



HAL
open science

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

Samir Adly, Hedy Attouch

► **To cite this version:**

Samir Adly, Hedy Attouch. Finite convergence of proximal-gradient inertial algorithms with dry friction damping. 2019. hal-02388038

HAL Id: hal-02388038

<https://hal.science/hal-02388038>

Preprint submitted on 30 Nov 2019

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

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



Distributed under a Creative Commons Attribution 4.0 International License

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

Samir ADLY* and Hedy ATTOUCH†

ABSTRACT. In a Hilbert space \mathcal{H} , based on inertial dynamics with dry friction damping, we introduce a new class of proximal-gradient algorithms with finite convergence properties. The function $f : \mathcal{H} \rightarrow \mathbb{R}$ to minimize is supposed to be differentiable (not necessarily convex), and enters the algorithm via its gradient. The dry friction damping function $\phi : \mathcal{H} \rightarrow \mathbb{R}_+$ 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. This algorithm naturally occurs as a discrete temporal version of an inertial differential inclusion involving viscous and dry friction together. 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 new 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.

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

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

1 Introduction and preliminary results

Throughout the paper \mathcal{H} is a real Hilbert space, endowed with the scalar product $\langle \cdot, \cdot \rangle$ and the associated norm $\|\cdot\|$, and $f : \mathcal{H} \rightarrow \mathbb{R}$ is a \mathcal{C}^1 function whose gradient is Lipschitz continuous. Several extensions of these hypothesis will be examined a little further in the paper, including the case where f is nonsmooth. We will analyze the finite convergence (within a finite number of steps) of several algorithms that can be obtained by temporal discretization of the differential inclusion

$$(HBDF) \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0, \quad t \in [t_0, +\infty[\quad (1.1)$$

where (HBDF) stands shortly for Heavy Ball system with Dry Friction. As a specific property, this inertial system combines two types of damping:

a) The term $\gamma(t)\dot{x}(t)$ in (1.1) models viscous damping with a positive coefficient $\gamma(t)$ that can vary with time (the vanishing case $\gamma(t) \rightarrow 0$ as $t \rightarrow +\infty$ is particularly interesting because of its link with the Nesterov acceleration method). An abundant litterature has been devoted to the link between damped

*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

inertial dynamics and optimization algorithms, see e.g. [11], [14], [24, 25], [37], [41], [42] for recent developments on the subject.

b) The set-valued term $\partial\phi(\dot{x}(t))$ in (1.1) models dry friction (also called Coulomb friction) in mechanics. The friction potential function ϕ is supposed to satisfy the Dry Friction property

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

As a model example, take $\phi(x) = r\|x\|$ with $r > 0$. The key assumption $0 \in \text{int}(\partial\phi(0))$ expresses that ϕ has a sharp minimum at the origin. This is specified in the following lemma (whose proof is straightforward and can be found e.g. in [1, Lemma 4.1 page 83]).

Lemma 1.1 *Let $\phi : \mathcal{H} \rightarrow \mathbb{R}$ 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) $0 \in \text{int}(\partial\phi(0))$;
- (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, $\|f\| \leq r \implies \partial\phi^*(f) \ni 0$.*

To enlight the crucial role played by the parameter r in our analysis, we introduce the following definition.

Definition 1.1 *We say that the friction potential function ϕ satisfies the property $(DF)_r$ 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. Indeed, this property gives the soft thresholding property satisfied by the proximal operator associated with a function ϕ having a sharp minimum at the origin.

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

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

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

Proof. We have $\lambda\partial\phi(0) \supset B(0, \lambda r)$. Since $\|x\| \leq \lambda r$, we deduce that

$$0 + \lambda\partial\phi(0) \ni x.$$

Indeed, the unique solution z of the inclusion

$$z + \lambda\partial\phi(z) \ni x,$$

is equal to $z = \text{prox}_{\lambda\phi}(x)$, hence the equality $\text{prox}_{\lambda\phi}(x) = 0$. ■

In Adly-Attouch-Cabot [3] it has been shown that, generically with respect to the initial data, dry friction generates solution trajectories that converge in finite time to equilibria. Based on this property, we will analyze the convergence properties of the corresponding proximal-based algorithms obtained by various temporal discretizations of (HBDF). Our main results concern the Inertial Proximal Gradient algorithm with Dry Friction

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

where prox_ϕ denotes the proximal mapping associated with the convex function ϕ . Recall that, for any $x \in \mathcal{H}$, for any $\lambda > 0$

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

In (IPGDF), we assume that the function ϕ satisfies the property (DF). In accordance with the classical proximal-gradient method, the smooth function f enters the algorithm through its gradient, and the non-smooth function ϕ via its proximal mapping. In many practical situations, prox_ϕ can be easily computed in a closed form, which makes (IPGDF) a splitted algorithm suitable for optimization problems of large dimensions. As a remarkable property, for any sequence (x_k) generated by the algorithm (IPGDF) we will show that, if the viscous damping parameter γ is large enough (this is made precise later), then

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

This property expresses that the trajectory has finite length, and therefore $\lim_{k \rightarrow \infty} x_k := x_\infty$ exists for the strong topology of \mathcal{H} . Moreover, we will show that, under a general condition on the function ϕ , there is finite convergence (*i.e.* within a finite number of steps) of the iterates generated by (IPGDF). The mechanical interpretation of (IPGDF) gives a natural explanation to this property. According to Lemma 1.1 (*iv*), when the resulting force acting on the system is below a prescribed threshold, then, because of the dry friction, the speed is set to zero, and the system stops. Dry friction acts as a closed-loop stopping rule.

We emphasize that the convergence properties for dry friction are valid for arbitrary f , whereas convergence of the iterates under the sole viscous friction requires a geometric assumption on f , such as convexity [5] or analyticity [32, 33]. However, these very good properties have a counterpart: the limit is just an “approximate” critical point of f . It satisfies indeed

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

It is like solving the optimization problem with the Ekeland variational principle, instead of the Fermat rule. The dynamic (HBDF) can be considered as an inertial dynamic approach of solutions of Ekeland variational principle. 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”.

The paper is organized as follows. In section 2, we state our main results, which concern the convergence properties of the inertial proximal-gradient algorithm with dry friction (IPGDF). In section 2.5 we examine a variant of the algorithm obtained by a different discretization of the viscous damping term, with quite similar results. In section 3 we examine the effect of the introduction of perturbations, errors in the algorithm (IPGDF). In section 4, we introduce dry friction effect in the Nesterov accelerated gradient method. In section 5, based on the variational properties of Moreau’s envelope, we extend these results to the case where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ 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 splitted form. In section 6 we extend our

analysis to the case of additive composite optimization problems, and obtain splitting methods with finite convergence. 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 and an appendix concerning the study of the continuous dynamic that supports our study.

2 Inertial proximal-based algorithms with dry friction

In this section, we assume that f is a \mathcal{C}^1 function whose gradient is L -Lipschitz continuous. We will consider a splitted algorithm with finite convergence property. We consider the following discretization of the differential inclusion (HBDF), where we have taken $\gamma(t)$ constant, namely $\gamma(t) \equiv \gamma$, $\gamma > 0$, as well as the time step $h > 0$

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \nabla f(x_k) \ni 0. \quad (2.1)$$

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, not the functions to minimize, hence significant differences! Solving (2.1) with respect to x_{k+1} gives the following algorithm where the function f to be minimized enters via its gradient and the potential friction function ϕ enters via its proximal mapping.

(IPGDF): Inertial Proximal Gradient algorithm with Dry Friction

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

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

2.1 Convergence: finite length property

Theorem 2.1 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ 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 $(\text{DF})_r$. Suppose that the parameters h, γ in the algorithm (IPGDF) satisfy the relation*

$$h \leq \frac{2\gamma}{L}.$$

Then, for any sequence (x_k) defined by the algorithm (IPGDF), 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,

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r} E_1 \quad \text{and} \quad \sum_{k=1}^{\infty} \|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq 2h^2 E_1,$$

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

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

(i) Let's first establish energy estimates. Without ambiguity, we write simply $\partial\phi$ to designate any element belonging to this set. Taking the dot product of (2.1) with $\frac{1}{h}(x_{k+1} - x_k)$, we obtain

$$\begin{aligned} & \left\langle \frac{1}{h}(x_{k+1} - x_k) - \frac{1}{h}(x_k - x_{k-1}), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 \\ & + h \left\langle \partial\phi \left(\frac{1}{h}(x_{k+1} - x_k) \right), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \langle \nabla f(x_k), x_{k+1} - x_k \rangle = 0. \end{aligned} \quad (2.2)$$

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

$$\langle X_{k+1} - X_k, X_{k+1} \rangle = \frac{1}{2} \|X_{k+1}\|^2 - \frac{1}{2} \|X_k\|^2 + \frac{1}{2} \|X_{k+1} - X_k\|^2. \quad (2.3)$$

Using the definition of $\partial\phi$, we get

$$\langle \partial\phi(X_{k+1}), X_{k+1} \rangle \geq \phi(X_{k+1}). \quad (2.4)$$

Taking into account (2.3) and (2.4), we deduce from (2.2) the following inequality

$$\frac{1}{2} \|X_{k+1}\|^2 - \frac{1}{2} \|X_k\|^2 + \frac{1}{2} \|X_{k+1} - X_k\|^2 + \gamma h \|X_{k+1}\|^2 + h\phi(X_{k+1}) + h \langle \nabla f(x_k), X_{k+1} \rangle \leq 0. \quad (2.5)$$

According to the assumption $(DF)_r$ on ϕ , for all $k \geq 1$

$$\phi(X_{k+1}) \geq r \|X_{k+1}\|. \quad (2.6)$$

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

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2. \quad (2.7)$$

Combining inequalities (2.6)-(2.7) with (2.5), we obtain, for all $k \geq 1$

$$\begin{aligned} & \frac{1}{2} \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 - \frac{1}{2} \left\| \frac{1}{h}(x_k - x_{k-1}) \right\|^2 + \frac{1}{2h^2} \|x_{k+1} - 2x_k + x_{k-1}\|^2 + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 \\ & + r \|x_{k+1} - x_k\| + (f(x_{k+1}) - f(x_k)) - \frac{L}{2} \|x_{k+1} - x_k\|^2 \leq 0. \end{aligned} \quad (2.8)$$

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

$$E_{k+1} - E_k + \left(\frac{\gamma}{h} - \frac{L}{2} \right) \|x_{k+1} - x_k\|^2 + \frac{1}{2h^2} \|x_{k+1} - 2x_k + x_{k-1}\|^2 + r \|x_{k+1} - x_k\| \leq 0. \quad (2.9)$$

According to the assumption $\frac{\gamma}{h} - \frac{L}{2} \geq 0$, we obtain

$$E_{k+1} - E_k + \frac{1}{2h^2} \|x_{k+1} - 2x_k + x_{k-1}\|^2 + r \|x_{k+1} - x_k\| \leq 0. \quad (2.10)$$

Hence, for all $k \geq 1$

$$\|x_{k+1} - x_k\| \leq \frac{1}{r} (E_k - E_{k+1}). \quad (2.11)$$

According to the nonnegativity of E_k , and $r > 0$, we deduce from (2.11) that

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r} E_1 < +\infty. \quad (2.12)$$

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 (2.10), we also get

$$\sum_{k=1}^{\infty} \|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq 2h^2 E_1 < +\infty. \quad (2.13)$$

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 immediately $\lim_k \|x_{k+1} - x_k\| = 0$. This in turn implies

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

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 (2.1), rewrite it as follows:

$$-\frac{1}{h^2} (x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h} (x_{k+1} - x_k) - \nabla f(x_k) \in \partial\phi \left(\frac{1}{h} (x_{k+1} - x_k) \right). \quad (2.14)$$

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

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

which gives item (ii). ■

2.2 Convergence rate: linear and finite convergence results

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

$$-\nabla f(x_\infty) \in \text{int}(\partial\phi(0)), \quad (2.15)$$

we then obtain linear convergence and finite convergence results. Note that condition (2.15) 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 $f : \mathcal{H} \rightarrow \mathbb{R}$ 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 $(\text{DF})_r$. Suppose that the parameters h, γ in the algorithm (IPGDF) satisfy the relation*

$$h \leq \frac{2\gamma}{L}.$$

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

(i) *Suppose that*

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

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

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

Therefore, 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

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

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

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

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

Proof.

(i) The assumption $-\nabla f(x_\infty) \in \text{int}(\partial\phi(0))$ 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:

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

So, for every $k \geq k_0$, we have

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

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

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

On the other hand, using inequality (2.5), we get

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

By combining the two inequalities above, we get, for every $k \geq k_0$

$$\frac{1}{2} \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 - \frac{1}{2} \left\| \frac{1}{h}(x_k - x_{k-1}) \right\|^2 + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 + \varepsilon \|x_{k+1} - x_k\| \leq 0. \quad (2.16)$$

a) Neglecting the nonnegative term $\varepsilon \|x_{k+1} - x_k\| \geq 0$ in (2.16), we obtain

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

Equivalently

$$(1 + 2h\gamma) \|(x_{k+1} - x_k)\|^2 \leq \|x_k - x_{k-1}\|^2,$$

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

$$\|x_{k+1} - x_k\| \leq \left(\frac{1}{\sqrt{1 + 2h\gamma}} \right)^k \|x_1 - x_0\|. \quad (2.17)$$

Set $q := \frac{1}{\sqrt{1 + 2h\gamma}}$ and $C := \|x_1 - x_0\|$. For, $p \geq 0$ we have

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

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

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

Hence, 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.

b) Neglecting the nonnegative term $\frac{\gamma}{h} \|x_{k+1} - x_k\|^2 \geq 0$ in (2.16), and setting $\alpha_k := \|x_k - x_{k-1}\|^2$, we obtain the discrete differential inequality:

$$\alpha_{k+1} - \alpha_k + 2\varepsilon h^2 \sqrt{\alpha_{k+1}} \leq 0 \quad \text{for all } k \geq k_0. \quad (2.18)$$

In the continuous case, the corresponding differential inequality

$$\dot{\alpha}(t) + 2\varepsilon \sqrt{\alpha(t)} \leq 0, \quad t \in [0, +\infty[\quad (2.19)$$

permits to conclude to the finite time convergence of the trajectories in general Hilbert spaces. But contrary to the continuous case, (2.18) does not make it possible to conclude to the finite convergence to zero of the velocities.

(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 (IPGDF) as follows:

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

Equivalently,

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

Set $\xi_k := \frac{1}{1 + \gamma h} \left(\frac{1}{h}(x_k - x_{k-1}) - h\nabla f(x_k) \right)$ and $\lambda := \frac{h}{1 + \gamma h}$. So, we have

$$\frac{1}{h}(x_{k+1} - x_k) = \text{prox}_{\lambda\phi}(\xi_k). \quad (2.20)$$

To show the finite convergence property, we need to show that $x_{k+1} - x_k = 0$ for k large enough. According to (2.20) and Lemma 1.2, it suffices to prove that

$$\frac{1}{\lambda} \|\xi_k\| \leq r. \quad (2.21)$$

By the triangle inequality and the L -Lipschitz continuity of ∇f we have

$$\frac{1}{\lambda} \|\xi_k\| = \left\| \frac{1}{h^2} (x_k - x_{k-1}) - \nabla f(x_k) \right\| \leq \frac{1}{h^2} \|x_k - x_{k-1}\| + \|\nabla f(x_\infty)\| + L \|x_k - x_\infty\|. \quad (2.22)$$

When $k \rightarrow +\infty$, the whole right-hand side of the inequality (2.22) tends to $\|\nabla f(x_\infty)\|$. So, condition (2.21) will be satisfied for k large enough if $\|\nabla f(x_\infty)\| < r$. Let us suppose this condition satisfied and further analyze (2.22). We will have $x_{k+1} - x_k = 0$ as soon as

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

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

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

This gives

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

which completes the proof. ■

Remark 2.1 Let's present another proof of the finite convergence property. On the one hand, it only assumes that $-\nabla f(x_\infty) \in \text{int}(\partial\phi(0))$, but it is valid only when \mathcal{H} is a finite dimensional space. It is similar to the argument developed by Baji-Cabot in [22]. Argue by contradiction and suppose that there is an infinite number of indices k such that $\|x_{k+1} - x_k\| \neq 0$. Set $\mathcal{N} := \{k \in \mathbb{N} : x_{k+1} \neq x_k\}$, and consider the sequence $(\omega_k)_k$ defined by

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

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). By the monotonicity property of $\partial\phi$ and the definition (2.1) of the algorithm (IPGDF) we have, for $k \in \mathcal{N}$

$$\left\langle -\frac{1}{h^2} (x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h} (x_{k+1} - x_k) - \nabla f(x_k) - \partial\phi(0), \frac{x_{k+1} - x_k}{\|x_{k+1} - x_k\|} \right\rangle \geq 0. \quad (2.23)$$

According to convergence properties shown above, by passing to the limit in (2.23), we obtain

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

Since $-\nabla f(x_\infty) \in \text{int}(\partial\phi(0))$, 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 contradiction with ω belonging to the unit sphere.

Remark 2.2 The case $\phi = 0$ gives the heavy ball with friction method initiated by Polyak [38], [39]. This case is excluded from our analysis because of the dry friction assumption (DF) on ϕ . We can compare our method advantageously with the method of the heavy ball, for which the results of convergence require restrictive assumptions on the parameters and the function f , see [31] for a recent account on the heavy ball method.

Remark 2.3 Let's comment the relationship between the geometric convergence property obtained in Theorem 2.2

$$\|x_k - x_\infty\| \leq Cq^k$$

and the classical linear convergence property, which is given by

$$\|x_{k+1} - x_\infty\| \leq q\|x_k - x_\infty\|,$$

where the parameter q belongs to $]0, 1[$. The above inequalities are supposed to be satisfied for k large enough. Of course, the linear convergence implies the geometric convergence. But the converse statement is not true in general. Take for example the sequence (a_k) in \mathbb{R} which is defined by

$$\begin{cases} a_{2k} = \frac{1}{3^{2k}}; \\ a_{2k+1} = \frac{1}{2^{2k+1}}. \end{cases} \quad (2.24)$$

Clearly, the sequence (a_k) converges geometrically to zero (with $q = \frac{1}{2}$). But it does not converge linearly, since there exists no $0 < q < 1$ such that

$$\frac{1}{2^{2k+1}} \leq q \frac{1}{3^{2k}}.$$

Indeed, the relation above would imply $q \geq \frac{1}{2} \left(\frac{3}{2}\right)^{2k}$, an obvious contradiction with the requirement $0 < q < 1$. A natural question would be to know if in Theorem 2.2 one could obtain the linear convergence instead of the geometric convergence.

Note also that the estimation of the number of steps to be stopped (finite convergence) depends on the geometric convergence result previously demonstrated. Without this estimate, the finite convergence property alone is less interesting from a numerical point of view.

2.3 Soft thresholding on the velocities

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

$$\partial\phi(x) = \begin{cases} r \frac{x}{\|x\|} & \text{if } x \neq 0; \\ B(0, r) & \text{if } x = 0. \end{cases} \quad (2.25)$$

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

$$\text{prox}_{\lambda\phi}(x) = \left(1 - \frac{\lambda r}{\max\{\lambda r, \|x\|\}}\right)x = \begin{cases} 0 & \text{if } \|x\| \leq \lambda r; \\ (\|x\| - \lambda r) \frac{x}{\|x\|} & \text{if } \|x\| \geq \lambda r. \end{cases} \quad (2.26)$$

This thresholding property is in accordance with Lemma 1.2.

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

$$T_{\lambda r}(x) = \text{sign}(x)(|x| - \lambda r)_+ = \begin{cases} x - \lambda r & \text{if } x \geq \lambda r; \\ 0 & \text{if } -\lambda r \leq x \leq \lambda r; \\ x + \lambda r & \text{if } x \leq -\lambda r. \end{cases} \quad (2.27)$$

b) In the multidimensional case $\mathcal{H} = \mathbb{R}^n$, take $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ 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: $\text{prox}_{\lambda\phi}(x)$ is the solution of the minimization problem

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

which can be decomposed with respect to each component. Hence

$$\left(\text{prox}_{\lambda r \|\cdot\|_1}(x) \right)_i = T_{\lambda r}(x_i) = \text{sign}(x_i)(|x_i| - \lambda r)_+ \quad \text{for } i = 1, 2, \dots, n. \quad (2.28)$$

When f is differentiable, the algorithm (IPGDF) reads componentwise as follows: setting $x_k = (x_{k,i})_{i=1,2,\dots,n}$, we have for $i = 1, 2, \dots, n$

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

The operator $T_{\frac{hr}{1+h\gamma}}$ acts as a thresholding operator on the velocities. By Theorem 2.1, under the condition $hL \leq 2\gamma$, where L is the Lipschitz constant of ∇f , the sequence (x_k) has a finite length and converges to x_∞ which verifies

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

Clearly, taking r small is the interesting situation for optimization. Theorem 2.2 tells us that if

$$\|\nabla f(x_\infty)\| < r$$

then, there is geometric convergence, and the sequence (x_k) is finitely convergent.

2.4 An example

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

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

Equivalently,

$$(x_{k+1} - x_k) + \frac{1}{1+\gamma} \partial\phi(x_{k+1} - x_k) \ni -\frac{1}{1+\gamma} x_{k-1}, \quad (2.30)$$

which gives

$$x_{k+1} - x_k = T_{\frac{r}{1+\gamma}} \left(-\frac{1}{1+\gamma} x_{k-1} \right). \quad (2.31)$$

According to (2.27) we get

$$x_{k+1} - x_k = \begin{cases} -\frac{1}{1+\gamma}(x_{k-1} + r) & \text{if } x_{k-1} \leq -r; \\ 0 & \text{if } |x_{k-1}| \leq r; \\ -\frac{1}{1+\gamma}(x_{k-1} - r) & \text{if } x_{k-1} \geq r; \end{cases} \quad (2.32)$$

Take as particular values: $r = 1$, $\gamma = 3$, and $x_0 = x_1 = 2$. Then as long as $x_{k-1} \geq 1$ we have

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

Therefore, the sequence $X_k := x_k - 1$ satisfies the linear recurrence relation of order two

$$4X_{k+1} - 4X_k + X_{k-1} = 0$$

with $X_0 = X_1 = 1$. This is true as long as $X_k \geq 0$. An elementary calculation gives $X_k = \frac{1}{2^k} + \frac{k}{2^k}$. Since this quantity remains positive for all k , we have $x_k = 1 + \frac{1}{2^k} + \frac{k}{2^k}$. In this particular situation we have linear convergence but not finite convergence. This is in accordance with $x_\infty = 1$ and $\nabla f(x_\infty) = 1$, which is not in the interior of the convex set $\partial\phi(0) = [-1, +1]$.

2.5 A proximal-gradient variant

Consider the following discretization of (HBDF)

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_k - x_{k-1}) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \nabla f(x_k) \ni 0, \quad (2.33)$$

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

<p>(IPGDF-variant):</p> <hr/> <p><i>Initialize</i> : $x_0 \in \mathcal{H}, x_1 \in \mathcal{H}$</p> <p>$x_{k+1} = x_k + h \operatorname{prox}_{h\phi}\left(\left(\frac{1-h\gamma}{h}\right)(x_k - x_{k-1}) - h\nabla f(x_k)\right).$</p>
--

We obtain results similar to those of Theorem 2.1 and Theorem 2.2. The proofs, which are quite similar to those of these theorems, are given in the appendix.

Theorem 2.3 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ 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 $(\text{DF})_r$. Suppose that the parameters h, γ in the algorithm (IPGDF – variant) satisfy the relation*

$$h < \inf \left\{ \frac{2\gamma}{L}; \frac{1}{\gamma} \right\}.$$

Then, for any sequence (x_k) defined by the algorithm (IPGDF – variant), 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,

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

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

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

Theorem 2.4 (geometric and finite convergence) Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a C^1 function whose gradient is L -Lipschitz continuous, and such that $\inf f > -\infty$. Assume that the potential friction function ϕ satisfies $(DF)_r$. Suppose that the parameters h, γ in the algorithm (IPGDF – variant) satisfy the relation

$$h < \inf \left\{ \frac{2\gamma}{L}; \frac{1}{\gamma} \right\}.$$

Let (x_k) be a sequence generated by (IPGDF – variant), and let x_∞ be its limit (given by Theorem 2.3).

(i) Suppose that

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

Then, there is geometric convergence of the velocities to zero. There exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$

$$\|x_k - x_\infty\| \leq \frac{1}{h\gamma} (1 - h\gamma)^k \|x_1 - x_0\|.$$

(ii) 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.

3 Errors, perturbations

Let's examine the effect of the introduction of perturbations, errors in the algorithm (IPGDF). According to the dynamic approach, let's start from the perturbed version of (HBDF)

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

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

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \nabla f(x_k) \ni e_k. \quad (3.2)$$

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

(IPGDF-pert)
<p><i>Initialize</i> : $x_0 \in \mathcal{H}, x_1 \in \mathcal{H}$</p> $x_{k+1} = x_k + h \text{prox}_{\frac{h}{1+h\gamma}\phi} \left(\frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{h}{1+h\gamma}\nabla f(x_k) + \frac{h}{1+h\gamma}e_k \right).$

We have the following convergence results that complement 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:*

$$\lim_k e_k = 0 \quad \text{as } k \rightarrow +\infty.$$

Then, for any sequence (x_k) defined by the algorithm (IPGDF – 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} .

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

(iii) Suppose that $-\nabla f(x_\infty) \in \text{int}(\partial\phi(0))$. 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 proof is similar to that of Theorem 2.1, and uses the sequence (E_k) of energy functions

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

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

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

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

$$\frac{1}{2} \|X_{k+1}\|^2 - \frac{1}{2} \|X_k\|^2 + \gamma h \|X_{k+1}\|^2 + h\phi(X_{k+1}) + h\langle \nabla f(x_k), X_{k+1} \rangle \leq \langle e_k, x_{k+1} - x_k \rangle. \quad (3.4)$$

According to the assumption $(DF)_r$ on ϕ , for all $k \geq 1$

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

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

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

Combining the two above inequalities with (3.4), and using Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \frac{1}{2} \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 - \frac{1}{2} \left\| \frac{1}{h}(x_k - x_{k-1}) \right\|^2 + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 \\ & + r \|x_{k+1} - x_k\| + (f(x_{k+1}) - f(x_k)) - \frac{L}{2} \|x_{k+1} - x_k\|^2 \leq \|e_k\| \|x_{k+1} - x_k\|. \end{aligned} \quad (3.5)$$

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

$$E_{k+1} - E_k + \left(\frac{\gamma}{h} - \frac{L}{2} \right) \|x_{k+1} - x_k\|^2 + (r - \|e_k\|) \|x_{k+1} - x_k\| \leq 0. \quad (3.6)$$

According to the assumption $\frac{\gamma}{h} - \frac{L}{2} \geq 0$, and $e_k \rightarrow 0$, we obtain that for k sufficiently large

$$E_{k+1} - E_k + \frac{r}{2} \|x_{k+1} - x_k\| \leq 0. \quad (3.7)$$

From this we deduce that

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| < +\infty. \quad (3.8)$$

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

$$-\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h}(x_{k+1} - x_k) - \nabla f(x_k) + e_k \in \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right). \quad (3.9)$$

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

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

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

$$\frac{1}{2} \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 - \frac{1}{2} \left\| \frac{1}{h}(x_k - x_{k-1}) \right\|^2 + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 + \varepsilon \|x_{k+1} - x_k\| \leq \|e_k\| \|x_{k+1} - x_k\|. \quad (3.10)$$

Since $e_k \rightarrow 0$, we obtain that, for k sufficiently large, $\|e_k\| \leq \frac{\varepsilon}{2}$. Therefore,

$$\frac{1}{2} \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 - \frac{1}{2} \left\| \frac{1}{h}(x_k - x_{k-1}) \right\|^2 + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 + \frac{\varepsilon}{2} \|x_{k+1} - x_k\| \leq 0 \quad (3.11)$$

from which we easily deduce the geometric convergence of the sequence (x_k) .

To prove the finite convergence property, we return to (2.20)

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

where now ξ_k is taken equal to $\xi_k := \frac{1}{1+\gamma h} \left(\frac{1}{h}(x_k - x_{k-1}) - h\nabla f(x_k) + he_k \right)$. According to Lemma 1.2, to show the finite convergence property, it suffices to prove that

$$\frac{1}{\lambda} \|\xi_k\| \leq r. \quad (3.12)$$

By the triangle inequality and the L -Lipschitz continuity of ∇f we have

$$\begin{aligned} \frac{1}{\lambda} \|\xi_k\| &= \left\| \frac{1}{h^2}(x_k - x_{k-1}) - \nabla f(x_k) + e_k \right\| \\ &\leq \frac{1}{h^2} \|x_k - x_{k-1}\| + \|\nabla f(x_{\infty})\| + L \|x_k - x_{\infty}\| + \|e_k\|. \end{aligned} \quad (3.13)$$

When $k \rightarrow +\infty$, the whole right-hand side of the inequality (3.13) tends to $\|\nabla f(x_{\infty})\|$. So, condition (3.12) will be satisfied for k large enough if $\|\nabla f(x_{\infty})\| < r$. ■

4 Combining Nesterov acceleration method with dry friction

We consider the algorithms obtained by the temporal discretization of (HBDF), and which correspond to the structure of the Nesterov accelerated gradient method. When discretizing (HBDF), 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 (IPGDF) studied in the previous section. Taking $y_k = x_{k+1}$, we obtain a proximal algorithm that will be studied in the next section. The Inertial Gradient method of Nesterov reads as follows:

$$(\text{IGN}) \begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} &= y_k - s\nabla f(y_k). \end{cases}$$

In this algorithm, y_k is an extrapolated point with $0 < \alpha_k < 1$ wisely chosen. We will come back to the choice of α_k later, which is a crucial point. One step of (IGN) is illustrated in Figure 1.

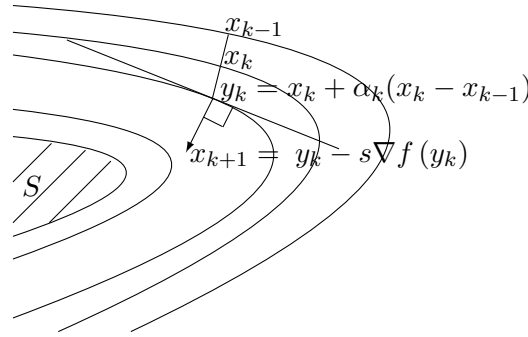


Figure 1: Accelerated gradient method of Nesterov.

We will successively examine the algorithms of this type which are obtained by discretization of (HBDF), first with a constant viscous damping coefficient $\gamma(t)$, then with a vanishing damping coefficient $\gamma(t) = \frac{\alpha}{t}$.

4.1 Fixed damping coefficient $\gamma(t) \equiv \gamma > 0$

Then (HBDF) is written as

$$\ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0. \quad (4.1)$$

Let's consider the following temporal discretization of (4.1)

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \nabla f(y_k) \ni 0, \quad (4.2)$$

where y_k will be chosen according to the Nesterov accelerated gradient method. To solve (4.2) with respect to $\frac{1}{h}(x_{k+1} - x_k)$, let's write it equivalently as

$$\frac{1}{h}(x_{k+1} - x_k) - \frac{1}{h}(x_k - x_{k-1}) + h\gamma\frac{1}{h}(x_{k+1} - x_k) + h\partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + h\nabla f(y_k) \ni 0. \quad (4.3)$$

Equivalently

$$(1 + h\gamma)\frac{1}{h}(x_{k+1} - x_k) + h\partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) \ni \frac{1}{h}(x_k - x_{k-1}) - h\nabla f(y_k), \quad (4.4)$$

which gives

$$\frac{1}{h}(x_{k+1} - x_k) + \frac{h}{1 + h\gamma}\partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) \ni \frac{1}{h(1 + h\gamma)}(x_k - x_{k-1}) - \frac{h}{1 + h\gamma}\nabla f(y_k). \quad (4.5)$$

Therefore

$$\frac{1}{h}(x_{k+1} - x_k) = \text{prox}_{\frac{h}{1+h\gamma}\phi} \left(\frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{h}{1+h\gamma}\nabla f(y_k) \right). \quad (4.6)$$

When $\phi = 0$, the proximal operator is the identity, and (4.6) is reduced to

$$x_{k+1} = x_k + \frac{1}{1+h\gamma}(x_k - x_{k-1}) - \frac{h^2}{1+h\gamma}\nabla f(y_k). \quad (4.7)$$

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:

(IPGDF-NF):

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

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

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

Theorem 4.1 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a 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 $(DF)_r$. Suppose that the parameters h, γ in the algorithm (IPGDF – NF) satisfy the relation*

$$h < \frac{2\gamma}{3L}.$$

Then, for any sequence (x_k) defined by the algorithm (IPGDF – NF), we have:

(i) $\sum_k \|x_{k+1} - x_k\| < +\infty$, and therefore $\lim_k x_k := x_\infty$ 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 (4.3) with $\frac{1}{h}(x_{k+1} - x_k)$, we obtain

$$\begin{aligned} & \left\langle \frac{1}{h}(x_{k+1} - x_k) - \frac{1}{h}(x_k - x_{k-1}), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + h\gamma \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 \\ & + h \left\langle \partial\phi \left(\frac{1}{h}(x_{k+1} - x_k) \right), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \langle \nabla f(y_k), x_{k+1} - x_k \rangle = 0. \end{aligned} \quad (4.8)$$

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

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

So, according to (4.8), we have

$$\langle X_{k+1} - X_k, X_{k+1} \rangle + \gamma h \|X_{k+1}\|^2 + hr \|X_{k+1}\| + \langle \nabla f(y_k), x_{k+1} - x_k \rangle \leq 0. \quad (4.9)$$

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

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) + \langle \nabla f(y_k), x_{k+1} - x_k \rangle + L \|y_k - x_k\| \|x_{k+1} - x_k\| + \frac{L}{2} \|x_{k+1} - x_k\|^2 \\ &\leq f(x_k) + \langle \nabla f(y_k), x_{k+1} - x_k \rangle + \frac{L}{1+h\gamma} \|x_k - x_{k-1}\| \|x_{k+1} - x_k\| + \frac{L}{2} \|x_{k+1} - x_k\|^2. \end{aligned} \quad (4.10)$$

Combining the above inequality with (4.9), we obtain

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

Equivalently

$$(1 + h\gamma - \frac{Lh^2}{2}) \|X_{k+1}\|^2 - (1 + \frac{Lh^2}{1+h\gamma}) \|X_k\| \|X_{k+1}\| + hr \|X_{k+1}\| + f(x_{k+1}) - f(x_k) \leq 0. \quad (4.11)$$

We have $1 + h\gamma - \frac{Lh^2}{2} > 0$ for $0 < h < \frac{1}{L}(\gamma + \sqrt{\gamma^2 + 2L})$. This last inequality is satisfied when $0 < h < \frac{2\gamma}{L}$, which in turn is satisfied under our assumption $0 < h < \frac{2\gamma}{3L}$. From (4.11) we infer

$$\begin{aligned} &\frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2})(\|X_{k+1}\|^2 - \|X_k\|^2) + \frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2})\|X_{k+1}\|^2 + \frac{1}{2}(1 + h\gamma - \frac{Lh^2}{2})\|X_k\|^2 \\ &\quad - (1 + \frac{Lh^2}{1+h\gamma})\|X_k\| \|X_{k+1}\| + hr \|X_{k+1}\| + (f(x_{k+1}) - f(x_k)) \leq 0, \end{aligned} \quad (4.12)$$

where we have used Cauchy-Schwarz inequality. Elementary algebra (sign of a polynomial of the second degree) gives that

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

under the condition

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

This is equivalent to

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

Since $\frac{1}{2} + \frac{1}{1+h\gamma} \leq \frac{3}{2}$, we end up with the condition

$$h \leq \frac{2\gamma}{3L},$$

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

$$E_{k+1} - E_k + r \|x_{k+1} - x_k\| \leq 0. \quad (4.13)$$

According to the nonnegativity of E_k , and $r > 0$, we deduce from (4.13) 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

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

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 $y_k - x_k = \frac{1}{1+h\gamma}(x_k - x_{k-1})$ in the definition of (IPGDF-NF), 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 on (4.2), rewrite it as follows:

$$-\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h}(x_{k+1} - x_k) - \nabla f(y_k) \in \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right). \quad (4.14)$$

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

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

which gives item (ii). ■

4.2 A variant

Let's return to (4.6), which is recalled below

$$\frac{1}{h}(x_{k+1} - x_k) = \text{prox}_{\frac{h}{1+h\gamma}\phi}\left(\frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{h}{1+h\gamma}\nabla f(y_k)\right), \quad (4.15)$$

and make a different choice of y_k . Taking $y_k - x_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1})$ gives the following algorithm

(IPGDF-NF-variant):
<hr/> Initialize : $x_0 \in \mathcal{H}, x_1 \in \mathcal{H}$
$y_k = x_k + \frac{1}{h(1+h\gamma)}(x_k - x_{k-1})$
$x_{k+1} = x_k + h \text{prox}_{\frac{h}{1+h\gamma}\phi}\left(y_k - x_k - \frac{h}{1+h\gamma}\nabla f(y_k)\right).$

When $\phi = 0$ 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 [35]. Note that (IPGDF-NF) and its variant (IPGDF-NF-variant) rely on the discretization of (HBDF)

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi \left(\frac{1}{h}(x_{k+1} - x_k) \right) + \nabla f(y_k) \ni 0, \quad (4.16)$$

where y_k is chosen differently. In both cases, there exists a certain positive constant C (independent of k) 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 (IPGDF – NF – variant). A slight adaptation of the proof of Theorem 4.1 gives the following convergence result.

Theorem 4.2 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ 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 $(DF)_r$. Suppose that the parameters h, γ in the algorithm (IPGDF – variant) satisfy the relation*

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

Then, for any sequence (x_k) defined by the algorithm (IPGDF – NF), we have:

(i) $\sum_k \|x_{k+1} - x_k\| < +\infty$, and therefore $\lim_k x_k := x_\infty$ 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. The proof follows the line of Theorem 4.1 until (4.10) where the choice of y_k appears explicitly. We now have

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

which gives

$$(1+h\gamma - \frac{Lh^2}{2})\|X_{k+1}\|^2 - (1 + \frac{Lh}{1+h\gamma})\|X_k\|\|X_{k+1}\| + hr\|X_{k+1}\| + f(x_{k+1}) - f(x_k) \leq 0. \quad (4.17)$$

We have $1+h\gamma - \frac{Lh^2}{2} > 0$ for $0 < h < \frac{1}{L}(\gamma + \sqrt{\gamma^2 + 2L})$. This last inequality is satisfied when $0 < h < \frac{2\gamma}{L}$. From (4.17) we infer

$$\begin{aligned} & \frac{1}{2}(1+h\gamma - \frac{Lh^2}{2})(\|X_{k+1}\|^2 - \|X_k\|^2) + \frac{1}{2}(1+h\gamma - \frac{Lh^2}{2})\|X_{k+1}\|^2 + \frac{1}{2}(1+h\gamma - \frac{Lh^2}{2})\|X_k\|^2 \\ & - (1 + \frac{Lh}{1+h\gamma})\|X_k\|\|X_{k+1}\| + hr\|X_{k+1}\| + (f(x_{k+1}) - f(x_k)) \leq 0. \end{aligned} \quad (4.18)$$

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

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

under the condition

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

This is equivalent to

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

Clearly, this is a stronger requirement than the previous one $h \leq \frac{2\gamma}{L}$. To summarize the results, in terms of

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

we have obtained

$$E_{k+1} - E_k + r\|x_{k+1} - x_k\| \leq 0. \quad (4.19)$$

The end of the proof is similar to that of Theorem 4.1. ■

4.3 Vanishing damping parameter

Let's start from the inertial dynamic with the damping coefficient $\gamma(t) = \frac{\alpha}{t}$

$$(\text{AVD} - \text{DF}) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0, \quad (4.20)$$

where α is a positive parameter. When $\phi = 0$, we obtain the inertial system

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

where, as a specific property, the damping coefficient $\gamma(t) = \frac{\alpha}{t}$ tends to zero as $t \rightarrow +\infty$, hence the terminology (AVD) for Asymptotic Vanishing Damping. The system $(\text{AVD})_{\alpha}$ was introduced in the context of convex optimization by Su-Boyd-Candès [42] in 2014. For a general convex differentiable function f , it provides a continuous version of the accelerated gradient method of Nesterov. For $\alpha \geq 3$, each trajectory $x(\cdot)$ of $(\text{AVD})_{\alpha}$ satisfies the asymptotic rate of convergence of the values $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(1/t^2)$. The convergence properties of the dynamic $(\text{AVD})_{\alpha}$ have been the subject of many recent studies, see [8], [10, 11], [12], [14], [18], [21], [36], [42]. 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 [15]). As a remarkable property, for $\alpha > 3$, it has been shown by Attouch-Chbani-Peypouquet-Redont [14] and May [36] that each trajectory converges weakly to a minimizer of f . The corresponding algorithmic result has been obtained by Chambolle-Dossal [27]. For $\alpha > 3$, it is shown in [18] and [36] that the asymptotic convergence rate of the values is $o(1/t^2)$. Let us consider the following discretization of (AVD-DF)

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \nabla f(y_k) \ni 0, \quad (4.21)$$

where y_k is chosen according to the Nesterov accelerated gradient method. A similar computation as above (just replace γ by $\frac{\alpha}{kh}$) leads to consider the following algorithm

(IPGDF-NV):

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

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

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

and its variant

(IPGDF-NV-variant):

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

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

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

The mathematical analysis of these algorithms is a research topic for future work, which is beyond the scope of this paper.

5 A splitted proximal algorithm with dry friction for nonsmooth functions

We assume that $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous and proper function such that $\inf f > -\infty$. The previous section deals with a differentiable function f , without convexity assumption on f . Now, when considering nonsmooth functions, we assume the convexity of f . To reduce to the previous situation, where $f : \mathcal{H} \rightarrow \mathbb{R}$ is a \mathcal{C}^1 function whose gradient is Lipschitz continuous, the idea is to replace f by its Moreau's envelope. Recall some classical facts. For any $\lambda > 0$, the Moreau's envelope of f of index λ is the function $f_\lambda : \mathcal{H} \rightarrow \mathbb{R}$ 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 [9, section 17.2.1], [23], [26] 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. Since $\nabla f_\lambda(x) = \frac{1}{\lambda} (x - \operatorname{prox}_{\lambda f}(x))$, the algorithm (IPGDF) becomes

(IPP-DF): Inertial Proximal-Proximal algorithm with Dry Friction

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

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

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

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

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

$$h \leq 2\gamma\lambda.$$

Then, for any sequence (x_k) defined by the algorithm (IPP – DF), we have:

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

Moreover,

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

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

$$(ii) \quad \text{The limit } x_\infty \text{ of the sequence } (x_k) \text{ satisfies } 0 \in \partial\phi(0) + \nabla f_\lambda(x_\infty).$$

Suppose moreover that $-\nabla f_\lambda(x_\infty) \in \text{int}(\partial\phi(0))$. Then

(iii) There is linear convergence to zero of the velocities: there exists $k_0 \geq 0$ such that for all $k \geq k_0$

$$\|x_{k+1} - x_k\| \leq \left(\frac{1}{\sqrt{1+2h\gamma}} \right)^k \|x_1 - x_0\|.$$

Hence there is linear convergence of the sequence (x_k) : setting $q = \frac{1}{\sqrt{1+2h\gamma}}$

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

(iv) Suppose that

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

Then (x_k) is finitely convergent. It stops at x_k when $q^{k-1} \leq \frac{r - \|\nabla f_\lambda(x_\infty)\|}{\left(\frac{1}{h^2} + \frac{h}{\lambda} \frac{q}{1-q}\right) \|x_1 - x_0\|}$.

Proof. The proof is immediate: just 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 $hL \leq 2\gamma$ becomes $h \leq 2\gamma\lambda$. ■

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 $h \leq 2\gamma\lambda$). Thus our approach differs from the classical approximation method which consists approaching f by f_λ as λ goes to zero. In [19] a similar device has been used concerning algorithms with Hessian-driven damping.

Remark 5.2 When using Moreau envelopes, besides the sequence (x_k) , another sequence occurs naturally, namely (p_k) with $p_k = \text{prox}_{\lambda f} x_k$. Since $\text{prox}_{\lambda f}$ 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 $p_\infty = \text{prox}_{\lambda f} x_\infty$. 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 Inertial algorithms with dry friction for composite 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 deal with such situations. They are naturally linked to inertial dynamics with viscous friction. We will analyze the effect of the introduction of dry friction in these algorithms.

6.1 Lasso problem

Take $\mathcal{H} = \mathbb{R}^n$ equipped with the usual Euclidean structure. Suppose that the function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ to be minimized has the additive structure

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

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 $n = N^2$). To deal with such situations, taking g nonsmooth plays a crucial role. It also makes f nonsmooth. A direct application of the algorithm (IPP-DF) 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 Lemarechal and Sagasatizábal in [34] to introduce efficient preconditioners into the proximal point algorithm for minimizing convex functions, for recent developments see [28, Section 4.6], [19]. 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 $x \in \mathbb{R}^n$

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

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

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

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

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

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

$$\nabla f_\lambda^M(x) = \frac{1}{\lambda} M \left(x - \text{prox}_{\lambda f}^M(x) \right). \quad (6.4)$$

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

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

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 $M = I_n - \lambda A^T A$. If $\lambda \in [0, \frac{1}{\|A\|_2^2}]$, then M is positive definite. In this case,

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

In fact, by formulation (6.3) of $\operatorname{prox}_{\lambda f}^M(x)$, we have

$$\begin{aligned} y = \operatorname{prox}_{\lambda f}^M(x) &\iff Mx \in My + \lambda \partial f(y) \\ &\iff x - \lambda A^T Ax + \lambda A^T b \in y + \lambda \partial g(y) \\ &\iff y = \operatorname{prox}_{\lambda g}\left(x - \lambda A^T(Ax - b)\right). \end{aligned}$$

Note that formula (6.6) for the composite optimization problem (6.1) was given in [28, Section 4.6 page 190]. Using (6.4) and (6.6), we get

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

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

(IPGDF) for the Lasso problem
<p style="margin: 0;"><i>Initialize</i> : $x_0 \in \mathbb{R}^n, x_1 \in \mathbb{R}^n, M = I_n - \lambda A^T A, 0 < \lambda \ A\ _2^2 < 1$</p> $y_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{h}{\lambda(1+h\gamma)} M \left(x_k - \operatorname{prox}_{\lambda g}\left(x_k - \lambda A^T(Ax_k - b)\right) \right)$ $x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma} \phi}(y_k)$

For the LASSO problem, $g(x) = \|x\|_1$, formula (2.28) can be used to compute $\operatorname{prox}_{\lambda g}$.

6.2 The general additive composite case

Consider the following general additive composite situation: the function to be minimized is written in the form $f + g$ where $f : \mathcal{H} \rightarrow \mathbb{R}$ is a C^1 function whose gradient is Lipschitz continuous, and $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function. The application of the algorithm (IPP-DF) would require that f be convex, and to compute (at least approximatively) the proximal mapping of $f + g$, which is not easy in general. Let's introduce another more practical algorithm for applications. Like the algorithm (IPP-DF), it relies on the continuous second-order differential inclusion (HBDF) associated with the nonsmooth nonconvex potential function $f + g$.

$$(\text{HBDF}) \quad \ddot{x}(t) + \gamma \dot{x}(t) + \partial \phi(\dot{x}(t)) + \nabla f(x(t)) + \partial g(x(t)) \ni 0. \quad (6.8)$$

Given $h > 0$ a fixed time step size, we consider the following time discretization of (HBDF),

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial \phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) + \nabla f(x_k) + \partial g(x_{k+1}) \ni 0. \quad (6.9)$$

Equation (6.9) is in accordance with the classical dynamic approach to the proximal-gradient methods: it is implicit with respect to the nonsmooth function g , and explicit with respect to the smooth function f .

Unlike the algorithm (IPP-DF), the two potentials f and g now play a non-symmetric role. Let's formulate (6.9) with the help of the new variable $X_k := \frac{1}{h}(x_{k+1} - x_k)$. We have

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

which gives

$$(1+\gamma h)X_k+h\partial\phi(X_k)+h\partial g(x_k+hX_k)\ni\frac{1}{h}(x_k-x_{k-1})-h\nabla f(x_k).$$

For each $k \in \mathbb{N}$, let's introduce the auxiliary convex function $g_k : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$g_k(x) := \frac{1}{h}g(x_k + hx). \quad (6.10)$$

We have $\partial g_k(x) = \partial g(x_k + hx)$, which gives $\partial g_k(X_k) = \partial g(x_k + hX_k)$. The algorithm becomes

$$(1+\gamma h)X_k+h(\partial\phi+\partial g_k)(X_k)\ni\frac{1}{h}(x_k-x_{k-1})-h\nabla f(x_k).$$

Since ϕ is continuous, we have $\partial\phi+\partial g_k=\partial(\phi+g_k)$. We obtain

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

Returning to the variable x_k , we finally obtain the following proximal-gradient algorithm:

(IPGDF-composite): Inertial Proximal-Gradient algorithm with Dry Friction-composite

Initialize : $x_0 \in \mathbb{R}^n, x_1 \in \mathbb{R}^n$;

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

A direct adaptation of Theorem 2.1 gives the following result:

Theorem 6.1 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a C^1 function whose gradient is Lipschitz continuous, and let $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function. Suppose $\inf_{\mathcal{H}}(f+g) > -\infty$. Assume that the potential friction function ϕ satisfies $(DF)_r$. Suppose that the parameters h, γ in the algorithm (IPGDF – composite) satisfy the relation*

$$h \leq \frac{2\gamma}{L}.$$

Then, for any sequence (x_k) generated by the algorithm (IPGDF – composite) 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} .
- (ii) The vector x_∞ satisfies $0 \in \partial\phi(0) + \nabla f(x_\infty) + \partial g(x_\infty)$.
- (iii) Suppose that \mathcal{H} is finite dimensional, and that $-(\nabla f(x_\infty) + \partial g(x_\infty)) \in \operatorname{int}(\partial\phi(0))$. Then, the sequence (x_k) is finitely convergent.

Proof. The proof follows the lines of Theorem 2.1. We will use an energetic argument based on the nonincreasing property of the sequence (E_k) of nonnegative global energy functions

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

(i) Taking the dot product of (6.9) with $\frac{1}{h}(x_{k+1} - x_k)$, we obtain

$$\begin{aligned} & \left\langle \frac{1}{h}(x_{k+1} - x_k) - \frac{1}{h}(x_k - x_{k-1}), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 \\ & + h \left\langle \partial\phi \left(\frac{1}{h}(x_{k+1} - x_k) \right), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \langle \partial g(x_{k+1}) + \nabla f(x_k), x_{k+1} - x_k \rangle = 0. \end{aligned} \quad (6.11)$$

Set $X_k := \frac{1}{h}(x_k - x_{k-1})$. By convexity of $\frac{1}{2} \|\cdot\|^2$

$$\langle X_{k+1} - X_k, X_{k+1} \rangle \geq \frac{1}{2} \|X_{k+1}\|^2 - \frac{1}{2} \|X_k\|^2. \quad (6.12)$$

According to the assumption $(DF)_r$ on ϕ , for all $k \geq 1$

$$\langle \partial\phi(X_{k+1}), X_{k+1} \rangle \geq \phi(X_{k+1}) \geq r \|X_{k+1}\|. \quad (6.13)$$

By convexity of g

$$\langle \partial g(x_{k+1}), x_{k+1} - x_k \rangle \geq g(x_{k+1}) - g(x_k). \quad (6.14)$$

Taking into account (6.12), (6.13), and (6.14), we deduce from (6.11) the following inequality

$$\frac{1}{2} \|X_{k+1}\|^2 - \frac{1}{2} \|X_k\|^2 + \gamma h \|X_{k+1}\|^2 + r \|x_{k+1} - x_k\| + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + g(x_{k+1}) - g(x_k) \leq 0. \quad (6.15)$$

Since ∇f is assumed to be L -Lipschitz continuous, the gradient descent lemma applied to f gives

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

Combining the above inequality with (6.15), we obtain, for all $k \geq 1$

$$\begin{aligned} & \frac{1}{2} \left\| \frac{1}{h}(x_{k+1} - x_k) \right\|^2 - \frac{1}{2} \left\| \frac{1}{h}(x_k - x_{k-1}) \right\|^2 + \frac{\gamma}{h} \|x_{k+1} - x_k\|^2 \\ & + r \|x_{k+1} - x_k\| + (f + g)(x_{k+1}) - (f + g)(x_k) - \frac{L}{2} \|x_{k+1} - x_k\|^2 \leq 0. \end{aligned} \quad (6.16)$$

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

$$E_{k+1} - E_k + \left(\frac{\gamma}{h} - \frac{L}{2} \right) \|x_{k+1} - x_k\|^2 + r \|x_{k+1} - x_k\| \leq 0. \quad (6.17)$$

According to the assumption $\frac{\gamma}{h} - \frac{L}{2} \geq 0$, we obtain

$$E_{k+1} - E_k + r \|x_{k+1} - x_k\| \leq 0, \quad (6.18)$$

from which we get

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| \leq \frac{1}{r} E_1 < +\infty. \quad (6.19)$$

Therefore, (x_k) has a finite length, which implies its strong convergence. Set $x_\infty := \lim x_k$.

(ii) To pass to the limit as $k \rightarrow +\infty$, write (6.9) as

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \xi_k + \nabla f(x_k) + \eta_k \ni 0, \quad (6.20)$$

with $\xi_k \in \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right)$ and $\eta_k \in \partial g(x_{k+1})$. According to (6.20), the sequence (η_k) remains bounded, because all the other terms are bounded (since ϕ is continuous, it is bounded on a neighborhood of the origin). Therefore, we can extract convergent subsequences for the weak topology (we keep the same notation for subsequences) such that

$$\xi_k \rightharpoonup \xi \text{ and } \eta_k \rightharpoonup \eta.$$

The graphs of $\partial\phi$ and ∂g are closed for the strong- \mathcal{H} -weak- \mathcal{H} topology of $\mathcal{H} \times \mathcal{H}$. According to the strong convergence of $\frac{1}{h}(x_{k+1} - x_k)$ to zero, and the strong convergence of x_{k+1} to x_∞ , we deduce that $\xi \in \partial\phi(0)$ and $\eta \in \partial g(x_\infty)$. Passing to the limit on (6.20), we get $\xi + \nabla f(x_\infty) + \eta \ni 0$, that is

$$\partial\phi(0) + \nabla f(x_\infty) + \partial g(x_\infty) \ni 0. \quad (6.21)$$

(iii) \mathcal{H} is now assumed to be a finite dimensional space. We follow a similar argument as in Remark 2.1. Argue by contradiction, and suppose that there is an infinite number of indices k such that $\|x_{k+1} - x_k\| \neq 0$. Set $\mathcal{N} := \{k \in \mathbb{N} : x_{k+1} \neq x_k\}$, and consider the sequence $(\omega_k)_k$ defined by

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

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. By the monotonicity property of $\partial\phi$ and the definition of the algorithm (IPG-DF-composite) we have, for $k \in \mathcal{N}$

$$\left\langle -\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h}(x_{k+1} - x_k) - \nabla f(x_k) - \partial g(x_{k+1}) - \partial\phi(0), \frac{x_{k+1} - x_k}{\|x_{k+1} - x_k\|} \right\rangle \geq 0,$$

with $\eta_k \in \partial g(x_{k+1})$. After extraction of another sequence we have $\eta_k \rightarrow \eta \in \partial g(x_\infty)$. According to the convergence properties shown above and passing to the limit, as $k \rightarrow +\infty$, in the above inequality, we obtain

$$\langle -\nabla f(x_\infty) - \eta - \partial\phi(0), \omega \rangle \geq 0.$$

Since $-\nabla f(x_\infty) - \eta \subset -\nabla f(x_\infty) - \partial g(x_\infty) \subset \text{int}(\partial\phi(0))$, there exists some $\rho > 0$ such that

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

Thus we would have $\langle \rho u, \omega \rangle \geq 0$ 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 contradiction with ω belonging to the unit sphere. ■

Remark 6.1 When $f = 0$ and $\gamma = 0$, that is in the case of the proximal algorithm without viscous friction, we are in the situation considered by Baji-Cabot in [23].

6.3 Application of (IPG-DF-composite) to the Lasso problem

Take $\mathcal{H} = \mathbb{R}^n$ equipped with the usual Euclidean structure. Suppose that the function $\theta : \mathcal{H} \rightarrow \mathbb{R}$ to minimize has the additive structure

$$\theta(x) = f(x) + \|x\|_1 \quad (6.22)$$

where $f : \mathcal{H} \rightarrow \mathbb{R}$ is a \mathcal{C}^1 function whose gradient is Lipschitz continuous. The Lasso problem considered in the previous section corresponds to $f(x) = \frac{1}{2}\|Ax - b\|_2^2$. We are in the framework of application of the algorithm (IPGDF-composite) by taking $g(x) = \|x\|_1$, which is a convex continuous function (but not differentiable). So, let us apply the algorithm (IPGDF-composite) with $\phi(x) = r\|x\|_1$. The crucial point is the computation of the proximal mapping of $\text{prox}_{\frac{h}{1+\gamma h}(\phi+g_k)}$, where $g_k(x) := \frac{1}{h}g(x_k + hx)$. Since the function $g = \|\cdot\|_1$ is homogenous of degree one, we have $g_k(x) := \|\frac{1}{h}x_k + x\|_1$. Setting $\lambda = \frac{h}{1+\gamma h}$, we have to compute $\text{prox}_{\lambda F}$ where

$$F(x) := r\|x\|_1 + \|\frac{1}{h}x_k + x\|_1.$$

By definition of the proximal mapping we have

$$\text{prox}_{\lambda F}(x) = \underset{(y_1, \dots, y_n) \in \mathbb{R}^n}{\text{argmin}} \sum_{i=1}^n \left(\frac{1}{2\lambda} |x_i - y_i|^2 + r|y_i| + \left| \frac{1}{h}x_{k,i} + y_i \right| \right).$$

Indeed, this is a completely decomposed problem. Setting $a = \frac{1}{h}x_k = (a_i)$, it can be reduce to the computation for each coordinate ($i = 1, 2, \dots, n$) of the solution of the one-dimensional problem

$$\min_{y \in \mathbb{R}} \left\{ \frac{1}{2\lambda} (x - y)^2 + r|y| + |y + a_i| \right\}$$

which can be solved in a closed form by elementary argument. For each $a \in \mathbb{R}$, set

$$T_a(x) := \underset{y \in \mathbb{R}}{\text{argmin}} \left\{ \frac{1}{2\lambda} |x - y|^2 + r|y| + |y + a| \right\}. \quad (6.23)$$

Observe that $T_a(x) = -T_{-a}(-x)$. So, we just need to consider the case $a \geq 0$. Let's write the optimality condition for the strongly convex minimization problem (6.23):

$$y + \lambda r \text{sgn}(y) + \lambda \text{sgn}(y + a) \ni x.$$

We can easily verify that the following formula gives the unique solution of the above equation.

$$T_a(x) = \begin{cases} x - \lambda(1 + r) & \text{if } x \geq \lambda(1 + r) \\ 0 & \text{if } \lambda(1 - r) \leq x \leq \lambda(1 + r) \\ x - \lambda(1 - r) & \text{if } \lambda(1 - r) - a \leq x \leq \lambda(1 - r) \\ -a & \text{if } -a - \lambda(1 + r) \leq x \leq -a + \lambda(1 - r) \\ x + \lambda(1 + r) & \text{if } x \leq -a - \lambda(1 + r) \end{cases}$$

This is a threshold operator with two critical values 0 and $-a$. From this, the algorithm (IPGDF-composite) for the Lasso problem can be easily implemented. Note that the method can be applied with f not necessarily convex. The following picture represents the graph of T_a in the case $a > 0$. Consequently, for each $i = 1, 2, \dots, n$, we have

$$(\text{prox}_{\lambda F}(x))_i = \begin{cases} T_{a_i}(x_i) & \text{if } a_i \geq 0 \\ -T_{-a_i}(-x_i) & \text{if } a_i \leq 0 \end{cases} \quad (6.24)$$

with a_i the i th component of the vector $a = \frac{1}{h}x_k \in \mathbb{R}^n$ and $\lambda = \frac{h}{1+\gamma h}$.

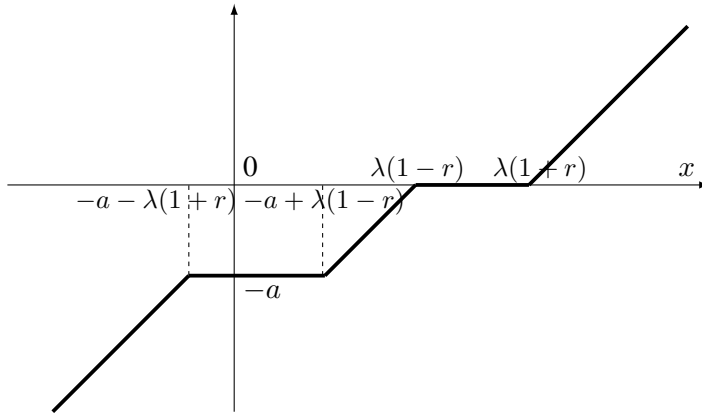


Figure 2: Threshold operator T_a with two levels.

7 Some numerical experiments

In this section, we perform numerical tests. Let's start by comparing the six algorithms IPGDF, IPGDF-Variant, IPGDF-NF, IPGDF-NF-Variant, IPGDF-NV and IPGDF-NV-Variant defined in sections 2 and 4. To realize this experiment, we chose the *performance profiles* developed by Dolan-Moré [30] 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 under comparison has a performance within the factor t of the best possible ratio.

To compare these algorithms, we choose the number of iterations and the cputime found by each solver as a performance measure. The function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is given by $x \mapsto \phi(x) = r\|x\|_2$ with $r = 0.1$, while the functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ are quadratic of the form $f(x) = \frac{1}{2}\langle Qx, x \rangle + \langle b, x \rangle$, with $Q \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ chosen randomly. The matrices Q in our set of tests come from the SuiteSparse Matrix Collection¹. We have chosen a set P of 70 different problems with matrices $Q \in \mathbb{R}^{n \times n}$ size ranging from $n = 24$ to $n = 44.609$. Let S be the set of the six solvers that will be compared and n_s the number of solvers. The *performance ratio* is defined by

$$r_{p,s} = \frac{t_{p,s}}{\min \{t_{p,s} : s \in S\}},$$

where $p \in P$, $s \in S$, and $t_{p,s}$ is the number of iterations corresponding to Figure 3. The performance of the solver $s \in S$ is defined by

$$\rho_s(t) = \frac{1}{n_p} \text{size} \{p \in P : r_{p,s} \leq t\},$$

where n_p is the number of problems, and t is a real factor. In this case, $\rho_s(t)$ is the probability, for solver $s \in S$, that the performance ratio $r_{p,s}$ is within a factor $t \in \mathbb{R}$ of the best possible ratio. For more details, we refer to [30].

By observing the performance profile, we can compare the performance of a solver compared to others by taking into consideration the aforementioned criterion. The value of $\rho_s(1)$ gives the probability that the solver s wins over the other five, while $\rho_s(t)$, for large value of t , measures the robustness of the solver s . The numerical experiments are carried out in an iMac with Mac OS 10.13 and a processor 4 GHz Intel Core i7 and 16Go memory. All the program codes are written and executed in Matlab R2018b. The stopping

¹<https://sparse.tamu.edu>

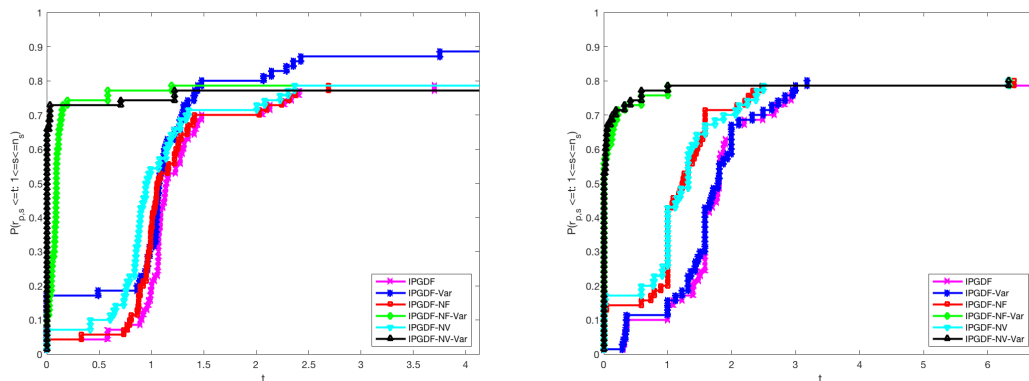


Figure 3: Performance profiles with $t_{p,s}$ the number of iterations (left) and cputime (right).

criterion is satisfied when either the number of iterations exceeds 10^5 or $\|\nabla f(x_k)\| \leq r$. We consider that an algorithm fails if the number of iterations exceeds 10^5 . Figure 3 represents the performance profiles of the six solvers corresponding to the final number of iterations of each. We observe first that all algorithms are robust and solve almost 80% of the problems. The algorithms IPGDF-NV-Variant and IPGDF-NF-Variant are efficient. In fact, in the interval $[0, \frac{1}{2}]$, both of them solved 75% of the problems, while IPGDF and IPGDF-NF do not reach 1%. We note that for $t \geq 3.5$, all solvers are robust. We conclude that, using the same initial points and under the same stopping criteria, IPGDF-NV-Variant is the winner closely followed by IPGDF-NF-Variant.

- In order to measure the effect of the introduction of the dry friction $\phi(x) = r\|x\|_2$ in IPGDF, we test this algorithm on a set of 70 problems with and without dry friction. The functions f in these tests are given by: $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, $A \in \mathbb{R}^{m \times n}$ (with $m \leq n$), $b \in \mathbb{R}^m$. We use the same initial points and the same stopping criterion i.e. either the number of iterations exceeds 10^5 or $\|\nabla f(x_k)\| \leq r$. The performance profiles, depicted in Figure 4 (left), shows clearly that IPGDF solves 80% of the problems in the interval $[0, 0.1]$ with less iterations, while the same algorithm without dry friction needs more times. Consistent with the theoretical part, we observe that the dry friction introduces some stability and robustness in the numerical algorithm.

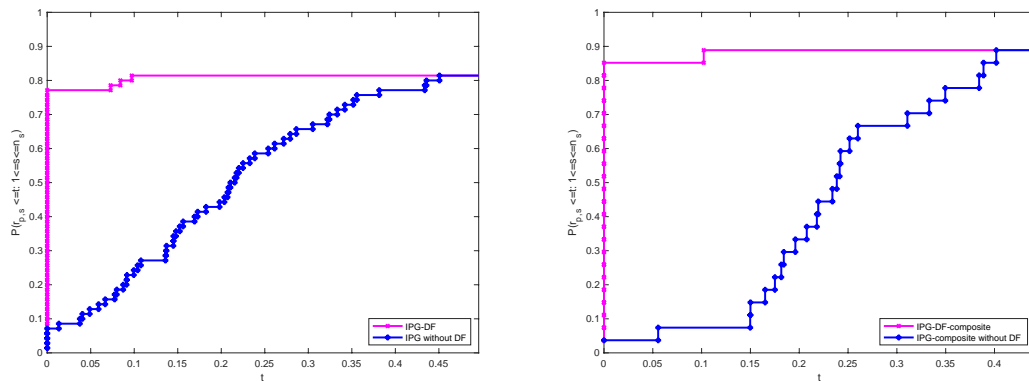


Figure 4: Comparison of IPG with and without dry friction (left) and IPG-composite with and without dry friction (right). Performance profiles with $t_{p,s}$ = the number of iterations.

• The LASSO method test: $\theta(x) = f(x) + g(x)$, with $f(x) = \frac{1}{2}\|Ax - b\|^2$, $A \in \mathbb{R}^{m \times n}$ (with $m \leq n$), $b \in \mathbb{R}^m$, and $g(x) = \alpha\|x\|_1$ with $\alpha > 0$. As before, we compare the algorithm IPGDF-composite with and without dry friction. In this case, we set $\phi(x) = r\|x\|_1$ and we use (6.24) to compute $\text{prox}_{\frac{h}{1+\gamma h}}(\phi + g_k)$, with $g_k(x) = \alpha\|x_k\|_1$. The performance profiles, depicted in Figure 4 (right), shows the efficiency of the introduction of the dry friction in the algorithm IPGDF-composite.

We observe that for the same initial points and the same stopping criterion, the algorithms with dry friction require less iterations than the same ones without dry friction. All the matrices $A \in \mathbb{R}^{m \times n}$ were taken from the SuiteSparse Matrix collection with size ranging from 24 to 44.609.

These numerical simulations have been introduced for illustrative purposes and it would be important to test these algorithms on concrete examples.

8 Conclusion, perspective

In this paper, we have studied a new class of inertial proximal-gradient algorithms with the specific property of dry friction damping. These algorithms are obtained by time discretization of a second-order differential inclusion (HBDF) involving both viscous and dry friction. It was shown in [3] that, under the assumption of dry friction, each trajectory of (HBDF) converges to an equilibrium in a finite time. We show in this paper that various discrete versions of this continuous dynamic follow the same pattern by converging geometrically and in a finite number of steps. Dry friction acts as a soft threshold on the velocity and forces the system to stop. From the optimization point of view, the limit point is not a critical point of the potential f to be minimized. It satisfies a stationary condition involving the gradient of the function f and the subdifferential of the dry friction ϕ at zero. If the dry friction function ϕ is chosen so that its subdifferential $\partial\phi(0)$ is relatively small, we obtain an approximate critical point of f . As a remarkable property, these results are valid without convexity assumption on the function f to be minimized, when f is differentiable. Extension to the nonsmooth case was obtained only when f is convex. It is an open question to consider the nonconvex and nonsmooth case. In addition, the algorithms have a very good property of stability, they tolerate errors just supposed to tend to zero. Several variants of the inertial proximal-based algorithms with dry friction have been considered in the paper. Particular attention has been paid to the association of dry friction with the Nesterov accelerated gradient method. Convergence proofs are based on Lyapunov's analysis and the use of appropriate energy functions. Many issues remain largely open and require further investigations, such as, for example, the study of the algorithm (IPGD-NV) combining dry friction with the Nesterov accelerated gradient method, in the case of the asymptotic vanishing viscosity. As the numerical experiments highlight, it would be interesting to carry out the same convergence analysis and prove the finite convergence of the algorithms (IPGDF-NV) and (IPGDF-NV-Variant). The study of inertial proximal-gradient optimization algorithms associating the Hessian-driven damping with the dry friction would be another important subject. In our context, splitting methods for additively structured optimization problems have been considered in the case of the Lasso problem. It would be interesting to develop splitting methods using dry friction for general structured problems. The study of all these issues is beyond the scope of this paper and requires further research.

A Auxiliary results

A.1 Finite time convergence of the continuous dynamic

Theorem A.1 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a C^1 function whose gradient is Lipschitz continuous, and let $\phi : \mathcal{H} \rightarrow \mathbb{R}$ be a convex continuous function that satisfies (DF). Suppose that the function $\gamma : [t_0, +\infty[\rightarrow \mathbb{R}_+$ belongs to $L^1([t_0, T])$ for any $T > t_0$. Then, the following properties hold:*

a) For any Cauchy data $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$, there exists a unique strong global solution of the Heavy Ball system with Dry Friction

$$\text{(HBDF)} \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0, \quad (\text{A.1})$$

satisfying $x(t_0) = x_0$, and $\dot{x}(t_0) = \dot{x}_0$.

b) For any solution trajectory x of (HBDF) we have:

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

(ii) The limit point x_∞ is an equilibrium point of (HBDF), i.e.

$$-\nabla f(x_\infty) \in \partial\phi(0). \quad (\text{A.2})$$

(iii) If

$$-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0)),$$

then there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$.

Proof. An existence proof based on a regularization technique, by using the Moreau-Yosida approximation of ϕ , was given in [3] in a finite dimensional setting. We present here an original proof of the existence and uniqueness part a) of Theorem A.1, in a general Hilbert space, which is based on the study of evolution equations governed by the Lipschitz perturbation of maximally monotone operators (see [26]). It uses in an essential way that ∇f is Lipschitz continuous over the entire space \mathcal{H} .

Write (HBDF) as

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \partial\phi(\dot{x}(t)) \ni -\nabla f\left(x_0 + \int_{t_0}^t \dot{x}(\tau)d\tau\right).$$

Setting $u(t) := \dot{x}(t)$, this amounts to solving the first-order evolution equation

$$\dot{u}(t) + \gamma(t)u(t) + \partial\phi(u(t)) + \nabla f\left(x_0 + \int_{t_0}^t u(\tau)d\tau\right) \ni 0$$

with the Cauchy data $u(t_0) = \dot{x}_0$. Let us introduce the (non-local) operator

$$F(u)(t) = \nabla f\left(x_0 + \int_{t_0}^t u(\tau)d\tau\right).$$

Thus, we have to solve

$$\dot{u}(t) + \gamma(t)u(t) + \partial\phi(u(t)) + F(u)(t) \ni 0. \quad (\text{A.3})$$

For any two trajectories u and v , we have

$$\|F(u)(t) - F(v)(t)\| \leq L \int_{t_0}^t \|u(\tau) - v(\tau)\|d\tau,$$

where L is the Lipschitz constant of ∇f . Following the approach developed in [26, Proposition 3.12, page 106], we consider the sequence (u_n) defined recursively by

$$\dot{u}_{n+1}(t) + \gamma(t)u_{n+1}(t) + \partial\phi(u_{n+1}(t)) + F(u_n)(t) \ni 0. \quad (\text{A.4})$$

Given u_n , the existence and uniqueness of u_{n+1} solution of (A.4) with $u_{n+1}(0) = \dot{x}_0$ is ensured by the classical results concerning the evolution equations governed by subdifferentials of convex functions (see

[26, Theorem 3.6, page 72], [9, Theorem 17.2.5]). Let's give $T > t_0$. According to the above Lipschitz continuity property of F , the monotonicity of $\partial\phi$, and $\gamma(t) \geq 0$, we have for all $0 \leq t \leq T$

$$\|u_{n+1}(t) - u_n(t)\| \leq L(t - t_0) \int_{t_0}^t \|u_n(\tau) - u_{n-1}(\tau)\| d\tau,$$

which gives

$$\|u_{n+1}(t) - u_n(t)\| \leq \frac{(L(t - t_0))^n}{n!} t^2 \|u_1 - u_0\|_{L^\infty(t_0, T)}.$$

This implies that (u_n) is a Cauchy sequence for the uniform convergence on $[t_0, T]$. Consequently, it converges uniformly on $[t_0, T]$ to a solution u of (A.3). So, this uniquely define $u = \dot{x}$, and at the same time x which is given by to $x(t) = x_0 + \int_{t_0}^t u(\tau) d\tau$.

For part b), we refer to [3, Theorem 3.2]. ■

Remark A.1 With the condition $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$, the finite time convergence of the trajectory to a stationary point of the dynamic (HBDF) is ensured, i.e. there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$. In addition, an estimate of the final time could be given. In fact, we can show, by integrating the differential inequality (2.19), that

$$t_1 \leq t_0 + \frac{2\|\dot{x}(t_0)\|}{\text{dist}\left(-\nabla f(x_\infty), \text{boundary}(\partial\phi(0))\right)},$$

where t_0 is the first time instant such that

$$\nabla f(x(t)) \in \nabla f(x_\infty) + B(0, \varepsilon), \text{ for all } t \geq t_0, \text{ with } \varepsilon = \frac{1}{2} \text{dist}\left(-\nabla f(x_\infty), \text{boundary}(\partial\phi(0))\right).$$

Remark A.2 The conclusions of Theorem A.1 are valid under the key assumption $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$. Since the boundary of the convex set $\partial\phi(0)$ has an empty interior, it is reasonable to think that the circumstances leading to the relation $-\nabla f(x_\infty) \in \text{boundary}(\partial\phi(0))$ are ‘‘exceptional’’. More precisely, we conjecture that generically with respect to the initial data $(x_0, \dot{x}_0) \in \mathbb{R}^n \times \mathbb{R}^n$, the point $x_\infty = \lim_{t \rightarrow +\infty} x(t)$ satisfies the condition $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$. Consequently, this would give a generic finite time stabilization result in the case of dry friction.

Let us give a counter-example to convergence in finite time when the condition $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$ is not satisfied, i.e. $-\nabla f(x_\infty) \in \text{boundary}(\partial\phi(0))$. For that purpose, take $\mathcal{H} = \mathbb{R}$, $\phi := |\cdot|$ (so that $\partial\phi(0) = [-1, 1]$), $\gamma = 2$ and $f := |\cdot|^2/2$. The differential inclusion (HBDF) then reads

$$\ddot{x}(t) + \text{sign}(\dot{x}(t)) + 2\dot{x}(t) + x(t) \ni 0.$$

Let us choose as initial conditions $x(0) = -2$ and $\dot{x}(0) = 1$. The unique solution of (HBDF) is given by $x(t) = -1 - e^{-t}$, $t \geq 0$. The trajectory tends toward the value $x_\infty = -1$, which satisfies $-f'(x_\infty) = 1 \in \text{boundary}(\partial\phi(0))$. However the convergence does not hold in a finite time.

Remark A.3 It is natural to know if convergence in finite time is specific to the dry friction situation $0 \in \text{int}(\partial\phi(0))$. To answer this question Amann-Diaz [7] and Diaz-Linan [29] considered the damped oscillator in $\mathcal{H} = \mathbb{R}$

$$\ddot{x}(t) + |\dot{x}(t)|^{\alpha-1} \dot{x}(t) + x(t) = 0,$$

where $\alpha \in]0, 1[$. This corresponds to a sub-linear friction, the case of dry friction corresponds to the limiting case $\alpha = 0$. They have shown the existence of two curves in the phase space such that, for the solution trajectories with initial data (x_0, \dot{x}_0) belonging to these two curves, there is finite time stabilization at the origin. Using both energetic and geometrical arguments, they showed that for many other initial data, the solution tends to zero in infinite time, at the rate $t^{-\frac{\alpha}{1-\alpha}}$.

A.2 Proof of Theorem 2.3

The proof is similar to that of Theorem 2.1, with some technical adaptation. First, note that taking $x = \frac{2\gamma}{L}$ and $y = \frac{1}{\gamma}$ into the general inequality $\inf\{x, y\} \leq \sqrt{xy}$ for $x, y \geq 0$, gives

$$\inf\left\{\frac{2\gamma}{L}, \frac{1}{\gamma}\right\} \leq \sqrt{\frac{2}{L}}.$$

Therefore the assumption $h < \inf\left\{\frac{2\gamma}{L}, \frac{1}{\gamma}\right\}$ implies

$$h < \sqrt{\frac{2}{L}}. \quad (\text{A.5})$$

(i) Let's first establish energy estimates. Without ambiguity, we write $\partial\phi$ to designate the element belonging to this set. Taking the dot product of (2.33) with $\frac{1}{h}(x_{k+1} - x_k)$, we obtain

$$\begin{aligned} & \left\langle \frac{1}{h}(x_{k+1} - x_k) - \frac{1}{h}(x_k - x_{k-1}), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \frac{\gamma}{h} \langle x_{k+1} - x_k, x_k - x_{k-1} \rangle \\ & + h \left\langle \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right), \frac{1}{h}(x_{k+1} - x_k) \right\rangle + \langle \nabla f(x_k), x_{k+1} - x_k \rangle = 0. \end{aligned} \quad (\text{A.6})$$

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

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

So, according to (A.6), we have

$$\langle X_{k+1} - X_k, X_{k+1} \rangle + \gamma h \langle X_{k+1}, X_k \rangle + hr\|X_{k+1}\| + h \langle \nabla f(x_k), X_{k+1} \rangle \leq 0. \quad (\text{A.7})$$

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

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

Combining the above inequality with (A.7), we obtain

$$\langle X_{k+1} - X_k, X_{k+1} \rangle + \gamma h \langle X_{k+1}, X_k \rangle + hr\|X_{k+1}\| + f(x_{k+1}) - f(x_k) - \frac{Lh^2}{2}\|X_{k+1}\|^2 \leq 0.$$

Equivalently

$$\left(1 - \frac{Lh^2}{2}\right)\|X_{k+1}\|^2 + (h\gamma - 1)\langle X_k, X_{k+1} \rangle + hr\|X_{k+1}\| + (f(x_{k+1}) - f(x_k)) \leq 0. \quad (\text{A.8})$$

By (A.5) we have $1 - \frac{Lh^2}{2} > 0$, and by assumption $h\gamma < 1$. From (A.8) we infer

$$\begin{aligned} & \frac{1}{2}\left(1 - \frac{Lh^2}{2}\right)(\|X_{k+1}\|^2 - \|X_k\|^2) + \frac{1}{2}\left(1 - \frac{Lh^2}{2}\right)\|X_{k+1}\|^2 + \frac{1}{2}\left(1 - \frac{Lh^2}{2}\right)\|X_k\|^2 \\ & - (1 - h\gamma)\|X_k\|\|X_{k+1}\| + hr\|X_{k+1}\| + (f(x_{k+1}) - f(x_k)) \leq 0, \end{aligned} \quad (\text{A.9})$$

where we used Cauchy-Schwarz inequality. Elementary algebra (sign of a polynomial of the second degree) gives

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

under the condition $\Delta = (1 - h\gamma)^2 - (1 - \frac{Lh^2}{2})^2 \leq 0$. This is equivalent to $h \leq \frac{2\gamma}{L}$, and hence is satisfied by assumption. To summarize the results, in terms of

$$E_k := \frac{1}{2} \left(1 - \frac{Lh^2}{2}\right) \left\| \frac{1}{h} (x_k - x_{k-1}) \right\|^2 + (f(x_k) - \inf f),$$

we have obtained

$$E_{k+1} - E_k + r \|x_{k+1} - x_k\| \leq 0. \quad (\text{A.10})$$

According to the nonnegativity of E_k , and $r > 0$, we deduce from (A.10) 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

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

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 (2.33), rewrite it as follows:

$$-\frac{1}{h^2} (x_{k+1} - 2x_k + x_{k-1}) - \frac{\gamma}{h} (x_k - x_{k-1}) - \nabla f(x_k) \in \partial\phi \left(\frac{1}{h} (x_{k+1} - x_k) \right). \quad (\text{A.11})$$

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

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

A.3 Proof of Theorem 2.4

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

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

On the other hand, using inequality (A.6), we get

$$\langle X_{k+1} - X_k, X_{k+1} \rangle + \gamma h \langle X_{k+1}, X_k \rangle + h\phi(X_{k+1}) + h \langle \nabla f(x_k), X_{k+1} \rangle \leq 0. \quad (\text{A.12})$$

By combining the two above inequalities we get, for every $k \geq k_0$

$$\|X_{k+1}\|^2 - (1 - h\gamma) \langle X_{k+1}, X_k \rangle + \varepsilon \|x_{k+1} - x_k\| \leq 0. \quad (\text{A.13})$$

a) Let's neglect the nonnegative term $\varepsilon \|x_{k+1} - x_k\| \geq 0$ in (A.13). According to the hypothesis $1 - h\gamma > 0$, and the Cauchy-Schwarz inequality, we deduce from (A.13) that

$$\|X_{k+1}\|^2 \leq (1 - h\gamma) \langle X_{k+1}, X_k \rangle \leq (1 - h\gamma) \|X_{k+1}\| \|X_k\|$$

which gives

$$\|X_{k+1}\| \leq (1 - h\gamma) \|X_k\|.$$

From this we immediately deduce the geometric convergence towards zero of the velocities:

$$\|x_{k+1} - x_k\| \leq (1 - h\gamma)^k \|x_1 - x_0\|. \quad (\text{A.14})$$

Set $q := 1 - h\gamma$ and $C = \|x_1 - x_0\|$. From (A.14) we easily deduce that

$$\|x_k - x_\infty\| \leq C \frac{q^k}{1-q} = \frac{1}{h\gamma} (1-h\gamma)^k \|x_1 - x_0\|.$$

Hence, the geometric convergence of (x_k) to its limit x_∞ , which ends item (i).

(ii) Let's show now that under the stronger assumption $\|\nabla f(x_\infty)\| < r$ where $B(0, r) \subset \partial\phi(0)$, then the finite convergence property is satisfied. Write the algorithm (IPGDF-variant) as follows:

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

Set $\xi_k := \frac{1}{h}(1-h\gamma)(x_k - x_{k-1}) - h\nabla f(x_k)$. So, we have

$$\frac{1}{h}(x_{k+1} - x_k) = \text{prox}_{h\phi}(\xi_k). \quad (\text{A.15})$$

To show the finite convergence property, we need to show that $x_{k+1} - x_k = 0$ for k large enough. According to (A.15) and Lemma 1.2 it is enough to prove that

$$\frac{1}{h}\|\xi_k\| \leq r. \quad (\text{A.16})$$

By the triangle inequality and the L -Lipschitz continuity of ∇f we have

$$\frac{1}{h}\|\xi_k\| = \left\| \frac{1}{h^2}(1-h\gamma)(x_k - x_{k-1}) - \nabla f(x_k) \right\| \leq \frac{1}{h^2}\|x_k - x_{k-1}\| + \|\nabla f(x_\infty)\| + L\|x_k - x_\infty\|. \quad (\text{A.17})$$

When $k \rightarrow +\infty$, the whole right-hand side of the inequality (A.17) tends to $\|\nabla f(x_\infty)\|$. Since $\|\nabla f(x_\infty)\| < r$, the condition (A.16) will be satisfied for k large enough, which gives the finite convergence.

References

- [1] S. Adly, *A variational approach to nonsmooth dynamics: applications in unilateral mechanics and electronics*, Springer Briefs in Mathematics, 2017.
- [2] S. Adly, D. Goeleven, *A stability theory for second order nonsmooth dynamical systems with application to friction problems*, Journal de Mathématiques Pures et Appliquées, 83 (2004), pp. 17–51.
- [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] S. Adly, B. Brogliato, B.K. Le, *Well-posedness, robustness and stability analysis of a set-valued controller for Lagrangian systems*, SIAM J. Control Optim., 51 (2) (2013), pp. 1592–1614.
- [5] 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.
- [6] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., 81 (8) (2002), pp. 747–779.

- [7] H. Amann, J. I. Díaz, *A note on the dynamics of an oscillator in the presence of strong friction*, *Nonlinear Analysis*, 55 (2003), pp. 209–216.
- [8] 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*.
- [9] 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).
- [10] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, *J. Differential Equations*, 263 (2017), pp. 5412–5458.
- [11] H. Attouch, A. Cabot, *Convergence rates of inertial forward-backward algorithms*, *SIAM J. Optim.*, 28 (1) (2018), pp. 849–874.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$* , *ESAIM-COCV*, 25 (2019), published electronically.
- [16] 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.
- [17] H. Attouch, P.E. Maingé, P. Redont, *A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws*, *Differential Equations and Applications*, 4 (1) (2012), pp. 27–65.
- [18] 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.
- [19] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, 2019. hal-02193846.
- [20] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. I The continuous dynamical system*, *Communications in Contemporary Math*, 2 (1) (2000), pp.1–34.
- [21] 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.
- [22] B. Baji and A. Cabot, *An inertial proximal algorithm with dry friction: finite convergence results*, *Set-Valued Analysis*, 9 (1) (2006), pp. 1–23.
- [23] H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).

- [24] 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).
- [25] 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.
- [26] H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, 1972.
- [27] A. Chambolle, Ch. Dossal, *On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm*, *J. Optim. Theory Appl.*, 166 (2015), pp. 968–982.
- [28] A. Chambolle, T. Pock, *An introduction to continuous optimization for imaging*, *Acta Numerica*, 25 (2016), pp. 161–319.
- [29] J. I. Díaz, A. Liñán, *On the asymptotic behavior of a damped oscillator under a sublinear friction term*, *Rev. R. Acad. Cien. Serie A. Mat.*, 95 (1) (2001), pp. 155–160.
- [30] E. D. Dolan and J. J. Moré, *Benchmarking Optimization Software with Performance Profiles*, *Math. Program.*, 91 (2002), pp. 201–213.
- [31] 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.
- [32] A. Haraux, M. A. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, *J. Differential Equations*, 144 (2) (1998), pp. 313–320.
- [33] A. Haraux, M. A. Jendoubi, *The convergence problem for dissipative autonomous systems*, *Classical methods and recent advances*, Springer, 2015.
- [34] C. Lemarechal, C. Sagastizábal. *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, *SIAM J. Optim.* 7 (2) (1997), pp. 367–385.
- [35] 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.
- [36] 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.
- [37] J. Peypouquet, S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, *J. Convex Anal.*, 17 (3-4) (2010), pp. 1113–1163.
- [38] B.T. Polyak, *Some methods of speeding up the convergence of iterative methods*, *Z. Vyslist Math. Fiz.*, 4 (1964), pp. 1–17.
- [39] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).
- [40] R.T. Rockafellar and R. Wets, *Variational analysis*, Springer, Berlin, 1998.
- [41] J. W. Siegel, *Accelerated first-order methods: Differential equations and Lyapunov functions*, arXiv: Optimization and Control (math.OC):1903.05671v1 (2019).
- [42] 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.