Newton method. Recent studies have been devoted to the study of the inertial dynamic

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

which combines asymptotic vanishing damping with Hessian-driven damping. The corresponding algorithms involve a correcting term in the Nesterov accelerated gradient method which reduces the oscillatory aspects, see Attouch-Peypouquet-Redont [16], Attouch-Chbani-Fadili-Riahi [17], Shi-Du-Jordan-Su [34].

1.3 Contents

The paper is organized as follows. In section 2, we state our main results, which concern the convergence properties of the inertial algorithm (IPAHDD). We then specialize our results in the case of the soft thresholding of velocities. In section 2.6 we examine a variant of (IPAHDD) which is based on another discretization of the viscous damping term. In section 3, we examine the effect of the introduction of perturbations, errors in the algorithm (IPAHDD). In section 4, we proceed with a similar analysis in the case of the Nesterov acceleration method. In section 5, based on the variational properties of Moreau's envelope, we extend these results to the case where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous and proper function such that $\inf f > -\infty$. Thus, we will obtain similar results for an algorithm in which the two nonsmooth functions f and ϕ enter the algorithm via their proximal mappings in a splitting form. In section 6, we extend our analysis to the case of additive composite optimization problems, and obtain splitting methods with finite convergence properties. Section 7 is devoted to numerical experiments, and comparing on the Lasso problem the performance of the different algorithms considered previously. We complete the paper with some perspectives.

2 Inertial proximal-based algorithms with dry friction and Hessian-driven damping

In this section, we assume that f is a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous. Unless otherwise indicated, no convexity assumption is made on the function f. We will consider a splitting algorithm with the finite convergence property, in which the function to be minimized f intervenes via its gradient, and the potential friction function ϕ via its proximal mapping. We denote by γ, β, r the three positive damping parameters. They can be respectively interpreted as

$$(\text{Damping parameters}) \qquad \begin{cases} \gamma \ \text{ is a viscous damping parameter ;} \\ \beta \ \text{ is attached to the Hessian-driven damping;} \\ r \ \text{ is a dry friction parameter, that is, } \phi(x) \geq r\|x\| \ \text{ and } \phi(0) = 0. \end{cases}$$

2.1 Proximal-Gradient algorithms with Hessian damping and dry friction

Given a constant time step $h > 0$, we consider the following temporal discretization of (IGDH)

$$\begin{aligned} &\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) \\ &+\frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) + \nabla f(x_k) \ni 0. \end{aligned} \tag{2.1}$$

It is implicit with respect to the nonsmooth function ϕ, and explicit with respect to the smooth function f. It is in line with the classical proximal-gradient methods that deal with additively structured minimization problems *smooth + nonsmooth*. But here, this structure involves the friction terms, hence significant differences! As a key ingredient, we used that $\nabla^2 f(x(t))\dot{x}(t) = \frac{d}{dt}\nabla f(x(t))$, which follows from the classical derivation chain rule. So, the correcting term $\nabla f(x_k) - \nabla f(x_{k-1})$ is directly related to the temporal discretization of the Hessian-driven damping term. It plays a central role to reduce the oscillatory effects which are attached to the inertial systems.
Solving (2.1) with respect to x_{k+1} gives the following first-order algorithm where dry friction enters via

the potential function ϕ, and the function to be minimized f enters via its gradient.

(IPAHDD): Inertial Proximal-gradient Algorithm with Hessian-Damping and Dry friction

Initialize: $x_0 \in \mathcal{H}$, $x_1 \in \mathcal{H}$

$$x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}\left(\tfrac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \tfrac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \tfrac{h}{1+h\gamma}\nabla f(x_k)\right)$$

We call it (IPAHDD), which stands for Inertial Proximal-gradient Algorithm with Hessian Damping and Dry friction. Consequently, given x_{k-1} and x_k, (IPAHDD) uniquely determines x_{k+1}. When $\phi = 0$, that is, without dry friction, we obtain the Inertial Gradient Algorithm with Hessian damping

$$\text{(IGAHD)} \quad x_{k+1} = x_k + \frac{1}{1+h\gamma}(x_k - x_{k-1}) - \frac{h\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h^2}{1+h\gamma}\nabla f(x_k).$$

(IGAHD) is based on the heavy ball with friction method. In the case of the accelerated gradient method of Nesterov, inertial algorithms involving a similar correcting term were studied recently by Attouch-Chbani-Fadili-Riahi [17] and Shi-Du-Jordan-Su [34]. We can now state the main results of the paper. In order not to make the statements too long, we expose separately the qualitative and the quantitative convergence results.

2.2 Convergence: finite length property

Theorem 2.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and such that $\inf_{\mathcal{H}} f > -\infty$. Assume that the potential friction function ϕ satisfies $(\mathrm{DF})_r$. Suppose that the parameters h, γ, β in the algorithm (IPAHDD) satisfy the relation*

$$\gamma \geq L\left(\frac{h}{2} + \beta\right).$$

Then, for any sequence (x_k) defined by the algorithm (IPAHDD)), we have:

(i) $\sum_k \|x_{k+1} - x_k\| < +\infty$, *and hence $\lim x_k := x_\infty$ exists for the strong topology of $\mathcal{H}$. Moreover,*

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r}\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right)$$
$$\sum_{k=1}^{\infty} \|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq 2h^2\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right),$$

where $E_1 := \frac{1}{2}\|\frac{1}{h}(x_1 - x_0)\|^2 + f(x_1) - \inf_{\mathcal{H}} f$.

(ii) *The limit x_∞ of the sequence (x_k) satisfies:* $0 \in \partial\phi(0) + \nabla f(x_\infty)$.

Proof. We will use an energetic argument based on the nonincreasing property of the sequence (E_k) of nonnegative global energy functions

$$E_k := \frac{1}{2}\|\frac{1}{h}(x_k - x_{k-1})\|^2 + f(x_k) - \inf_{\mathcal{H}} f.$$

Let's formulate (2.1) in terms of the discrete velocity vectors $\frac{1}{h}(x_k - x_{k-1})$. After multiplication by h, we obtain the equivalent formulation

$$\begin{aligned}&\frac{1}{h}(x_{k+1}-x_k)-\frac{1}{h}(x_k-x_{k-1})+\gamma h\frac{1}{h}(x_{k+1}-x_k)+h\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right)\\&+\beta\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)+h\nabla f(x_k)\ni 0.\end{aligned}\tag{2.2}$$

(i) Let's first establish energy estimates. Without ambiguity, we write simply $\partial\phi$ to designate any element belonging to this set. Taking the scalar product of (2.2) with $\frac{1}{h}(x_{k+1} - x_k)$, we obtain

$$\begin{aligned}&\left\langle\frac{1}{h}(x_{k+1}-x_k)-\frac{1}{h}(x_k-x_{k-1}),\,\frac{1}{h}(x_{k+1}-x_k)\right\rangle+\gamma h\|\frac{1}{h}(x_{k+1}-x_k)\|^2+\langle\nabla f(x_k),x_{k+1}-x_k\rangle\\&+h\left\langle\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right),\frac{1}{h}(x_{k+1}-x_k)\right\rangle+\beta\left\langle\nabla f(x_k)-\nabla f(x_{k-1}),\frac{1}{h}(x_{k+1}-x_k)\right\rangle=0.\end{aligned}\tag{2.3}$$

Set $X_k := \frac{1}{h}(x_k - x_{k-1})$. The following elementary relation reflects the strong convexity of $\frac{1}{2}\|\cdot\|^2$

$$\langle X_{k+1}-X_k,X_{k+1}\rangle=\frac{1}{2}\|X_{k+1}\|^2-\frac{1}{2}\|X_k\|^2+\frac{1}{2}\|X_{k+1}-X_k\|^2.\tag{2.4}$$

According to the convexity of ϕ and $\phi(0) = 0$, we have

$$\langle\partial\phi\left(X_{k+1}\right),X_{k+1}\rangle\geq\phi\left(X_{k+1}\right).\tag{2.5}$$

Taking into account (2.4) and (2.5), we deduce from (2.3) the following inequality

$$\begin{aligned}&\frac{1}{2}\|X_{k+1}\|^2-\frac{1}{2}\|X_k\|^2+\frac{1}{2}\|X_{k+1}-X_k\|^2+\gamma h\|X_{k+1}\|^2+h\phi(X_{k+1})\\&+\beta\left\langle\nabla f(x_k)-\nabla f(x_{k-1}),\frac{1}{h}(x_{k+1}-x_k)\right\rangle+\langle\nabla f(x_k),x_{k+1}-x_k\rangle\leq 0.\end{aligned}\tag{2.6}$$

Let's now use the assumptions on the potential functions ϕ and f. According to the assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1, for all $k \geq 1$

$$\phi\left(X_{k+1}\right)\geq r\|X_{k+1}\|.\tag{2.7}$$

Since ∇f is L-Lipschitz continuous, the classical gradient descent lemma gives, for all $k \geq 1$

$$f(x_{k+1})\leq f(x_k)+\langle\nabla f(x_k),x_{k+1}-x_k\rangle+\frac{L}{2}\|x_{k+1}-x_k\|^2.\tag{2.8}$$

According to the Cauchy-Schwarz inequality, and using again that ∇f is L-Lipschitz continuous,

$$\left|\left\langle\nabla f(x_k)-\nabla f(x_{k-1}),\frac{1}{h}(x_{k+1}-x_k)\right\rangle\right|\leq hL\|X_k\|\|X_{k+1}\|\leq\frac{hL}{2}(\|X_k\|^2+\|X_{k+1}\|^2).\tag{2.9}$$

Combining inequalities (2.7)-(2.8)-(2.9) with (2.6), we obtain, for all $k \geq 1$

$$\begin{aligned}&\frac{1}{2}\|\frac{1}{h}(x_{k+1}-x_k)\|^2-\frac{1}{2}\|\frac{1}{h}(x_k-x_{k-1})\|^2+\frac{1}{2h^2}\|x_{k+1}-2x_k+x_{k-1}\|^2+\frac{\gamma}{h}\|x_{k+1}-x_k\|^2\\&+r\|x_{k+1}-x_k\|+f(x_{k+1})-f(x_k)-\frac{L}{2}\|x_{k+1}-x_k\|^2\leq\frac{\beta L}{2h}(\|x_k-x_{k-1}\|^2+\|x_{k+1}-x_k\|^2).\end{aligned}$$

In terms of $E_k := \frac{1}{2}\|\frac{1}{h}(x_k - x_{k-1})\|^2 + f(x_k) - \inf_{\mathcal{H}} f$, this is equivalent to

$$\begin{aligned}&E_{k+1}-E_k+\left(\frac{\gamma}{h}-\frac{L}{2}-\frac{\beta L}{2h}\right)\|x_{k+1}-x_k\|^2+\frac{1}{2h^2}\|x_{k+1}-2x_k+x_{k-1}\|^2\\&+r\|x_{k+1}-x_k\|\leq\frac{\beta L}{2h}\|x_k-x_{k-1}\|^2.\end{aligned}$$

According to the assumption $\gamma \geq L\left(\frac{h}{2}+\beta\right)$, we have $\frac{\gamma}{h}-\frac{L}{2}-\frac{\beta L}{2h} \geq \frac{\beta L}{2h}$. Therefore,

$$E_{k+1}-E_k+\frac{1}{2h^2}\|x_{k+1}-2x_k+x_{k-1}\|^2+r\|x_{k+1}-x_k\|+\frac{\beta L}{2h}\|x_{k+1}-x_k\|^2 \leq \frac{\beta L}{2h}\|x_k-x_{k-1}\|^2. \quad (2.10)$$

Set $\tilde{E}_k := E_k + \frac{\beta L}{2h}\|x_k - x_{k-1}\|^2$. We have

$$\tilde{E}_{k+1} - \tilde{E}_k + \frac{1}{2h^2}\|x_{k+1} - 2x_k + x_{k-1}\|^2 + r\|x_{k+1} - x_k\| \leq 0. \quad (2.11)$$

Adding the above inequalities, and according to $E_k \geq 0$, and $r > 0$, we deduce from (2.11) that

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r}\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right) < +\infty. \quad (2.12)$$

Therefore, the sequence (x_k) has a finite length, which implies that the strong limit of the sequence (x_k) exists. Set $x_\infty := \lim x_k$. Moreover, according to (2.11), we also get

$$\sum_{k=1}^{\infty} \|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq 2h^2\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right) < +\infty. \quad (2.13)$$

Estimation (2.13) gives more accurate information than (2.12) when the step size h is small.

(ii) From $\sum_k \|x_{k+1} - x_k\| < +\infty$, we get $\lim_k \|x_{k+1} - x_k\| = 0$. This in turn implies

$$\lim_k \frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) = \lim_k \frac{1}{h^2}\left((x_{k+1} - x_k) - (x_k - x_{k-1})\right) = 0.$$

Moreover, since ∇f is Lipschitz continuous and (x_k) converges strongly to x_∞, we have

$$\lim_k \nabla f(x_k) = \nabla f(x_\infty) \quad \text{and} \quad \nabla f(x_k) - \nabla f(x_{k-1}) \to 0.$$

To pass to the limit on (2.1), rewrite it as follows:

$$-\frac{1}{h^2}(x_{k+1}-2x_k+x_{k-1})-\frac{\gamma}{h}(x_{k+1}-x_k)-\frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right)-\nabla f(x_k) \in \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right).$$

According to the above convergence results and the closedness of the graph of $\partial\phi$, we deduce that

$$-\nabla f(x_\infty) \in \partial\phi\left(0\right),$$

which gives item (ii). ■

2.3 Convergence rate: linear and finite convergence results

We have shown that the limit of the iterates x_∞ satisfies $-\nabla f(x_\infty) \in \partial\phi\left(0\right)$. We will show that, when it happens that x_∞ satisfies the stronger property

$$-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0)), \quad (2.14)$$

we then obtain linear convergence and finite convergence results. Note that condition (2.14) involves the limit of the iterates x_∞, which is a priori unknown. But practically, this condition is almost always satisfied, making it a valuable numerical result.

Theorem 2.2 *(linear convergence, finite convergence) Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and such that $\inf_{\mathcal{H}} f > -\infty$. Assume that the potential friction function ϕ satisfies $(\mathrm{DF})_r$. Suppose that the parameters h, γ, β in the algorithm (IPAHDD) satisfy the relation*

$$\gamma \geq L\left(\frac{h}{2}+\beta\right).$$

Let (x_k) be a sequence generated by the algorithm (IPAHDD), and let x_∞ be its strong limit (as given by Theorem 2.1).

(i) *Suppose that*

$$-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0)).$$

Then, there is geometric convergence of the velocities to zero. Set $q := \dfrac{1}{\sqrt{1+\frac{2h(\gamma-\beta L)}{1+\beta h L}}}$ which satisfies $0<q<1$: there exists $k_0 \geq 0$ such that for all $k \geq k_0$

$$\|x_{k+1}-x_k\| \leq q^k \|x_1-x_0\|.$$

There is geometric convergence of the sequence (x_k): for all $k \geq k_0$

$$\|x_k - x_\infty\| \leq \frac{q^k}{1-q}\|x_1-x_0\|.$$

(ii) *Suppose that*

$$\|\nabla f(x_\infty)\| < r \quad \textit{where} \quad B(0,r) \subset \partial\phi(0).$$

Then the sequence (x_k) is finitely convergent. The iteration stops at x_k when $k \geq k_0$ and

$$q^{k-1} \leq \frac{r-\|\nabla f(x_\infty)\|}{\left(\frac{1}{h^2}+\frac{\beta L}{h}+L\frac{q}{1-q}\right)\|x_1-x_0\|},$$

which is satisfied for k large enough, because of $q<1$.

Proof.

(i) The assumption $-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0))$ implies the existence of $\varepsilon > 0$ such that

$$-\nabla f(x_\infty) + B(0,2\,\varepsilon) \subset \partial\phi(0).$$

On the other hand, since $\lim_k \nabla f(x_k) = \nabla f(x_\infty)$, there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$

$$\nabla f(x_k) \in \nabla f(x_\infty) + B(0,\varepsilon).$$

Hence,

$$-\nabla f(x_k) + B(0,\varepsilon) \subset -\nabla f(x_\infty) + B(0,2\,\varepsilon) \subset \partial\phi(0).$$

Equivalently, for every $k \geq k_0$ and for every $u \in B(0,1)$, we have:

$$-\nabla f(x_k) + \varepsilon\, u \in \partial\phi(0).$$

Let's write the corresponding subdifferential inequality at the origin (recall that $\phi(0)=0$)). For every $k \geq k_0$, we have

$$\forall u \in B(0,1), \quad \phi(\frac{1}{h}(x_{k+1}-x_k)) \geq \langle -\nabla f(x_k) + \varepsilon\, u, \frac{1}{h}(x_{k+1}-x_k)\rangle.$$

Taking the supremum over $u \in B(0,1)$, we obtain that, for every $k \geq k_0$,

$$\phi(\frac{1}{h}(x_{k+1}-x_k)) + \langle \nabla f(x_k), \frac{1}{h}(x_{k+1}-x_k)\rangle \geq \varepsilon \|\frac{1}{h}(x_{k+1}-x_k)\|. \tag{2.15}$$

Let's return to inequality (2.6). According to (2.9), we have

$$\begin{aligned}&\frac{1}{2}\|\frac{1}{h}(x_{k+1}-x_k)\|^2 - \frac{1}{2}\|\frac{1}{h}(x_k-x_{k-1})\|^2 + \left(\frac{\gamma}{h}-\frac{\beta L}{2h}\right)\|x_{k+1}-x_k\|^2\\ &+h\phi(\frac{1}{h}(x_{k+1}-x_k)) + \langle \nabla f(x_k), x_{k+1}-x_k\rangle \leq \frac{\beta L}{2h}\|x_k-x_{k-1}\|^2.\end{aligned}$$

According to $\gamma \geq L\left(\frac{h}{2}+\beta\right)$, we have $\gamma > \beta L$. Hence $\frac{\gamma}{h} - \frac{\beta L}{2h} = \frac{\beta L}{2h} + \frac{1}{h}(\gamma - \beta L) > \frac{\beta L}{2h}$, which gives

$$\begin{aligned}&\frac{1}{2}\|\frac{1}{h}(x_{k+1}-x_k)\|^2 - \frac{1}{2}\|\frac{1}{h}(x_k-x_{k-1})\|^2 + \frac{\beta L}{2h}\|x_{k+1}-x_k\|^2 + \frac{1}{h}(\gamma-\beta L)\|x_{k+1}-x_k\|^2\\ &+h\phi(\frac{1}{h}(x_{k+1}-x_k)) + \langle \nabla f(x_k), x_{k+1}-x_k\rangle \leq \frac{\beta L}{2h}\|x_k-x_{k-1}\|^2.\end{aligned}$$

Combining the inequality above with (2.15), we obtain, for every $k \geq k_0$

$$\begin{aligned}&\frac{1}{2}(1+\beta hL)\|\frac{1}{h}(x_{k+1}-x_k)\|^2 - \frac{1}{2}(1+\beta hL)\|\frac{1}{h}(x_k-x_{k-1})\|^2\\ &+\frac{1}{h}(\gamma-\beta L)\|x_{k+1}-x_k\|^2 + \varepsilon\|x_{k+1}-x_k\| \leq 0.\end{aligned} \tag{2.16}$$

Neglecting the nonnegative term $\varepsilon\|x_{k+1}-x_k\| \geq 0$ in the above inequality, we obtain

$$\frac{1}{2}(1+\beta hL)\|\frac{1}{h}(x_{k+1}-x_k)\|^2 - \frac{1}{2}(1+\beta hL)\|\frac{1}{h}(x_k-x_{k-1})\|^2 + \frac{1}{h}(\gamma-\beta L)\|x_{k+1}-x_k\|^2 \leq 0.$$

Equivalently

$$\Big(1+\beta hL+2h(\gamma-\beta L)\Big)\|x_{k+1}-x_k\|^2 \leq \Big(1+\beta hL\Big)\|x_k-x_{k-1}\|^2.$$

which gives the geometric convergence of velocities towards zero: for $k \geq k_0$

$$\|x_{k+1}-x_k\| \leq q^k\|x_1-x_0\| \tag{2.17}$$

with $q := \dfrac{1}{\sqrt{1+\frac{2h(\gamma-\beta L)}{1+\beta hL}}}$. Set $C := \|x_1-x_0\|$. For $p \geq 0$ we have

$$\|x_k - x_{k+p}\| \leq Cq^k(1+q+...+q^{p-1}) \leq C\frac{q^k}{1-q}.$$

By making p go to infinity in the inequality above, and using that (x_k) converges to x_∞, we obtain

$$\|x_k - x_\infty\| \leq C\frac{q^k}{1-q}.$$

This formula expresses the geometric convergence of the sequence (x_k) to its limit x_∞. This is a remarkable property because there can be a continuum of possible limits of the sequence.

(ii) Let us show that the finite convergence property holds under the assumption $\|\nabla f(x_\infty)\| < r$ where $B(0,r) \subset \partial\phi(0)$. Write the algorithm (IPAHDD) as follows:

$$\frac{1}{h}(x_{k+1}-x_k)+\gamma(x_{k+1}-x_k)+h\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right) \ni \frac{1}{h}(x_k-x_{k-1})-h\nabla f(x_k)$$
$$-\beta\left(\nabla f(x_k)-\nabla f(x_{k-1})\right). \tag{2.18}$$

Equivalently,

$$(1+\gamma h)\left(\frac{1}{h}(x_{k+1}-x_k)\right)+h\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right) \ni \xi_k,$$

where $\xi_k := \frac{1}{h}(x_k-x_{k-1})-h\nabla f(x_k)-\beta\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)$. Set $\lambda := \frac{h}{1+\gamma h}$. We have

$$\frac{1}{h}(x_{k+1}-x_k) = \mathrm{prox}_{\lambda\phi}\left(\frac{1}{1+\gamma h}\xi_k\right). \tag{2.19}$$

To show the finite convergence property, we need to show that $x_{k+1}-x_k=0$ for k large enough. According to (2.19) and Lemma 1.2, it suffices to prove that

$$\frac{1}{\lambda}\|\frac{1}{1+\gamma h}\xi_k\| \leq r, \tag{2.20}$$

which, by definition of λ gives $\frac{1}{h}\|\xi_k\| \leq r$. By the triangle inequality and the L-Lipschitz continuity of ∇f we have

$$\begin{aligned}\frac{1}{h}\|\xi_k\| &= \|\frac{1}{h^2}(x_k-x_{k-1})-\nabla f(x_k)-\frac{\beta}{h}\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)\| \\ &\leq \frac{1}{h^2}\|x_k-x_{k-1}\|+\|\nabla f(x_\infty\|+L\|x_k-x_\infty\|+\frac{\beta L}{h}\|x_k-x_{k-1}\|. \end{aligned}\tag{2.21}$$

When $k\to+\infty$, the right-hand side of the inequality (2.21) tends to $\|\nabla f(x_\infty)\|$. So, condition (2.20) will be satisfied for k large enough if $\|\nabla f(x_\infty)\|<r$. Let us suppose this condition satisfied, and further analyze (2.21). We will have $x_{k+1}-x_k=0$ as soon as

$$\left(\frac{1}{h^2}+\frac{\beta L}{h}\right)\|x_k-x_{k-1}\|+L\|x_k-x_\infty\| \leq r-\|\nabla f(x_\infty)\|.$$

According to the geometric convergence rate obtained in (i), this will be satisfied when $k\geq k_0$ and

$$\left(\frac{1}{h^2}+\frac{\beta L}{h}+L\frac{q}{1-q}\right)q^{k-1}\|x_1-x_0\| \leq r-\|\nabla f(x_\infty)\|.$$

This gives

$$q^{k-1} \leq \frac{r-\|\nabla f(x_\infty)\|}{\left(\frac{1}{h^2}+\frac{\beta L}{h}+L\frac{q}{1-q}\right)\|x_1-x_0\|},$$

which completes the proof. ■

Remark 2.1 Let's give another proof of the finite convergence property. On the one hand, it only assumes that $-\nabla f(x_\infty) \in \mathrm{int}(\partial\phi(0))$, but it is valid only when $\mathcal{H}$ is a finite dimensional space. It is similar to the argument developed by Baji-Cabot in [20]. Argue by contradiction, and suppose that there is an

infinite number of indices $k \in \mathbb{N}$ such that $\|x_{k+1} - x_k\| \neq 0$. Set $\mathcal{N} := \{k \in \mathbb{N} : \|x_{k+1} - x_k\| \neq 0\}$, and consider the sequence $(\omega_k)_k$ defined by

$$\omega_k := \frac{x_{k+1} - x_k}{\|x_{k+1} - x_k\|} \quad \text{for} \quad k \in \mathcal{N}.$$

The sequence (ω_k) belongs to the unit sphere of $\mathcal{H}$, and since $\mathcal{H}$ is assumed to have a finite dimension, we can extract a convergent sequence (still denoted (ω_k)) that converges to a point ω which belongs to the unit sphere (in an infinite dimensional space, we would only have weak convergence towards a point of the unit ball). According to the monotonicity property of $\partial\phi$ and the definition of the algorithm (IPAHDD), we have, for $k \in \mathcal{N}$

$$\left\langle a_k - \frac{\gamma}{h}(x_{k+1} - x_k) - \frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) - \nabla f(x_k) - \partial\phi(0), \omega_k \right\rangle \geq 0, \tag{2.22}$$

with $a_k = -\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1})$.
According to convergence properties shown above, by passing to the limit in (2.22), we obtain

$$\langle \nabla f(x_\infty) + \partial\phi(0), \omega \rangle \leq 0.$$

Since $-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0))$, there exists some $\rho > 0$ such that:

$$B(0, \rho) \subset \nabla f(x_\infty) + \partial\phi(0).$$

Therefore, we would have $\langle \rho u, \omega \rangle \leq 0$ for all $u \in B(0,1)$. Taking $u = \omega$ (since $\|\omega\| = 1$), gives $\rho\|\omega\|^2 \leq 0$, and hence $\omega = 0$, a clear contraction with ω belonging to the unit sphere.

Remark 2.2 The case $\phi = 0$ gives the heavy ball with friction method initiated by Polyak [32], [33]. This case is excluded from our analysis because of the dry friction hypothesis $(\mathrm{DF})_r$ on ϕ. We can advantageously compare our method with the heavy-ball method, for which the results of convergence require restrictive assumptions on the parameters and the function f, see [28] for a recent account on the heavy ball method. Note that, compared to the restart method, we get a geometric convergence for a general function f, which might be nonconvex.

2.4 Soft thresholding on the velocities

As a model situation for dry friction, take $\phi : \mathcal{H} \to \mathbb{R}$ given by $\phi(x) = r\|x\|$, with $r > 0$. We have

$$\partial\phi(x) = \begin{cases} r\frac{x}{\|x\|} & \text{if } x \neq 0; \\ B(0,r) & \text{if } x = 0. \end{cases} \tag{2.23}$$

By definition of the proximal operator, we obtain, for all $\lambda > 0$,

$$\operatorname{prox}_{\lambda\phi}(x) = \Big(1 - \frac{\lambda r}{\max\{\lambda r, \|x\|\}}\Big)x = \begin{cases} 0 & \text{if } \|x\| \leq \lambda r; \\ (\|x\| - \lambda r)\frac{x}{\|x\|} & \text{if } \|x\| \geq \lambda r. \end{cases} \tag{2.24}$$

a) When $\mathcal{H} = \mathbb{R}$, we get the classical soft thresholding operator $\operatorname{prox}_{\lambda\phi} = T_{\lambda r}$, which is used in the FISTA method for sparse optimization:

$$T_{\lambda r}(x) = \operatorname{sign}(x)(|x| - \lambda r)_+ = \begin{cases} x - \lambda r & \text{if } x \geq \lambda r; \\ 0 & \text{if } -\lambda r \leq x \leq \lambda r; \\ x + \lambda r & \text{if } x \leq -\lambda r. \end{cases} \tag{2.25}$$

b) In the multidimensional case $\mathcal{H} = \mathbb{R}^n$, take $\phi : \mathbb{R}^n \to \mathbb{R}$ given by $\phi(x) = r\|x\|_1 = r\sum_{i=1}^n |x_i|$. The proximal mapping of ϕ can be computed componentwise by applying the one-dimensional soft thresholding operator $T_{\lambda r}$ to each component. This is transparent from the variational formulation of the proximal operator: $\mathrm{prox}_{\lambda\phi}(x)$ is the solution of the minimization problem

$$\min_{\xi\in\mathbb{R}^n}\left\{\frac{1}{2}\|\xi - x\|^2 + \lambda r\|\xi\|_1\right\} = \min_{\xi_1\in\mathbb{R},\dots,\xi_n\in\mathbb{R}}\left\{\sum_i\left(\frac{1}{2}|\xi_i - x_i|^2 + \lambda r|\xi_i|\right)\right\}$$

which can be decomposed with respect to each component. Hence

$$\left(\mathrm{prox}_{\lambda r\|\cdot\|_1}(x)\right)_i = T_{\lambda r}(x_i) = \mathrm{sign}(x_i)(|x_i| - \lambda r)_+ \quad \text{for } i = 1, 2, \dots n. \tag{2.26}$$

The algorithm (IGAHDD) is a splitting algorithm which reads componentwise as follows: setting $x_k = (x_{k,i})_{i=1,2,\dots n}$, we have for $i = 1, 2, \dots n$

(IPAHDD) with soft thresholding on the velocities

Initialize: $x_0 \in \mathcal{H}$, $x_1 \in \mathcal{H}$

$for\ i = 1, 2, \dots n$

$$x_{k+1,i} = x_{k,i} + hT_{\frac{hr}{1+h\gamma}}\left(\frac{1}{h(1+h\gamma)}(x_{k,i} - x_{k-1,i}) - \frac{\beta}{1+h\gamma}\left(\frac{\partial f}{\partial x_i}(x_k) - \frac{\partial f}{\partial x_i}(x_{k-1})\right) - \frac{h}{1+h\gamma}\frac{\partial f}{\partial x_i}(x_k)\right)$$

$T_{\frac{hr}{1+h\gamma}}$ acts as a soft thresholding operator on the velocities. A direct application of Theorem 2.1 and Theorem 2.2 gives the following result:

Corollary 2.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and such that $\inf_{\mathcal{H}} f > -\infty$. Assume that the potential friction function ϕ is given by $\phi(x) = r\|x\|_1$. Suppose that the parameters h, γ, β in the algorithm* (IPAHDD) *satisfy the relation*

$$\gamma \geq L\left(\frac{h}{2} + \beta\right).$$

Let (x_k) be a sequence generated by the algorithm (IPAHDD) *with soft thresholding on the velocities.*

(i) *Then, the sequence (x_k) has a finite length and converges to x_∞ that verifies $\|\nabla f(x_\infty)\| \leq r$.*

(ii) *Suppose that $\|\nabla f(x_\infty)\| < r$. Then, there is geometric convergence, and the sequence (x_k) is finitely convergent.*

Clearly, taking r small is the interesting situation for optimization.

2.5 An example

Take $\mathcal{H} = \mathbb{R}$, $\phi(x) = r|x|$, and $f(x) = \frac{1}{2}x^2$. With $h = 1$, the algorithm (IPAHDD) reads as follows

$$(x_{k+1} - x_k) - (x_k - x_{k-1}) + \gamma(x_{k+1} - x_k) + \beta(x_k - x_{k-1}) + \partial\phi\,(x_{k+1} - x_k) + x_k \ni 0. \tag{2.27}$$

Equivalently, $(x_{k+1} - x_k) + \frac{1}{1+\gamma}\partial\phi\,(x_{k+1} - x_k) \ni -\frac{1}{1+\gamma}\left(x_{k-1} + \beta(x_k - x_{k-1})\right)$, which gives

$$x_{k+1} - x_k = T_{\frac{r}{1+\gamma}}\left(-\frac{1}{1+\gamma}x_{k-1} - \frac{\beta}{1+\gamma}(x_k - x_{k-1})\right). \tag{2.28}$$

According to (2.25), with $\lambda = \frac{1}{1+\gamma}$, we obtain

$$x_{k+1} - x_k = \begin{cases} -\frac{1}{1+\gamma}(x_{k-1} + r) - \frac{\beta}{1+\gamma}(x_k - x_{k-1}) & \text{if } x_{k-1} + \beta(x_k - x_{k-1}) \leq -r; \\ 0 & \text{if } |x_{k-1} + \beta(x_k - x_{k-1})| \leq r; \\ -\frac{1}{1+\gamma}(x_{k-1} - r) - \frac{\beta}{1+\gamma}(x_k - x_{k-1}) & \text{if } x_{k-1} + \beta(x_k - x_{k-1}) \geq r. \end{cases}$$

Take $r = 1$, $\gamma = 3$, $\beta = 1$. We have $L = 1$, and the condition $\gamma \geq L\left(\frac{h}{2} + \beta\right)$ of Theorem 2.2 is satisfied. So, as long as $x_k \geq 1$, according to the above formula, we have

$$x_{k+1} - x_k = -\frac{1}{1+\gamma}(x_{k-1} - r) - \frac{\beta}{1+\gamma}(x_k - x_{k-1}) = -\frac{1}{4}(x_k - 1).$$

The sequence $(x_k - 1)$ satisfies the geometric recurrence relation $x_{k+1} - 1 = \frac{3}{4}(x_k - 1)$, which gives

$$x_k = 1 + \left(\frac{3}{4}\right)^{k-1}(x_1 - 1).$$

By taking $x_1 \geq 1$, the condition $x_k \geq 1$ is satisfied. So, in this particular situation we have linear convergence but not finite convergence. This is in accordance with the fact that $x_\infty = 1$ and that $\nabla f(x_\infty) = 1$, which is not in the interior of the convex set $\partial\phi(0) = [-1, +1]$. Note that taking $\beta = h = 1$, as we did above, makes the study of the algorithm (2.27) quite simple, since, it this case, it reduces to the first-order algorithm

$$(1+\gamma)(x_{k+1} - x_k) + \partial\phi\left(x_{k+1} - x_k\right) + x_k \ni 0. \tag{2.29}$$

2.6 A variant

Consider the following discretization of the differential inclusion (IGDH)

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_k - x_{k-1}) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) + \nabla f(x_k) \ni 0, \tag{2.30}$$

where the temporal discretization of the viscous damping term is taken equal to $\frac{\gamma}{h}(x_k - x_{k-1})$ instead of $\frac{\gamma}{h}(x_{k+1} - x_k)$. Solving (2.30) with respect to x_{k+1} gives the following algorithm:

(IPAHDD-Var):

$Initialize : x_0 \in \mathcal{H},\ x_1 \in \mathcal{H}$

$$x_{k+1} = x_k + h\,\mathrm{prox}_{h\phi}\left(\left(\tfrac{1-h\gamma}{h}\right)(x_k - x_{k-1}) - \beta\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) - h\nabla f(x_k)\right).$$

We obtain the following convergence results that are parallel to Theorem 2.1 and Theorem 2.2.

Theorem 2.3 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and such that $\inf_{\mathcal{H}} f > -\infty$. Assume that the potential friction function ϕ satisfies $(\mathrm{DF})_r$. Suppose that the parameters h, γ in the algorithm* (IPAHDD-Var) *satisfy the relation*

$$\gamma \geq L\left(\beta + \frac{1}{2}h\right) + \frac{1}{2}\gamma^2 h.$$

Then, for any sequence (x_k) *defined by the algorithm* (IPAHDD-Var) *, we have:*

(i) $\sum_k \|x_{k+1}-x_k\| < +\infty$, *and* $\lim_k x_k := x_\infty$ *exists for the strong topology of* $\mathcal{H}$. *Moreover,*

$$\sum_{k=1}^{\infty} \|x_{k+1}-x_k\| \leq \frac{E_1}{r}$$

where $E_1 := \frac{1}{2}\Big(1+\beta hL\Big)\|\frac{1}{h}(x_1-x_0)\|^2 + f(x_1) - \inf_{\mathcal{H}} f$.

(ii) *The limit* x_∞ *of the sequence* (x_k) *satisfies:* $0 \in \partial\phi(0) + \nabla f(x_\infty)$.

Proof. (i) Without ambiguity, we write $\partial\phi$ to designate any element belonging to this set. Set $X_k := \frac{1}{h}(x_k - x_{k-1})$. Taking the dot product of (2.30) with $\frac{1}{h}(x_{k+1}-x_k)$, we obtain

$$\begin{aligned}&\langle X_{k+1}-X_k, X_{k+1}\rangle + \gamma h\,\langle X_{k+1}, X_k\rangle + h\,\langle \partial\phi\,(X_{k+1})\,, X_{k+1}\rangle \\ &+\beta\,\langle \nabla f(x_k) - \nabla f(x_{k-1}), X_{k+1}\rangle + \langle \nabla f(x_k), x_{k+1}-x_k\rangle = 0.\end{aligned} \tag{2.31}$$

The following elementary relation is related to the strong convexity of $\|\cdot\|^2$

$$\langle X_{k+1}-X_k, X_{k+1}\rangle = \frac{1}{2}\|X_{k+1}\|^2 - \frac{1}{2}\|X_k\|^2 + \frac{1}{2}\|X_{k+1}-X_k\|^2. \tag{2.32}$$

On the other hand

$$\langle X_{k+1}, X_k\rangle = \|X_{k+1}\|^2 - \langle X_{k+1}-X_k, X_{k+1}\rangle \geq \|X_{k+1}\|^2 - \|X_{k+1}\|\|X_{k+1}-X_k\|. \tag{2.33}$$

By convexity of ϕ, and $\phi(0)=0$

$$\langle \partial\phi\,(X_{k+1})\,, X_{k+1}\rangle \geq \phi\,(X_{k+1})\,. \tag{2.34}$$

Taking into account (2.32), (2.33) and (2.34), we deduce from (2.31) the following inequality

$$\begin{aligned}&\frac{1}{2}\|X_{k+1}\|^2 - \frac{1}{2}\|X_k\|^2 + \frac{1}{2}\|X_{k+1}-X_k\|^2 + \gamma h\|X_{k+1}\|^2 - \gamma h\|X_{k+1}\|\|X_{k+1}-X_k\| + h\phi(X_{k+1}) \\ &+\beta\left\langle \nabla f(x_k) - \nabla f(x_{k-1}), \frac{1}{h}(x_{k+1}-x_k)\right\rangle + \langle \nabla f(x_k), x_{k+1}-x_k\rangle \leq 0.\end{aligned} \tag{2.35}$$

Let's now use the assumptions on the potential functions ϕ and f. According to the assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1, for all $k \geq 1$

$$\phi\,(X_{k+1}) \geq r\|X_{k+1}\|. \tag{2.36}$$

Since ∇f is L-Lipschitz continuous, the classical gradient descent lemma gives, for all $k \geq 1$

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1}-x_k\rangle + \frac{L}{2}\|x_{k+1}-x_k\|^2. \tag{2.37}$$

According to the Cauchy-Schwarz inequality, and using again that ∇f is L-Lipschitz continuous,

$$\left|\left\langle \nabla f(x_k) - \nabla f(x_{k-1}), \frac{1}{h}(x_{k+1}-x_k)\right\rangle\right| \leq hL\|X_k\|\|X_{k+1}\| \leq \frac{hL}{2}(\|X_k\|^2 + \|X_{k+1}\|^2). \tag{2.38}$$

Combining inequalities (2.36)-(2.37)-(2.38) with (2.35), we obtain, for all $k \geq 1$

$$\begin{aligned}&\frac{1}{2}\|X_{k+1}\|^2 - \frac{1}{2}\|X_k\|^2 + \Big(\frac{1}{2}\|X_{k+1}-X_k\|^2 - \gamma h\|X_{k+1}\|\|X_{k+1}-X_k\|\Big) + \gamma h\|X_{k+1}\|^2 \\ &+r\|x_{k+1}-x_k\| + f(x_{k+1}) - f(x_k) - \frac{Lh^2}{2}\|X_{k+1}\|^2 \leq \frac{\beta hL}{2}(\|X_k\|^2 + \|X_{k+1}\|^2).\end{aligned} \tag{2.39}$$

According to the elementary inequality

$$\frac{1}{2}\|X_{k+1}-X_k\|^2-\gamma h\|X_{k+1}\|\|X_{k+1}-X_k\|\geq-\frac{1}{2}\gamma^2h^2\|X_{k+1}\|^2$$

we obtain

$$\begin{aligned}&\frac{1}{2}\|X_{k+1}\|^2-\frac{1}{2}\|X_k\|^2-\frac{1}{2}\gamma^2h^2\|X_{k+1}\|^2+\gamma h\|X_{k+1}\|^2+r\|x_{k+1}-x_k\|\\&+f(x_{k+1})-f(x_k)-\frac{Lh^2}{2}\|X_{k+1}\|^2\leq\frac{\beta hL}{2}(\|X_k\|^2+\|X_{k+1}\|^2).\end{aligned}\tag{2.40}$$

Equivalently

$$\frac{1}{2}\Big(1+2\gamma h-\gamma^2h^2-Lh^2-\beta hL\Big)\|X_{k+1}\|^2-\frac{1}{2}\Big(1+\beta hL\Big)\|X_k\|^2+r\|x_{k+1}-x_k\|+f(x_{k+1})-f(x_k)\leq0.\tag{2.41}$$

Let's assume that $1+2\gamma h-\gamma^2h^2-Lh^2-\beta hL\geq1+\beta hL$. After simplification, this gives

$$\gamma\geq L\left(\beta+\frac{1}{2}h\right)+\frac{1}{2}\gamma^2h,$$

which is our assumption. Under this condition, we get

$$\frac{1}{2}\Big(1+\beta hL\Big)\|X_{k+1}\|^2-\frac{1}{2}\Big(1+\beta hL\Big)\|X_k\|^2+r\|x_{k+1}-x_k\|+f(x_{k+1})-f(x_k)\leq0.\tag{2.42}$$

Thus, in terms of

$$E_k:=\frac{1}{2}\Big(1+\beta hL\Big)\|\frac{1}{h}(x_k-x_{k-1})\|^2+(f(x_k)-\inf f),$$

we have obtained

$$E_{k+1}-E_k+r\|x_{k+1}-x_k\|\leq0.\tag{2.43}$$

Using the nonnegativity of E_k, and $r>0$, we deduce from (2.43) that $\sum_{k=1}^{\infty}\|x_{k+1}-x_k\|\leq\frac{1}{r}E_1<+\infty$. Therefore, the strong limit of the sequence (x_k) exists. Set $x_\infty:=\lim x_k$, which ends item (i).

(ii) From $\sum_k\|x_{k+1}-x_k\|<+\infty$, we get immediately $\lim_k\|x_{k+1}-x_k\|=0$. This in turn implies

$$\lim_k\frac{1}{h^2}\left(x_{k+1}-2x_k+x_{k-1}\right)=\lim_k\frac{1}{h^2}\left((x_{k+1}-x_k)-(x_k-x_{k-1})\right)=0.$$

Moreover, since ∇f is continuous and (x_k) converges strongly to x_∞, we have $\lim_k\nabla f(x_k)=\nabla f(x_\infty)$. To pass to the limit on (2.30), rewrite it as follows:

$$-\frac{1}{h^2}(x_{k+1}-2x_k+x_{k-1})-\frac{\gamma}{h}(x_k-x_{k-1})-\frac{\beta}{h}\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)-\nabla f(x_k)\in\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right).$$

According to the above convergence results and the closedness of the graph of $\partial\phi$ in $\mathcal{H}\times\mathcal{H}$, we deduce that

$$-\nabla f(x_\infty)\in\partial\phi\left(0\right).$$

which ends the proof. ■

Remark 2.3 *Let us analyze the condition on the parameters*

$$\gamma\geq L\left(\beta+\tfrac{1}{2}h\right)+\tfrac{1}{2}\gamma^2h$$

under which the convergence properties of Theorem 2.3 are satisfied. First, it is a slightly stronger condition that in Theorem 2.1 where we assume that $\gamma \geq L\left(\beta + \frac{1}{2}h\right)$. *Second, it is satisfied when* γ *is taken large enough and* h *is taken small enough. For example take*

$$\gamma \geq 2\beta L + \sqrt{L} \quad \textit{and} \quad h \leq \frac{1}{\gamma}.$$

Theorem 2.4 *(geometric and finite convergence) Under the assumptions of Theorem 2.3, suppose that*

$$\gamma > L\left(\beta + \frac{1}{2}h\right) + \frac{1}{2}\gamma^2 h.$$

Let (x_k) *be a sequence generated by* (IPAHDD-variant)*, and let* x_∞ *be its limit.*

(i) *Suppose that*

$$-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0)).$$

Then, there is geometric convergence of the velocities to zero.

(ii) *Suppose that*

$$\|\nabla f(x_\infty)\| < r \quad \textit{where} \quad B(0,r) \subset \partial\phi(0).$$

Then the sequence (x_k) *is finitely convergent.*

Proof. Under the assumption $-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0))$, a similar argument as in Theorem 2.2 gives the existence of $\varepsilon > 0$, and $k_0 \in \mathbb{N}$ such that for every $k \geq k_0$,

$$\phi(\frac{1}{h}(x_{k+1} - x_k)) + \langle \nabla f(x_k), \frac{1}{h}(x_{k+1} - x_k)\rangle \geq \varepsilon \|\frac{1}{h}(x_{k+1} - x_k)\|.$$

Based on the above inequality, similar arguments as in the proof of theorem 2.3 give

$$\frac{1}{2}\Big(1 + \beta h L\Big)\|X_{k+1}\|^2 - \frac{1}{2}\Big(1 + \beta h L\Big)\|X_k\|^2 + h\left(\gamma - L\left(\beta + \frac{1}{2}h\right) - \frac{1}{2}\gamma^2 h\right)\|X_{k+1}\|^2 + \varepsilon\|x_{k+1} - x_k\| \leq 0.$$

By hypothesis, $\gamma - L\left(\beta + \frac{1}{2}h\right) - \frac{1}{2}\gamma^2 h > 0$. From this, we easily derive the geometric convergence. The proof of the linear convergence is very similar to the proof of Theorem 2.2. ■

3 Errors, perturbations

Let's introduce perturbations, errors in the algorithm (IPAHDD). According to the dynamic approach, we start from the perturbed version of (IGDH)

$$\ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(\dot{x}(t)) + \nabla f(x(t)) \ni e(t), \tag{3.1}$$

where the second member $e(\cdot)$ takes into account perturbations, errors. A similar temporal discretization as in section 2 gives

$$\begin{aligned}&\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right)\\ &+\frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) + \nabla f(x_k) \ni e_k.\end{aligned} \tag{3.2}$$

Solving (3.2) with respect to x_{k+1} gives the following algorithm

(IPAHDD-pert)

Initialize: $x_0 \in \mathcal{H}$, $x_1 \in \mathcal{H}$

$y_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f(x_k) + \frac{h}{1+h\gamma}e_k$

$x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}(y_k)$

The following convergence results parallel Theorem 2.1 and Theorem 2.2.

Theorem 3.1 *Let's make the assumptions of Theorem* 2.1*, and suppose that the sequence* (e_k) *of perturbations, errors satisfies:* $\lim_k \|e_k\| = 0$*. Then, for any sequence* (x_k) *defined by the algorithm* (IPAHDD-pert)*, we have:*

(i) $\sum_k \|x_{k+1} - x_k\| < +\infty$*, and therefore* $\lim x_k := x_\infty$ *exists for the strong topology of* $\mathcal{H}$*. Suppose that* $\|e_k\| \leq \frac{r}{2}$*. Then*

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{2}{r}\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right).$$

(ii) *The limit* x_∞ *of the sequence* (x_k) *satisfies:* $0 \in \partial\phi(0) + \nabla f(x_\infty)$.

(iii) *Suppose that* $-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0))$*. Then, there is geometric convergence of the velocities to zero. Set* $q = \frac{1}{\sqrt{1+2h\gamma}}$*. There exists* $k_0 \geq 0$ *such that for all* $k \geq k_0$

$$\|x_k - x_\infty\| \leq \frac{q^k}{1-q}\|x_1 - x_0\|.$$

(iv) *Suppose that* $\|\nabla f(x_\infty)\| < r$ *where* $B(0,r) \subset \partial\phi(0)$*. Then* (x_k) *is finitely convergent.*

Proof. The beginning of the proof is similar to that of Theorem 2.1, and uses the sequence (E_k) of energy functions

$$E_k := \frac{1}{2}\|\frac{1}{h}(x_k - x_{k-1})\|^2 + f(x_k) - \inf_{\mathcal{H}} f.$$

Taking the scalar product of (3.2) with $\frac{1}{h}(x_{k+1} - x_k)$, we obtain

$$\begin{aligned}
&\left\langle \frac{1}{h}(x_{k+1} - x_k) - \frac{1}{h}(x_k - x_{k-1}),\, \frac{1}{h}(x_{k+1} - x_k)\right\rangle + \gamma h\|\frac{1}{h}(x_{k+1} - x_k)\|^2 \\
&+h\left\langle \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right), \frac{1}{h}(x_{k+1} - x_k)\right\rangle + \beta\left\langle \nabla f(x_k) - \nabla f(x_{k-1}), \frac{1}{h}(x_{k+1} - x_k)\right\rangle \\
&+\langle \nabla f(x_k), x_{k+1} - x_k\rangle = \langle e_k, x_{k+1} - x_k\rangle. \qquad (3.3)
\end{aligned}$$

Set $X_k := \frac{1}{h}(x_k - x_{k-1})$. Using convex subdifferential inequalities, we obtain

$$\begin{aligned}
&\frac{1}{2}\|X_{k+1}\|^2 - \frac{1}{2}\|X_k\|^2 + \gamma h\|X_{k+1}\|^2 + h\phi(X_{k+1}) + \langle \nabla f(x_k), x_{k+1} - x_k\rangle \\
&+\beta\left\langle \nabla f(x_k) - \nabla f(x_{k-1}), \frac{1}{h}(x_{k+1} - x_k)\right\rangle \leq \langle e_k, x_{k+1} - x_k\rangle. \qquad (3.4)
\end{aligned}$$

According to the assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1, for all $k \geq 1$

$$\phi\left(X_{k+1}\right) \geq r\|X_{k+1}\|. \tag{3.5}$$

Since ∇f is L-Lipschitz continuous, the classical gradient descent gives, for all $k \geq 1$

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2. \tag{3.6}$$

According to the Cauchy-Schwarz inequality, and using again that ∇f is L-Lipschitz continuous,

$$\left| \left\langle \nabla f(x_k) - \nabla f(x_{k-1}), \frac{1}{h}(x_{k+1} - x_k) \right\rangle \right| \leq hL\|X_k\|\|X_{k+1}\| \leq \frac{hL}{2}(\|X_k\|^2 + \|X_{k+1}\|^2). \tag{3.7}$$

Combining inequalities (3.5)-(3.6)-(3.7) with (3.4), and using Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}&\frac{1}{2}\|\frac{1}{h}(x_{k+1} - x_k)\|^2 - \frac{1}{2}\|\frac{1}{h}(x_k - x_{k-1})\|^2 + \frac{\gamma}{h}\|x_{k+1} - x_k\|^2 + r\|x_{k+1} - x_k\| \\ &+f(x_{k+1}) - f(x_k) - \frac{L}{2}\|x_{k+1} - x_k\|^2 \leq \frac{\beta L}{2h}(\|x_k - x_{k-1}\|^2 + \|x_{k+1} - x_k\|^2) + \|e_k\|\|x_{k+1} - x_k\|.\end{aligned}$$

In terms of $E_k := \frac{1}{2}\|\frac{1}{h}(x_k - x_{k-1})\|^2 + (f(x_k) - \inf f)$, this is equivalent to

$$E_{k+1} - E_k + \left(\frac{\gamma}{h} - \frac{L}{2} - \frac{\beta L}{2h}\right)\|x_{k+1} - x_k\|^2 + (r - \|e_k\|)\|x_{k+1} - x_k\| \leq \frac{\beta L}{2h}\|x_k - x_{k-1}\|^2. \tag{3.8}$$

According to the assumption $\gamma \geq L\left(\frac{h}{2} + \beta\right)$, we have $\frac{\gamma}{h} - \frac{L}{2} - \frac{\beta L}{2h} \geq \frac{\beta L}{2h}$. Therefore,

$$E_{k+1} - E_k + (r - \|e_k\|)\|x_{k+1} - x_k\| + \frac{\beta L}{2h}\|x_{k+1} - x_k\|^2 \leq \frac{\beta L}{2h}\|x_k - x_{k-1}\|^2. \tag{3.9}$$

Set $\tilde{E}_k := E_k + \frac{\beta L}{2h}\|x_k - x_{k-1}\|^2$. We have

$$\tilde{E}_{k+1} - \tilde{E}_k + (r - \|e_k\|)\|x_{k+1} - x_k\| \leq 0. \tag{3.10}$$

Adding the above inequalities, and according to $E_k \geq 0$, and $e_k \to 0$, we deduce from (2.11) that

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| < +\infty.$$

So, the sequence (x_k) has a finite length, which implies that the strong limit of the sequence (x_k) exists. Set $x_\infty := \lim x_k$. Therefore, $\lim_k \|x_{k+1} - x_k\| = 0$, $\lim_k \frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) = 0$, and $\lim_k \nabla f(x_k) = \nabla f(x_\infty)$. To pass to the limit on (3.2), rewrite it as follows: $A_k \in \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right)$

$$\text{with } A_k = -\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h}(x_{k+1} - x_k) - \frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) - \nabla f(x_k) + e_k.$$

According to the above convergence results and the closedness of the graph of $\partial\phi$, we deduce that

$$-\nabla f(x_\infty) \in \partial\phi\left(0\right),$$

which gives item (i) and (ii).
The proof of (iii) and (iv) follows the lines of the proof of Theorem 2.2. Estimation (2.16) becomes

$$\begin{aligned}&\frac{1}{2}(1 + \beta hL)\|\frac{1}{h}(x_{k+1} - x_k)\|^2 - \frac{1}{2}(1 + \beta hL)\|\frac{1}{h}(x_k - x_{k-1})\|^2 + \frac{1}{h}\left(\gamma - \beta L\right)\|x_{k+1} - x_k\|^2 \\ &+\varepsilon\|x_{k+1} - x_k\| \leq \|e_k\|\|x_{k+1} - x_k\|.\end{aligned} \tag{3.11}$$

Since $\|e_k\| \to 0$, we obtain that for k sufficiently large

$$\frac{1}{2}(1+\beta hL)\|\frac{1}{h}(x_{k+1}-x_k)\|^2 - \frac{1}{2}(1+\beta hL)\|\frac{1}{h}(x_k-x_{k-1})\|^2$$
$$+\frac{1}{h}(\gamma-\beta L)\|x_{k+1}-x_k\|^2 + \frac{\varepsilon}{2}\|x_{k+1}-x_k\| \leq 0$$

According to $\gamma - \beta L > 0$, we easily deduce the geometric convergence of the sequence (x_k).
To prove the finite convergence, we return to the definition of the algorithm (IPAHDD-pert):

$$\frac{1}{h}(x_{k+1}-x_k) = \operatorname{prox}_{\lambda\phi}(\xi_k),$$

where $\lambda = \frac{h}{1+h\gamma}$, and ξ_k is given by

$$\xi_k := \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f(x_k) + \frac{h}{1+h\gamma}e_k.$$

According to Lemma 1.2, the finite convergence will result from the proof of the following inequality

$$\frac{1}{\lambda}\|\xi_k\| \leq r. \tag{3.12}$$

Since $\frac{1}{\lambda}\|\xi_k\| \to \|\nabla f(x_\infty)\|$, (3.12) will be satisfied for k large enough if $\|\nabla f(x_\infty)\| < r$. ■

4 Combining Nesterov acceleration method with dry friction

We will construct algorithms obtained by the temporal discretization of the differential inclusion

$$\text{(IGDH)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,$$

and which have an analogous structure to the Nesterov accelerated gradient method. Indeed, when discretizing (IGDH), there is some flexibility in the choice of the point y_k where the gradient is computed. Taking $y_k = x_k$, we get the algorithm (IPAHDD) studied in the previous sections. Taking $y_k = x_{k+1}$, we obtain a proximal algorithm that will be studied in the next section. Precisely, we consider the following temporal discretization of (IGDH)

$$\frac{1}{h^2}(x_{k+1}-2x_k+x_{k-1}) + \frac{\gamma}{h}(x_{k+1}-x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right) + \frac{\beta}{h}(\nabla f(x_k) - \nabla f(x_{k-1})) + \nabla f(y_k) \ni 0, \tag{4.1}$$

where y_k will be chosen consistently with the accelerated gradient method of Nesterov. To solve (4.1) with respect to $\frac{1}{h}(x_{k+1}-x_k)$, let's write it equivalently as

$$\frac{1}{h}(x_{k+1}-x_k) - \frac{1}{h}(x_k - x_{k-1}) + h\gamma\frac{1}{h}(x_{k+1}-x_k) + h\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right)$$
$$+\beta(\nabla f(x_k) - \nabla f(x_{k-1})) + h\nabla f(y_k) \ni 0. \tag{4.2}$$

Equivalently

$$(1+h\gamma)\frac{1}{h}(x_{k+1}-x_k) + h\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right) \ni \frac{1}{h}(x_k-x_{k-1}) - \beta(\nabla f(x_k) - \nabla f(x_{k-1})) - h\nabla f(y_k),$$

which gives

$$\begin{aligned}\frac{1}{h}(x_{k+1}-x_k)+\frac{h}{1+h\gamma}\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right) \;&\ni\; \frac{1}{h(1+h\gamma)}(x_k-x_{k-1})\\ &-\; \frac{\beta}{1+h\gamma}\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)-\frac{h}{1+h\gamma}\nabla f(y_k).\end{aligned}$$

Therefore

$$x_{k+1}=x_k+h\,\mathrm{prox}_{\frac{h}{1+h\gamma}\phi}\left(z_k\right), \tag{4.3}$$

$$\text{with}\quad z_k=\frac{1}{h(1+h\gamma)}(x_k-x_{k-1})-\frac{\beta}{1+h\gamma}\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)-\frac{h}{1+h\gamma}\nabla f(y_k).$$

When $\phi=0$, and $\beta=0$, the proximal operator is the identity, and we obtain

$$x_{k+1}=x_k+\frac{1}{1+h\gamma}(x_k-x_{k-1})-\frac{h^2}{1+h\gamma}\nabla f(y_k).$$

To recover the accelerated gradient method of Nesterov, we must take $y_k=x_k+\frac{1}{1+h\gamma}(x_k-x_{k-1})$. In doing so, we obtain the following algorithm:

(IPAHDD-N):

Initialize: $x_0\in\mathcal{H}$, $x_1\in\mathcal{H}$

$y_k=x_k+\frac{1}{1+h\gamma}(x_k-x_{k-1})$

$x_{k+1}=x_k+h\,\mathrm{prox}_{\frac{h}{1+h\gamma}\phi}\left(\frac{1}{h}(y_k-x_k)-\frac{\beta}{1+h\gamma}\left(\nabla f(x_k)-\nabla f(x_{k-1})\right)-\frac{h}{1+h\gamma}\nabla f(y_k)\right).$

Theorem 4.1 *Let $f:\mathcal{H}\to\mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and such that $\inf_{\mathcal{H}} f>-\infty$. Assume that the potential friction function ϕ satisfies $(\mathrm{DF})_r$. Suppose that the parameters h, γ, β in the algorithm* (IPAHDD-N) *satisfy the relation*

$$\gamma\geq\frac{3L}{2}(h+\beta)\quad\textit{and}\quad Lh^2\leq 1.$$

Then, for any sequence (x_k) defined by the algorithm (IPAHDD-N)*, we have:*

(i) $\sum_k\|x_{k+1}-x_k\|<+\infty$*, and hence $\lim_k x_k:=x_\infty$ exists for the strong topology of $\mathcal{H}$. Moreover,*

$$\sum_{k=1}^{\infty}\|x_{k+1}-x_k\|\leq\frac{1}{r}E_1$$

where $E_1=\frac{1}{2h^2}(1+h\gamma-\frac{Lh^2}{2})\|x_1-x_0\|^2+f(x_1)-\inf_{\mathcal{H}} f$.

(ii) *The limit x_∞ of the sequence (x_k) satisfies:* $0\in\partial\phi(0)+\nabla f(x_\infty)$.

Proof. (i) Taking the dot product of (4.2) with $\frac{1}{h}(x_{k+1}-x_k)$, we obtain

$$\begin{aligned}&\left\langle\frac{1}{h}(x_{k+1}-x_k)-\frac{1}{h}(x_k-x_{k-1}),\,\frac{1}{h}(x_{k+1}-x_k)\right\rangle+\gamma h\|\frac{1}{h}(x_{k+1}-x_k)\|^2\\ &+\langle\nabla f(y_k),x_{k+1}-x_k\rangle+h\left\langle\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right),\frac{1}{h}(x_{k+1}-x_k)\right\rangle\\ &+\beta\left\langle\nabla f(x_k)-\nabla f(x_{k-1}),\frac{1}{h}(x_{k+1}-x_k)\right\rangle=0.\end{aligned}\tag{4.4}$$

Set $X_k := \frac{1}{h}(x_k - x_{k-1})$. According to the assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1, for all $k \geq 1$

$$\langle \partial\phi\left(X_{k+1}\right), X_{k+1}\rangle \geq \phi\left(X_{k+1}\right) \geq r\|X_{k+1}\|.$$

According to the Cauchy-Schwarz inequality, and using that ∇f is L-Lipschitz continuous,

$$\left| \left\langle \nabla f(x_k) - \nabla f(x_{k-1}), \frac{1}{h}(x_{k+1} - x_k) \right\rangle \right| \leq hL\|X_k\|\|X_{k+1}\| \leq \frac{hL}{2}(\|X_k\|^2 + \|X_{k+1}\|^2).$$

Combining (4.4) with the two above inequalities, we obtain

$$\begin{aligned} &\langle X_{k+1} - X_k, X_{k+1}\rangle + \gamma h\|X_{k+1}\|^2 + hr\|X_{k+1}\| \\ &+ \langle \nabla f(y_k), x_{k+1} - x_k\rangle \leq \frac{\beta hL}{2}(\|X_k\|^2 + \|X_{k+1}\|^2). \end{aligned} \tag{4.5}$$

Using successively the gradient descent lemma for f, the L-Lipschitz continuity of ∇f, and the equality $y_k = x_k + \frac{1}{1+h\gamma}(x_k - x_{k-1})$ (by definition of (IPAHDD-N)), we get

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k\rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) + \langle \nabla f(y_k), x_{k+1} - x_k\rangle + L\|y_k - x_k\|\|x_{k+1} - x_k\| + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) + \langle \nabla f(y_k), x_{k+1} - x_k\rangle + \frac{L}{1+h\gamma}\|x_k - x_{k-1}\|\|x_{k+1} - x_k\| + \frac{L}{2}\|x_{k+1} - x_k\|^2 \end{aligned} \tag{4.6}$$

Combining the above inequality with (4.5), we obtain

$$\begin{aligned} &\langle X_{k+1} - X_k, X_{k+1}\rangle + \gamma h\|X_{k+1}\|^2 + hr\|X_{k+1}\| + f(x_{k+1}) - f(x_k) \\ &-\frac{Lh^2}{1+h\gamma}\|X_k\|\|X_{k+1}\| - \frac{Lh^2}{2}\|X_{k+1}\|^2 \leq \frac{\beta hL}{2}(\|X_k\|^2 + \|X_{k+1}\|^2). \end{aligned}$$

Therefore,

$$\begin{aligned} &(1 + h\gamma - \frac{Lh^2}{2} - \frac{\beta hL}{2})\|X_{k+1}\|^2 - (1 + \frac{Lh^2}{1+h\gamma})\|X_k\|\|X_{k+1}\| - \frac{\beta hL}{2}\|X_k\|^2 \\ &+hr\|X_{k+1}\| + f(x_{k+1}) - f(x_k) \leq 0. \end{aligned} \tag{4.7}$$

According to the assumptions $\gamma \geq \frac{3L}{2}(h + \beta)$, and $Lh^2 \leq 1$, we get $1 + h\gamma - \frac{Lh^2}{2} - \frac{\beta hL}{2} \geq 0$. From (4.7) we infer

$$\begin{aligned} &\frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2} - \frac{\beta hL}{2})(\|X_{k+1}\|^2 - \|X_k\|^2) + \frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2} - \frac{\beta hL}{2})\|X_{k+1}\|^2 \\ &+\frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2} - \frac{3\beta hL}{2})\|X_k\|^2 - (1 + \frac{Lh^2}{1+h\gamma})\|X_k\|\|X_{k+1}\| + hr\|X_{k+1}\| + f(x_{k+1}) - f(x_k) \leq 0. \end{aligned}$$

Therefore,

$$\begin{aligned} &\frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2} - \frac{\beta hL}{2})(\|X_{k+1}\|^2 - \|X_k\|^2) + \frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2} - \frac{3\beta hL}{2})\|X_{k+1}\|^2 \\ &+\frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2} - \frac{3\beta hL}{2})\|X_k\|^2 - (1 + \frac{Lh^2}{1+h\gamma})\|X_k\|\|X_{k+1}\| + hr\|X_{k+1}\| + f(x_{k+1}) - f(x_k) \leq 0. \end{aligned}$$

Elementary algebra (sign of a polynomial of the second degree) gives that the inequality

$$\frac{1}{2}(1+h\gamma-\frac{Lh^2}{2}-\frac{3\beta hL}{2})\|X_{k+1}\|^2-(1+\frac{Lh^2}{1+h\gamma})\|X_k\|\|X_{k+1}\|+\frac{1}{2}(1+h\gamma-\frac{Lh^2}{2}-\frac{3\beta hL}{2})\|X_k\|^2\geq 0$$

is satisfied under the condition

$$\Delta=\left(1+\frac{Lh^2}{1+h\gamma}\right)^2-\left(1+h\gamma-\frac{Lh^2}{2}-\frac{3\beta hL}{2}\right)^2\leq 0.$$

This is equivalent to

$$\frac{\gamma}{L}\geq\frac{h}{1+h\gamma}+\frac{h}{2}+\frac{3\beta}{2}.$$

Since $\frac{1}{2}+\frac{1}{1+h\gamma}\leq\frac{3}{2}$, we end up with the condition $\frac{\gamma}{L}\geq\frac{3}{2}(h+\beta)$, which is satisfied by assumption. To summarize the results, in terms of

$$E_k:=\frac{1}{2}(1+h\gamma-\frac{Lh^2}{2})\|\frac{1}{h}(x_k-x_{k-1})\|^2+(f(x_k)-\inf_{\mathcal{H}}f),$$

we have obtained

$$E_{k+1}-E_k+r\|x_{k+1}-x_k\|\leq 0. \tag{4.8}$$

According to the nonnegativity of E_k, and $r>0$, we deduce from (4.8) that

$$\sum_{k=1}^{\infty}\|x_{k+1}-x_k\|\leq\frac{1}{r}E_1<+\infty.$$

Therefore, the sequence (x_k) has a finite length, which implies that the strong limit of the sequence (x_k) exists. Set $x_\infty:=\lim x_k$, which ends item (i).

(ii) From $\sum_k\|x_{k+1}-x_k\|<+\infty$, we get immediately $\lim_k\|x_{k+1}-x_k\|=0$. This in turn implies

$$\lim_k\frac{1}{h^2}\left(x_{k+1}-2x_k+x_{k-1}\right)=\lim_k\frac{1}{h^2}\left((x_{k+1}-x_k)-(x_k-x_{k-1})\right)=0.$$

Moreover, since ∇f is continuous and (x_k) converges strongly to x_∞, we have

$$\lim_k\nabla f(x_k)=\nabla f(x_\infty).$$

According the L-Lipschitz continuity of ∇f, and $y_k-x_k=\frac{1}{1+h\gamma}(x_k-x_{k-1})$, we have

$$\|\nabla f(y_k)-\nabla f(x_k)\|\leq L\|y_k-x_k\|\leq\frac{L}{1+h\gamma}\|x_k-x_{k-1}\|.$$

Therefore,

$$\lim_k\nabla f(y_k)=\nabla f(x_\infty).$$

To pass to the limit in (4.1), rewrite it as follows:

$$-\frac{1}{h^2}(x_{k+1}-2x_k+x_{k-1})-\frac{\gamma}{h}(x_{k+1}-x_k)-\nabla f(y_k)\in\partial\phi\left(\frac{1}{h}(x_{k+1}-x_k)\right). \tag{4.9}$$

According to the above convergence results and the closedness of the graph of $\partial\phi$ in $\mathcal{H}\times\mathcal{H}$, we deduce that:

$$-\nabla f(x_\infty)\in\partial\phi\left(0\right),$$

which gives item (ii). Item (iii) is obtained by a similar argument as in Theorem 2.2. ■

4.1 A variant

Let's go back to (4.3), which is recalled below

$$x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}(z_k),$$

with $z_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f(y_k)$, and make a different choice of the extrapolated point y_k. Taking $y_k - x_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1})$ gives the algorithm

(IPAHDD-N-Var):

Initialize: $x_0 \in \mathcal{H}$, $x_1 \in \mathcal{H}$

$y_k = x_k + \frac{1}{h(1+h\gamma)}(x_k - x_{k-1})$

$x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}\left(y_k - x_k - \frac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f(y_k)\right).$

When $\phi = 0$ and $\beta = 0$, we obtain

$$x_{k+1} = x_k + \frac{1}{1+h\gamma}(x_k - x_{k-1}) - \frac{h^2}{1+h\gamma}\nabla f\left(x_k + \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}\right).$$

This corresponds to a variant of the Nesterov accelerated gradient method, with two different extrapolation coefficients $\alpha_{k,1} = \frac{1}{1+h\gamma}$ and $\alpha_{k,2} = \frac{1}{h(1+h\gamma)}$. This type of situation has been studied by Liang-Fadili-Peyré in [30]. Note that (IPAHDD-N) and its variant (IPAHDD-N-Var) rely on the discretization of (IGDH)

$$\begin{aligned}&\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right)\\ &+\frac{\beta}{h}(\nabla f(x_k) - \nabla f(x_{k-1})) + \nabla f(x_y) \ni 0,\end{aligned} \tag{4.10}$$

where y_k is chosen differently. In both cases, there exists a positive constant C such that

$$\|y_k - x_k\| \leq C\|x_k - x_{k-1}\|.$$

These are the main constitutive ingredients of the proof of Theorem 4.1. Therefore, similar convergence properties are valid for (IPAHDD-N-Var).

5 (IPAHDD) for nonsmooth functions

We assume that $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous and proper function such that $\inf f > -\infty$. The preceding sections deal with a differentiable function f, without convexity assumption on f. Now, when considering nonsmooth functions, we assume the convexity of f. This allows us to use the regularity properties of the Moreau envelope in the convex case. Indeed, to reduce to the previous situation, where $f : \mathcal{H} \to \mathbb{R}$ is a $\mathcal{C}^1$ function whose gradient is Lipschitz continuous, the idea is to replace f by its Moreau envelope. Recall some classical facts. For any $\lambda > 0$, the Moreau envelope of f of index λ is the function $f_\lambda : \mathcal{H} \to \mathbb{R}$ defined by: for all $x \in \mathcal{H}$,

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

The function f_λ is convex, of class $\mathcal{C}^{1,1}$, and such that $\inf_{\mathcal{H}} f_\lambda = \inf_{\mathcal{H}} f$, $\operatorname{argmin}_{\mathcal{H}} f_\lambda = \operatorname{argmin}_{\mathcal{H}} f$. One can consult [7, section 17.2.1], [21], [24] for an in-depth study of the properties of the Moreau envelope in a Hilbert framework. Since the infimal value and the set of minimizers are preserved by taking the Moreau envelope, the idea is to replace f by f_λ in the previous algorithm, and take advantage of the fact that f_λ is continuously differentiable. The algorithm (IPAHDD) becomes

$$x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}(y_k)$$

with

$$y_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(\nabla f_\lambda(x_k) - \nabla f_\lambda(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f_\lambda(x_k).$$

According to $\nabla f_\lambda(x) = \frac{1}{\lambda}\left(x - \operatorname{prox}_{\lambda f}(x)\right)$, we obtain

(IPAHDD-nonsmooth)

Initialize: $x_0 \in \mathcal{H}$, $x_1 \in \mathcal{H}$

$$y_k = \tfrac{1}{1+h\gamma}\left(\tfrac{\lambda-\beta h}{\lambda h}(x_k - x_{k-1}) + \tfrac{\beta}{\lambda}\left(\operatorname{prox}_{\lambda f}(x_k) - \operatorname{prox}_{\lambda f}(x_{k-1})\right) - \tfrac{h}{\lambda}\left(x_k - \operatorname{prox}_{\lambda f}(x_k)\right)\right)$$

$$x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}(y_k)$$

Note that the two nonsmooth functions f and ϕ enter the algorithm via their proximal mappings. In addition, these proximal steps are computed independently, which makes (IPAHDD-nonsmooth) a splitting algorithm. Based on the properties of the Moreau envelope, a direct adaptation of Theorem 2.1 gives the following convergence results for the algorithm (IPAHDD-nonsmooth).

Theorem 5.1 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous and proper function such that $\inf f > -\infty$. Assume that the potential friction function ϕ satisfies $(\mathrm{DF})_r$. Suppose that the parameters h, γ, β, λ in the algorithm* (IPAHDD-nonsmooth) *satisfy the relation*

$$\gamma \geq \frac{1}{\lambda}\left(\frac{h}{2} + \beta\right).$$

Then, for any sequence (x_k) defined by the algorithm (IPAHDD-nonsmooth)*, we have:*

(i) $\sum_k \|x_{k+1} - x_k\| < +\infty$. *Hence $\lim x_k := x_\infty$ exists for the strong topology of $\mathcal{H}$. Moreover,*

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r}\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right)$$

$$\sum_{k=1}^{\infty} \|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq 2h^2\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right),$$

where $E_1 := \frac{1}{2}\|\frac{1}{h}(x_1 - x_0)\|^2 + f(x_1) - \inf_{\mathcal{H}} f$.

(ii) *The limit x_∞ of the sequence (x_k) satisfies:* $0 \in \partial\phi(0) + \nabla f_\lambda(x_\infty)$.

Suppose moreover that

$$-\nabla f_\lambda(x_\infty) \in \operatorname{int}(\partial\phi(0)).$$

Then there is geometric convergence of the velocities to zero. Set $q := \dfrac{1}{\sqrt{1+\frac{2h(\gamma\lambda-\beta)}{\lambda+\beta h}}}$ *which satisfies* $0<q<1$*: there exists* $k_0 \geq 0$ *such that for all* $k \geq k_0$

$$\|x_{k+1}-x_k\| \leq q^k\|x_1-x_0\| \ \text{ and } \ \|x_k-x_\infty\| \leq \frac{q^k}{1-q}\|x_1-x_0\|.$$

(iii) *Suppose that*

$$\|\nabla f(x_\infty)\| < r \quad \text{where} \ \ B(0,r) \subset \partial\phi(0).$$

Then the sequence (x_k) *is finitely convergent. The iteration stops at* x_k *when* $k \geq k_0$ *and*

$$q^{k-1} \leq \frac{r-\|\nabla f(x_\infty)\|}{\left(\frac{1}{h^2}+\frac{\beta}{h\lambda}+\frac{q}{\lambda(1-q)}\right)\|x_1-x_0\|},$$

which is satisfied for k *large enough, because of* $q<1$.

Proof. The proof is immediate: replace f by f_λ in Theorem 2.1 and Theorem 2.2 , and use that ∇f_λ is $\frac{1}{\lambda}$-Lipschitz continuous. Taking $L=\frac{1}{\lambda}$, the condition $\gamma \geq L\left(\frac{h}{2}+\beta\right)$ becomes $\gamma \geq \frac{1}{\lambda}\left(\frac{h}{2}+\beta\right)$. ■

Remark 5.1 In the above approach, the parameter λ is fixed. Indeed, it could be possible to make it vary, but as a key property, it has to be bounded away from zero (because of the assumption $\lambda \geq \frac{1}{\gamma}\left(\frac{h}{2}+\beta\right)$). Thus our approach differs from the classical approximation method which consists approaching f by f_λ as λ goes to zero. In [17] a similar device has been used.

Remark 5.2 When using Moreau envelopes, besides the sequence (x_k), another sequence occurs naturally, namely (p_k) with $p_k = \operatorname{prox}_{\lambda f} x_k$. Since $\operatorname{prox}_{\lambda f}$ is a nonexpansive mapping, we have

$$\sum_k \|p_{k+1}-p_k\| \leq \sum_k \|x_{k+1}-x_k\| < +\infty.$$

Therefore, the sequence (p_k) has a finite length, it converges strongly to $p_\infty = \operatorname{prox}_{\lambda f} x_\infty$. Using the relation $\nabla f_\lambda(x_\infty) \in \partial f(p_\infty)$, we obtain the approximate optimality property:

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

6 Splitting algorithms for the Lasso-type problems

In many situations, the minimization problem has an additive composite structure $\min_{\mathcal{H}}(f+g)$, with f smooth and g nonsmooth. Accelerated proximal-gradient algorithms are effective splitting methods to treat these problems. We will show how to adapt the (IPAHDD) algorithm to such composite setting, in the case of the Lasso-type problems.
Take $\mathcal{H}=\mathbb{R}^n$ equipped with the usual Euclidean structure. Suppose that the function $f : \mathcal{H} \to \mathbb{R}\cup\{+\infty\}$ to be minimized has the additive structure

$$f(x) = \frac{1}{2}\|Ax-b\|_2^2 + g(x), \tag{6.1}$$

where $A \in \mathbb{R}^{m\times n}$ (with $m \leq n$), $b \in \mathbb{R}^m$ and $g \in \Gamma_0(\mathbb{R}^n)$ (set of closed proper and convex functions). Minimizing such function f occurs in a variety of fields ranging from inverse problems in signal/image processing, to machine learning and statistics. Typical examples of function g include the ℓ_1 norm (Lasso),

the $\ell_1-\ell_2$ norm (group Lasso), the total variation, or the nuclear norm (the ℓ_1 norm of the singular values of $x \in \mathbb{R}^{N\times N}$ identified with a vector in $\mathbb{R}^n$ with $n = N^2$). In all these such situations, g is nonsmooth which also makes f nonsmooth. A direct application of the algorithm (IPAHDD-nonsmooth) would require calculating (at least approximately) the proximal operator of f. It's not easy in general. To work around this difficulty, we use a change of metric. This technique was initiated by Lemaréchal and Sagasatizábal in [29] to introduce efficient preconditioners into the proximal point algorithm for minimizing convex functions, for recent developments see [2], [17], [26, Section 4.6]. For a symmetric and positive definite matrix $M \in \mathbb{R}^{n\times n}$, we denote by $\langle \cdot, \cdot \rangle_M = \langle M\cdot, \cdot\rangle$ the scalar product on $\mathbb{R}^n$ induced by M and by $\|\cdot\|_M$ the associated norm. For a given $f \in \Gamma_0(\mathbb{R}^n)$, the Moreau's envelope f_λ^M of index $\lambda > 0$ associated with the metric induced by M is defined by: for $x \in \mathbb{R}^n$

$$f_\lambda^M(x) = \min_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2\lambda}\|x-y\|_M^2 \right\}. \tag{6.2}$$

Let us denote by $\operatorname{prox}_{\lambda f}^M(x)$ the unique minimizer in (6.2), which is the proximal point of x, of index λ, for the metric induced by M. The first-order optimality condition for this strongly convex minimization problem gives

$$\operatorname{prox}_{\lambda f}^M(x) = (M + \lambda\partial f)^{-1}(Mx). \tag{6.3}$$

When $M = I_n$ (the identity matrix), we find the classical definitions. It is easy to prove that

$$\|\operatorname{prox}_{\lambda f}^M(x_1) - \operatorname{prox}_{\lambda f}^M(x_2)\| \leq \frac{\mu_{\max}(M)}{\mu_{\min}(M)}\|x_1 - x_2\|,$$

where $\mu_{\max}(M)$ and $\mu_{\min}(M)$ are respectively the largest and the smallest eigenvalue of M. The Moreau envelope f_λ^M is of class $C^{1,1}$ and its gradient for the Euclidean structure is given by

$$\nabla f_\lambda^M(x) = \frac{1}{\lambda}M\Big(x - \operatorname{prox}_{\lambda f}^M(x)\Big). \tag{6.4}$$

As a classical result, ∇f_λ^M is $\frac{1}{\lambda}$-Lipschitz continuous for the norm $\|\cdot\|_M$. From this, by using classical linear algebra, we easily deduce that

$$\|\nabla f_\lambda^M(x_1) - \nabla f_\lambda^M(x_2)\| \leq \frac{1}{\lambda}\sqrt{\frac{\mu_{\max}(M)}{\mu_{\min}(M)}}\|x_1 - x_2\|, \ \ \forall x_1,\ x_2 \in \mathbb{R}^n. \tag{6.5}$$

On the other hand, one can check easily that

$$\operatorname{argmin}(f_\lambda^M) = \operatorname{argmin}(f).$$

With the particular choice of f in (6.1), we set $M = I_n - \lambda A^TA$. If $\lambda \in [0, \frac{1}{\|A\|_2^2}[$, then M is positive definite. In this case,

$$\operatorname{prox}_{\lambda f}^M(x) = \operatorname{prox}_{\lambda g}\Big(x - \lambda A^T(Ax - b)\Big). \tag{6.6}$$

Note that formula (6.6) for the composite optimization problem (6.1) was given in [26, Section 4.6 page 190]. Using (6.4) and (6.6), we get

$$\nabla f_\lambda^M(x) = \frac{1}{\lambda}M\Big(x - \operatorname{prox}_{\lambda g}\Big(x - \lambda A^T(Ax - b)\Big)\Big). \tag{6.7}$$

Replacing f with f_λ^M in (IPGDF), we obtain the following splitting algorithm applicable to (6.1):

(IPAHDD) for the Lasso problem

Initialize: $x_0 \in \mathbb{R}^n$, $x_1 \in \mathbb{R}^n$, $M = I_n - \lambda A^T A$, $0 < \lambda \|A\|_2^2 < 1$

$z_k = \frac{1}{\lambda} M\Big(x_k - \mathrm{prox}_{\lambda g}\Big(x_k - \lambda A^T(Ax_k - b)\Big)\Big)$

$y_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(z_k - z_{k-1}) - \frac{h}{1+h\gamma} z_k.$

$x_{k+1} = x_k + h\,\mathrm{prox}_{\frac{h}{1+h\gamma}\phi}(y_k)$

For the LASSO problem, $g(x) = \|x\|_1$, formula (2.26) can be used to compute $\mathrm{prox}_{\lambda g}$.

Theorem 6.1 *Assume that the potential friction function ϕ satisfies $(\mathrm{DF})_r$. Suppose that the parameters h, γ, β, λ in the algorithm* (IPAHDD) *for the Lasso problem satisfy the relation*

$$\gamma \geq \frac{1}{\lambda}\sqrt{\frac{\mu_{\max}(M)}{\mu_{\min}(M)}}\left(\frac{h}{2} + \beta\right).$$

Then, for any sequence (x_k) defined by the algorithm (IPAHDD)) *for the Lasso problem, we have $\sum_k \|x_{k+1} - x_k\| < +\infty$, and therefore $\lim x_k := x_\infty$ exists for the strong topology of $\mathcal{H}$. Moreover,*

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r}\left(E_1 + \frac{\beta L}{2h}\|x_1 - x_0\|^2\right)$$

where $E_1 := \frac{1}{2}\|\frac{1}{h}(x_1 - x_0)\|^2 + (f(x_1) - \inf_{\mathcal{H}} f)$. The limit x_∞ satisfies $\|A^T(Ax_\infty - b) + \partial g(x_\infty)\| \leq r$.

7 Some numerical experiments

In this section, we perform some numerical tests to compare the four algorithms IPAHDD, IPAHDD-Var, IPAHDD-N, and IPAHDD-N-Var defined in the last sections. We use the *performance profiles* developed by Dolan-Moré [27] as a tool for comparing the solvers. The performance profiles give for each $t \in \mathbb{R}$, the proportion $\rho_s(t)$ of test problems on which each solver s under comparison has a performance within the factor t of the best possible ratio. For more details, we refer to [27].

To compare these algorithms, we choose the number of iterations and the cputime found by each solver as a performance measure. The function $\phi : \mathbb{R}^n \to \mathbb{R}$ is given by $x \mapsto \phi(x) = r\|x\|_2$ with $r = 0.1$, while the functions $f : \mathbb{R}^n \to \mathbb{R}$ are quadratic of the form $f(x) = \frac{1}{2}\|Ax - b\|^2$, $A \in \mathbb{R}^{m\times n}$ (with $m \leq n$) and $b \in \mathbb{R}^m$ are chosen randomly. The matrices A in our set of tests come from the SuiteSparse Matrix Collection[1]. We have chosen a set P of 42 different problems with matrices $A \in \mathbb{R}^{m\times n}$ size ranging from $m = 24$ to $m = 1309$ and from $n = 1309$ to $n = 1706$. The numerical experiments are carried out in an iMac with Mac OS 10.14 and a processor 3.2 GHz Intel Xeon W with 64Go memory. All the codes are written and executed in Matlab R2018b. We use the same initial points and the same stopping criterion *i.e.* either the number of iterations exceeds 10^5 or $\|\nabla f(x_k)\| \leq r$. We observe that both solvers are robust and solved all problems. The algorithm IPAHDD-N-Var is the most efficient. The algorithms IPAHDD and IPAHDD-Var solve 80% of the problems in the interval $[0, 1.5]$, while IPAHDD-N is robust after $t \geq 4.5$.

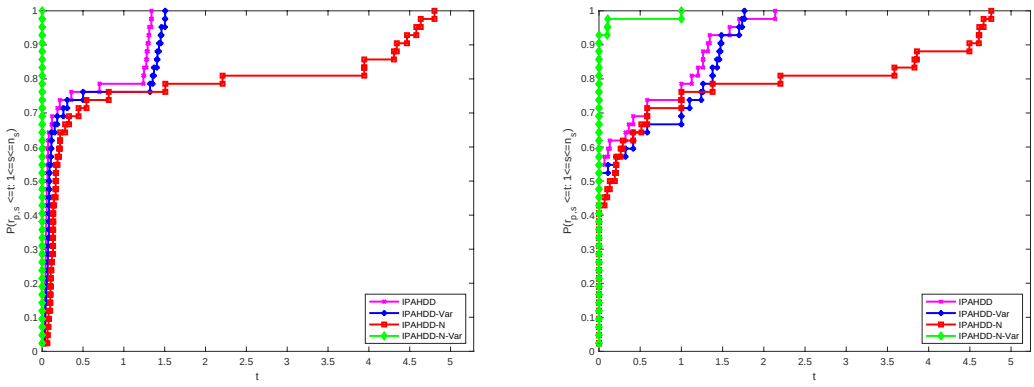

Figure 1: Performance profiles with $t_{p,s}$ the number of iterations (left) and cputime (right).

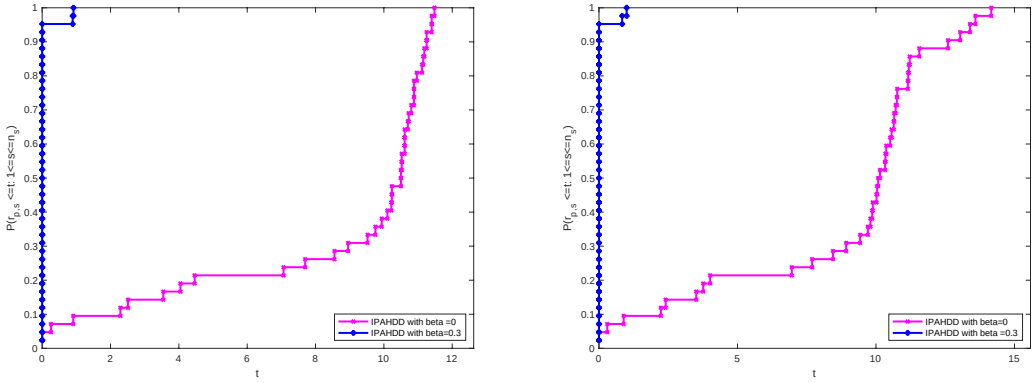

Figure 2: Performance profiles for IPAHDD with $t_{p,s}$ the number of iterations (left) and cputime (right) with and without the Hessian-driven viscous damping.

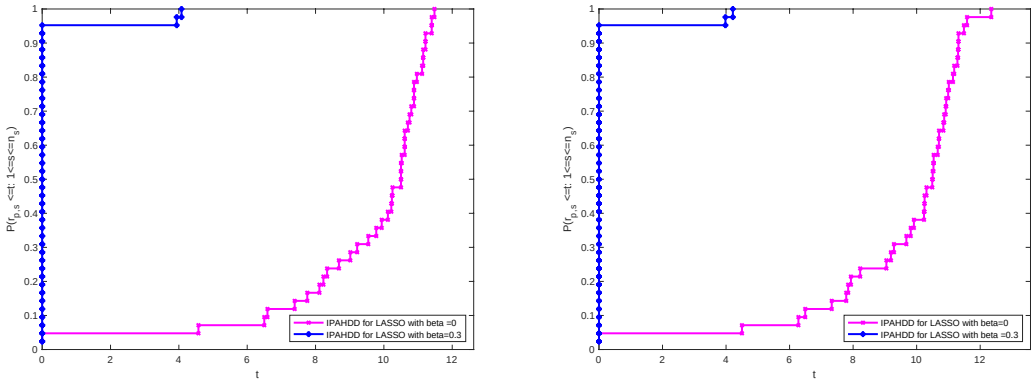

Figure 3: Performance profiles for IPAHDD for the LASSO problem with $t_{p,s}$ the number of iterations (left) and cputime (right) with and without the Hessian-driven viscous damping.

We conclude that, using the same initial points and under the same stopping criteria, IPAHDD-N-Var is the winner, followed by IPAHDD and IPAHDD-Var. In order to measure the effect of the introduction of the Hessian-driven viscous damping β, we also tested the four algorithms with $\beta = 0$ and $\beta = 0.3$. We observe that for all four methods, the introduction of the Hessian-driven viscous damping $\beta > 0$ has favorable effects not only for the convergence of the algorithm but also for the acceleration of the convergence. We only reported the performance profiles on the solver IPAHDD with $\beta = 0$ and $\beta = 0.3$ (see Figure 2), the results for the other solvers are very similar. We also compared the algorithm IPAHDD for the LASSO problem with $g(x) = \|x\|_1$ where formula (2.26) is used to compute $\mathrm{prox}_{\lambda g}$. Figure 3 shows the effect of the introduction of the Hessian-driven viscous damping. Consistent with the theoretical part, we observe that the introduction of both the dry friction coefficient $r > 0$ and the Hessian-driven viscous damping coefficient $\beta > 0$ introduces some stability, robustness and acceleration of the convergence in the numerical algorithms studied in this paper.

8 Perspectives

Let's list some of the many directions of research for the future:

1. The algorithm (IPAHDD) works with a general smooth function f, without any convexity assumption on f. We have been able to extend our study to the case of a nonsmooth convex function, by using the properties of the Moreau envelope in the convex case. For a general nonconvex nonsmooth function, a natural idea would be to use the Lasry-Lions regularization for which similar regularity properties are valid. The price to pay would be to perform a two-step proximal procedure.

2. Another remarkable property of the algorithm (IPAHDD) is its robustness and its tolerance to perturbations and errors. This naturally suggests developing corresponding stochastic gradient methods, combining dry friction with Hessian damping.

3. Many applications in optimization involve a composite additive structure. We have developed our method in the case of Lasso-type problems. It would be interesting to further develop the method in order to capture a larger class of composite problems.

4. Our analysis of (IPAHDD) crucially depends on the fact that the viscous damping coefficient γ is fixed, or more generally that it is bounded from below by a positive constant. It would be interesting to consider the case where it tends to zero like $\frac{\alpha}{t}$, $\alpha \geq 3$. This would allow dry friction to be considered together with Nesterov's fast gradient method.

A Auxiliary results

A.1 Finite time convergence of the continuous dynamic

Let's analyze the asymptotic behavior as $t \to +\infty$, and the finite convergence property, of the solution trajectories of the second-order differential inclusion

$$\text{(IGDH)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,\ t \in [t_0, +\infty[$$

which is at the origin of the algorithms studied in the paper. We take for granted the existence and uniqueness, for given initial data $x(t_0) = x_0,\ \dot{x}(t_0) = x_1$, of the solution trajectory of the corresponding Cauchy problem.

[1] https://sparse.tamu.edu

Theorem A.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^2$ function whose gradient is L-Lipschitz continuous, and let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function that satisfies $(\mathrm{DF})_{\mathrm{r}}$ and which is bounded on the bounded sets. Suppose that*

$$\gamma > \beta L.$$

Then, for any solution trajectory $x(\cdot)$ of (IGDH) *we have:*

(i) $\|\dot{x}\| \in L^1([t_0, +\infty[, \mathbb{R})$, and therefore the strong limit $x_\infty := \lim_{t\to+\infty} x(t)$ exists.

(ii) The limit point x_∞ is an equilibrium point of (IGDH)*, i.e. $-\nabla f(x_\infty) \in \partial\phi(0)$.*

(iii) If $-\nabla f(x_\infty) \notin$ boundary$(\partial\phi(0))$, then there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$.

Proof. (i) Take the scalar product of (IGDH) with $\dot{x}(t)$. We obtain

$$\langle \ddot{x}(t), \dot{x}(t)\rangle + \gamma\|\dot{x}(t)\|^2 + \langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle + \beta\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle + \langle \nabla f(x(t)), \dot{x}(t)\rangle = 0, \quad \text{(A.1)}$$

which gives

$$\frac{1}{2}\frac{d}{dt}\|\dot{x}(t)\|^2 + \gamma\|\dot{x}(t)\|^2 + \langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle + \beta\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle + \frac{d}{dt}(f(x(t) - inf_{\mathcal{H}} f) = 0.$$

According to the L-Lipschitz continuity of ∇f, and the Cauchy-Schwarz inequality, we have

$$|\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle| \leq L\|\dot{x}(t)\|^2.$$

According to the assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1,

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

Collecting the above results, we obtain

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t)) - inf_{\mathcal{H}} f\right) + (\gamma - \beta L)\|\dot{x}(t)\|^2 + r\|\dot{x}(t)\| \leq 0. \quad \text{(A.2)}$$

According to the hypothesis $\gamma > \beta L$, we deduce that the global energy

$$E(t) = \frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t) - inf_{\mathcal{H}} f$$

is non increasing. Morever, by integrating (A.2) we obtain

$$\int_{t_0}^{\infty} \|\dot{x}(t)\|^2 dt < +\infty \quad \text{and} \quad \int_{t_0}^{\infty} \|\dot{x}(t)\| dt < +\infty. \quad \text{(A.3)}$$

This last property expresses that the trajectory has a finite length, and hence $\lim_{t\to+\infty} x(t) := x_\infty$ exists.

(ii) Since $E(\cdot)$ is non increasing, we have that the velocity $\dot{x}(t)$ remains bounded. Since ϕ is bounded on the bounded sets, so is $\partial\phi$. Therefore, from equation (IGDH) we deduce that the acceleration $\ddot{x}(t)$ remains bounded. This combined with $\int_{t_0}^{\infty} \|\dot{x}(t)\| dt < +\infty$ implies that the velocity $\dot{x}(t)$ converges strongly to zero, as $t \to +\infty$. Let us now pass to the limit on (IGDH). Set $u(t) = \dot{x}(t)$. Let us write (IGDH) equivalently as

$$\dot{u}(t) + (\gamma I + \partial\phi)(u(t)) = h(t)$$

with $h(t) := -\beta\nabla^2 f(x(t))\dot{x}(t) - \nabla f(x(t))$. The operator $A = \gamma I + \partial\phi$ is strongly monotone because $\gamma > 0$. According to the above results, we have that $h(t)$ converges strongly to $-\nabla f(x_\infty)$. We now

apply the Theorem 3.9 of Brezis [24], which tells us that the strong limit of $u(t)$, that's zero, satisfies $A(0) \ni -\nabla f(x_\infty)$. Equivalently

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

(iii) The assumption $-\nabla f(x_\infty) \in \operatorname{int}(\partial\phi(0))$ implies the existence of $\varepsilon > 0$ such that

$$-\nabla f(x_\infty) + B(0, 2\,\varepsilon) \subset \partial\phi(0).$$

On the other hand, since $\lim_{t\to+\infty} \nabla f(x(t)) = \nabla f(x_\infty)$, there exists $t_1 \geq t_0$ such that for all $t \geq t_1$

$$\nabla f(x(t)) \in \nabla f(x_\infty) + B(0, \varepsilon).$$

Hence,

$$-\nabla f(x(t)) + B(0, \varepsilon) \subset -\nabla f(x_\infty) + B(0, 2\,\varepsilon) \subset \partial\phi(0).$$

Equivalently, for every $t \geq t_1$ and for every $u \in B(0, 1)$, we have:

$$-\nabla f(x(t))) + \varepsilon\, u \in \partial\phi(0).$$

Let's write the corresponding subdifferential inequality at the origin (recall that $\phi(0) = 0$)). For every $t \geq t_1$

$$\forall u \in B(0, 1), \quad \phi(\dot{x}(t)) \geq \langle -\nabla f(x(t)) + \varepsilon\, u, \dot{x}(t)\rangle.$$

Taking the supremum over $u \in B(0, 1)$, we obtain that, for every $t \geq t_1$,

$$\phi(\dot{x}(t)) + \langle \nabla f(x(t)), \dot{x}(t)\rangle \geq \varepsilon\, \|\dot{x}(t)\|. \tag{A.4}$$

Let's return to (A.1). According to the above results, we obtain

$$\frac{d}{dt}\frac{1}{2}\|\dot{x}(t)\|^2 + (\gamma - \beta L)\|\dot{x}(t)\|^2 + \varepsilon\|\dot{x}(t)\| \leq 0. \tag{A.5}$$

a) Neglecting the nonnegative term $\varepsilon\|\dot{x}(t)\|$ we obtain

$$\frac{d}{dt}\frac{1}{2}\|\dot{x}(t)\|^2 + (\gamma - \beta L)\|\dot{x}(t)\|^2 \leq 0, \tag{A.6}$$

whose integration gives

$$\|\dot{x}(t)\| \leq \|\dot{x}(t_0)\|e^{-(\gamma-\beta L)t}.$$

b) Neglecting the nonnegative term $(\gamma - \beta L)\|\dot{x}(t)\|^2$ we obtain

$$\frac{d}{dt}\|\dot{x}(t)\|^2 + 2\varepsilon\|\dot{x}(t)\| \leq 0, \tag{A.7}$$

Set $v(t) = \|\dot{x}(t)\|^2$. We have $\dot{v}(t) + 2\varepsilon\sqrt{v(t)} \leq 0$. As long as $v(t) > 0$ we will have $\frac{d}{dt}\sqrt{v(t)} \leq -\epsilon$. This forces $v(t)$ to be equal to zero after some finite time. ■

Remark A.1 *There are significant differences bewteen the continuous and the discrete case. The most important is that in the continuous case, when f is assumed to be convex, there is no more restrictive assumption on the parameters, since $\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle \geq 0$.*

References

[1] S. Adly, *A variational approach to nonsmooth dynamics: applications in unilateral mechanics and electronics*, Springer Briefs in Mathematics, 2017.

[2] S. Adly, H. Attouch, *Finite convergence of proximal-gradient inertial algorithms with dry friction damping*, (2019) `https://hal.archives-ouvertes.fr/hal-02388038`.

[3] S. Adly, H. Attouch, A. Cabot, *Finite time stabilization of nonlinear oscillators subject to dry friction*, Nonsmooth Mechanics and Analysis, Adv. Mech. Math., 12 (2006), Springer, New York, pp. 289–304.

[4] F. Alvarez, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM J. Control Optim., **38** (4) (2000), pp. 1102–1119.

[5] F. Álvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping*, J. Math. Pures Appl., 81 (8) (2002), pp. 747–779.

[6] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule*, HAL-01551873, (2017), to appear in Mathematical Programming.

[7] H. Attouch, G. Buttazzo, G. Michaille, *Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization.* Second Edition, MOS/SIAM Series on Optimization, MO 17, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, (2014).

[8] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, 263 (2017), pp. 5412–5458.

[9] H. Attouch, A. Cabot, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., 28 (1) (2018), pp. 849–874.

[10] H. Attouch, A. Cabot, Z. Chbani, H. Riahi, *Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient*, Evolution Equations and Control Theory, 7 (3) (2018), pp. 353–371.

[11] H. Attouch, Z. Chbani, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim., 29 (3) 2019, pp. 2227–2256.

[12] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Math. Program. Ser. B 168 (2018), pp. 123–175.

[13] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case* $\alpha \leq 3$, arXiv:1706.05671v1 [math.OC] 2017, ESAIM-COCV (2019) published electronically.

[14] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system*, Commun. Contemp. Math., 2 (1) (2000), pp. 1–34.

[15] H. Attouch, J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$, SIAM J. Optim., 26 (3) (2016), pp. 1824–1834.

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

[17] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, 2019. hal-02193846.

[18] J.-F. Aujol, Ch. Dossal, *Stability of over-relaxations for the Forward-Backward algorithm, application to FISTA*, SIAM J. Optim., 25 (4) (2015), pp. 2408–2433.

[19] J.-F. Aujol, Ch. Dossal, *Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for* $b > 0$, 2017, https://hal.inria.fr/hal-01547251v2.

[20] B. Baji, A. Cabot, *An inertial proximal algorithm with dry friction: finite convergence results*, Set-Valued Analysis, 9 (1) (2006), pp. 1–23.

[21] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).

[22] R. I. Bot, E. R. Csetnek, S.C. László, *A second order dynamical approach with variable damping to nonconvex smooth minimization*, to appear in Applicable Analysis, (2018).

[23] R. I. Bot, E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., 54 (3) (2016), pp. 1423–1443.

[24] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, 1972.

[25] A. Chambolle, Ch. Dossal, *On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm*, J. Optim. Theory Appl., 166 (2015), pp. 968–982.

[26] A. Chambolle, T. Pock, *An introduction to continuous optimization for imaging*, Acta Numerica, 25 (2016), pp. 161-319.

[27] E. D. Dolan, J. J. Moré, *Benchmarking Optimization Software with Performance Profiles*, Math. Program., 91 (2002), pp. 201–213.

[28] E. Ghadimi, H. R. Feyzmahdavian, M. Johansson, *Global convergence of the heavy-ball method for convex optimization*, in 2015 European Control Conference, July 2015, pp. 310–315.

[29] C. Lemaréchal, C. Sagastizábal. *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries.* SIAM J. Optim. 7 (1997), no. 2, 367-385.

[30] J. Liang, J. Fadili, G. Peyré, *Local linear convergence of forward-backward under partial smoothness*, Advances in Neural Information Processing Systems, 2014, pp. 1970–1978.

[31] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Math., 41 (3) (2017), pp. 681–685.

[32] B.T. Polyak, *Some methods of speeding up the convergence of iterative methods*, Z. Vylist Math. Fiz., 4 (1964), pp. 1–17.

[33] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).

[34] B. Shi, S. S. Du, M. I. Jordan, W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, arXiv:submit/2440124[cs.LG] 21 Oct 2018.

[35] J. W. Siegel, *Accelerated first-order methods: Differential equations and Lyapunov functions*, arXiv: Optimization and Control (math.OC):1903.05671v1 (2019).

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