Math. Program., Ser. B (2009) 116:173–191
DOI 10.1007/s10107-007-0109-5

FULL LENGTH PAPER

The gradient and heavy ball with friction dynamical systems: the quasiconvex case

X. Goudou · J. Munier

Received: 20 June 2005 / Accepted: 11 December 2006 / Published online: 11 May 2007

Abstract We consider the gradient system $\dot{x}(t) + \nabla\Phi(x(t)) = 0$ and the so-called heavy ball with friction dynamical system $\ddot{x}(t) + \lambda\dot{x}(t) + \nabla\Phi(x(t)) = 0$, as well as an implicit discrete (proximal) version of it, and study the asymptotic behavior of their solutions in the case of a smooth and quasiconvex objective function Φ. Minimization properties of trajectories are obtained under various additional assumptions. We finally show a minimizing property of the heavy ball method which is not shared by the gradient method: the genericity of the convergence of each trajectory, at least when Φ is a Morse function, towards local minimum of Φ.

Keywords Quasiconvex function · Minimization · Continuous dynamical system · Proximal algorithm · Asymptotic behavior

Mathematics Subject Classification (2000) 37N40 · 46N10 · 37L05

1 Introduction

We consider here the problem of minimizing a function $\Phi : H \to \mathbb{R}$ defined on H a real Hilbert space, with real values. This writes

$$\min_{x \in H} \Phi(x). \tag{1}$$

We assume that the set of minimizers $S := \text{argmin}\ \Phi$ is nonempty. Let us denote by $\langle \cdot, \cdot \rangle$ the scalar product on H and $|\cdot|$ the associated norm. Classically $\nabla\Phi(x)$ will be

X. Goudou (✉) · J. Munier
Institut de Mathématiques et Modélisation de Montpellier, UMR CNRS 5149,
Université Montpellier II, case courrier 051, Place Eugène Bataillon,
34095 Montpellier Cedex 5, France
e-mail: krystal@math.univ-montp2.fr

the gradient of Φ at x according to this scalar product. As this work deals with the case of a quasiconvex objective function Φ, let us recall here the basic notions about that concept.

Quasiconvexity of a function f (from some convex subset C of a vector space to $\mathbb{R}$) admits various definitions (see [17,18], or [11]). For example, in geometric terms, the function f is said quasiconvex if

$$\forall x, y \in C, \quad f(tx + (1-t)y) \leq \max(f(x), f(y))$$

which means that each sublevel sets $L_\alpha(f) := \{x \in C \text{ such that } f(x) \leq \alpha\}$ is convex.

When C is open, and f is continuously differentiable, an equivalent condition (see [10] for more details) is that for all x and y in C,

$$f(y) \leq f(x) \quad \Rightarrow \quad \langle \nabla f(x), y - x \rangle \leq 0. \tag{2}$$

Remark 1 An other equivalent condition is that, for all x and y in C, $f(y) < f(x) \Rightarrow \langle \nabla f(x), y-x \rangle \leq 0$. This condition must not be confused with the following: $\forall x, y \in C,\ f(y) < f(x) \Rightarrow \langle \nabla f(x), y - x \rangle < 0$, which claims that f is pseudoconvex: pseudoconvexity implies quasiconvexity, and a quasiconvex function f on some open convex subset is pseudoconvex if each critical points of f is a local minimum. See for example [9].

So quasiconvexity generalizes convexity. The interest of quasiconvex functions is that, while retaining some important aspects of convex functions, they present the advantage of having some stability properties that convex functions do not have. For example, for each constant λ, the function $\min\{f, \lambda\}$ is quasiconvex when f is quasiconvex, but may not be convex even if f is convex.

The notion of quasiconvexity considered here must not be confused with the Morrey quasiconvexity used in calculus of variations (see [20] for further details).

The notion of quasiconvexity appeared recently in control theory (see [13]). But its main domain of application is the value theory in economics (see [3,19,21,22,24,33] or [39]). In this area, each consumer is characterized by a preference $\succeq$ (which is a reflexive and transitive binary relation on consumptions). A natural psychological assumption is that the consumer tends to share out his consumption among all goods: if x and y are two consumptions, the consumption $tx + (1-t)y$ is, for all t in [0, 1], at least as desirable as the least desirable consumption between x and y. In other words, for each consumption x, the set $\{y \,/\, y \succeq x\}$ is convex. A function U from the consumption set to $\mathbb{R}$ is said to represent the preordering $\succeq$ if for every consumption x and y, y is preferred to x if and only if $U(y) \geq U(x)$. Such a function is called a utility function relatively to the consumer.

A utility function may not exist (consider for example the case of the lexicographic ordering). Some additional assumptions are required to ensure existence of a utility function. For example, it is shown in [22] that, when relation $\succeq$ is complete, the preference $\succeq$ can be represented by a continuous utility function if and only if, for each x, the sets $\{y \,/\, x \succeq y\}$ and $\{y \,/\, y \succeq x\}$ are closed. Of course, there is no uniqueness: if U represents $\succeq$, every function $h \circ U$ represents also $\succeq$, as soon as function h strictly increases from $\mathbb{R}$ to $\mathbb{R}$.

A consequence of the preceding psychological assumption is that the utility function (when exists) is quasiconcave (i.e., $-U$ is quasiconvex), and, for example, the classical consumer problem consists, for each consumer, in maximizing his utility function among the set of accessible commodities, under some budget constraints.

In this paper, we only deal with a regular (this will be precised later) quasiconvex function, defined from some real Hilbert space H to $\mathbb{R}$: Φ could be the difference between a desired level of utility, the goal, and the current utility

$$\Phi(x) = \bar{u} - U(x)$$

and we want to minimize it.

This framework is interesting in decision sciences too. Such links between optimization and decision theory are developed in [7].

Classically, dynamical systems can be used to solve the minimization problem (1), the most famous being the steepest descent equation

$$\dot{x}(t) + \nabla\Phi(x(t)) = 0 \tag{3}$$

in its continuous form. Many results exist relatively to that method, let us cite [14,15] at least for the case of a convex function Φ.

More recently, second order methods have been introduced in link with the same problem (1). We are working in this paper with the following one

$$\ddot{x}(t) + \lambda\dot{x}(t) + \nabla\Phi(x(t)) = 0 \tag{4}$$

together with initial conditions $x(0) = x_0$ and $\dot{x}(0) = v_0$. Here λ is a positive constant.

A complete reference on that method is [6], it presents the literature and the existing results for a smooth Φ (existence and uniqueness of a solution, asymptotic convergence in the case of a Morse function Φ, applications in unconstrained and global optimization), as well as a mechanical interpretation of the model. The convex case, which is the closest to our case is studied in [1]. Let us present the other works on the topic. This continuous dynamical system appeared in the literature in the last twenty years. Before, Polyack [36] introduced the terminology and studied an explicit discrete version of it. Some results for an analytic function Φ are proved in [28]. Still for a convex function, [5] deals with the asymptotic behavior of the solutions of a controlled version of (4). Finally [4] generalizes (4) in finite dimensional case for a convex and only lower semicontinuous Φ. In this case, the gradient $\nabla\Phi$ has to be replaced by its generalization in convex analysis, namely the subdifferential $\partial\Phi$. Applications of the nonsmooth case are first elastic shocks in mechanics, and optimization and control of constrained problems.

Interest of that system lies in the inertial effects, which are quite clear in the mechanical interpretation: this system approximately describes the motion of a material point $M(t) = (x(t), \Phi(x(t)))$ ("a ball") on a profile defined by Φ. While the *first order* steepest descent system (3) is unable to cross any non-minimum critical point of Φ, the presence in (4) of an inertial (second order in time) term captures some exploratoring properties of the ball's motion. It is a step towards global optimization, as shown in [6]. Nevertheless, despite their explorating properting, (4) cannot ensure convergence to a global minimizer in general. We obtain also in some cases a generic (with respect to initial conditions: position and velocity) convergence of the trajectories to local minima (see the Appendix), which is not the case of first order methods. That is furthermore interesting in the framework of economics or decision sciences, in the sense that it furnishes a model including inertia for agents behavior. Such models are quite new and for more details, see [7].

In this paper, we also study the implicit discretization of (4), which leads to an inertial proximal method. Indeed, given $x_0, x_1 \in H \times H$, it writes for all $k \geq 1$

$$\frac{\frac{x_{k+1}-x_k}{h_k} - \frac{x_k - x_{k-1}}{h_{k-1}}}{h_k} + \lambda \frac{x_{k+1} - x_k}{h_k} + \nabla \Phi(x_{k+1}) = 0$$

for adaptative positive time steps $(h_k)_{k\in\mathbb{N}}$. We can see the sequence $(x_k)_{k\in\mathbb{N}}$ as a discrete approximation of the sequence $(x(t_k))_{k\in\mathbb{N}}$, where $x(\cdot)$ is the solution of (4) with the initial conditions $x(0) = x_0$ and $\dot{x}(0) = \frac{x_1 - x_0}{h_0}$, and $t_k = \sum_{i=0}^{k-1} h_i$. The previous equality is equivalent to

$$x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) + \beta_k \nabla \Phi(x_{k+1}) = 0 \tag{5}$$

with for every $k \geq 1$, $\alpha_k = \frac{h_{k-1}}{h_k} \frac{1}{1+\lambda h_k}$ and $\beta_k = \frac{h_k^2}{1+\lambda h_k}$. This is indeed a proximal method, since it can also be written

$$x_{k+1} \in \text{argmin}\,\{\Phi(x) + \frac{1}{2\beta_k}|x - x_k - \alpha_k(x_k - x_{k-1})|^2, x \in H\}.$$

As guideline we follow the direction of the paper [1], that introduced the algorithm for nonsmooth convex minimisation, as well as [2] that extended it to general maximal monotone operators, and introduced the “prox-inertial” terminology. As for the continuous system, the trajectories have good asymptotic properties, in regards with the problem (1). The original proximal method appeared in the seminal works [32,37] (see also [31]), as a sort of regularization technique in witch the original problem is replaced with a sequence of better behaved problems of same type. Although it was not its motivation, the proximal method can be viewed as an implicit discretization of the steepest descent method (3), and is called proximal because the current iterate can be seen as a solution of a minimization problem involving a distance function.

The results presented here have been obtained under similar hypotheses for an integrodifferential equation of Volterra type (see [25])

$$\dot{x}(t) + \nabla\Phi(x(t)) + \int_0^t a(t-s)\nabla\Phi(x(s))ds = 0. \tag{6}$$

The paper organizes as follows. We introduce in Sect. 2 the assumptions and the main result in the continuous second order quasiconvex case. The proof of this result is the topic of Sect. 5. This section is preceded by a recall of the proof of the convergence of the second order system (4) in the convex case (Sect. 3), and by an original proof of the convergence of the first order system (3) in the quasiconvex case (Sect. 4). Then Sect. 6 deals with the discrete version, that gives a "prox-inertial" method, and the result in that case is proved in Sect. 7. The end of the paper is devoted to the genericity of the convergence to a local minimum of Φ if the latter is a Morse function.

2 The continuous dynamical system

Throughout this case, we assume that the following hold

$$\Phi \text{ is } C^{1,1}, \text{ quasiconvex and } S := \operatorname{argmin} \Phi \neq \emptyset. \tag{7}$$

The first means that Φ is continuously differentiable with $\nabla\Phi$ Lipschitz continuous on the bounded subsets of H. These assumptions are those of Theorem 3.1 in [6], whose conclusions are recalled here, using classical notations: $L^\infty(0,+\infty;H) = \{f : [0,+\infty) \to H, \sup_{t\geq 0} |f(t)| < +\infty\}$ and $L^2(0,+\infty;H) = \{f : [0,+\infty) \to H, \int_0^\infty |f(t)|^2 dt < +\infty\}$

Proposition 1 *Under Assumptions 7 the following hold.*

(i) *For all (x_0, x_1) in $H \times H$, there exists a unique solution $x(t)$ of the Cauchy problem associated with (4). This solution is defined on the whole interval $[0,+\infty)$ and is of class C^2.*
(ii) *For every trajectory $x(t)$ of (4), the corresponding energy $E(t) = \frac{1}{2}|\dot{x}(t)|^2 + \Phi(x(t))$ is decreasing on $[0,+\infty)$ and bounded from below, and hence converges to some real value E_∞. Moreover,*

$$\dot{x} \in L^\infty(0,+\infty;H) \cap L^2(0,+\infty;H).$$

(iii) *Assuming moreover that x is in $L^\infty(0,+\infty;H)$, then we have*
- *$\dot{x}$ and $\ddot{x}$ belong to $L^\infty(0,+\infty;H)$,*
- $\lim_{t\to+\infty} \dot{x}(t) = 0$ *and* $\lim_{t\to+\infty} \ddot{x}(t) = 0$,
- $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$ *and* $\lim_{t\to+\infty} \Phi(x(t)) = E_\infty$.

This result holds under our assumptions. It ensures that trajectories of (4) exist globally. Quasiconvexity has not yet been used. It brings much more. Indeed the main result stated here goes further.

Theorem 1 *Under Assumptions 7, for every trajectory $x(\cdot)$ of the second order system (4) the following properties hold.*

(i) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$ *and* $\lim_{t\to+\infty} \Phi(x(t)) = E_\infty$.
(ii) *$x(t)$ weakly converges in H towards some x_∞.*
(iii) $\nabla\Phi(x_\infty) = 0$.
(iv) *If the limit point x_∞ does not belong to $S = \operatorname{argmin}\Phi$, the convergence is strong.*

Moreover

(v) *If S has a nonempty interior, the convergence is strong.*
(vi) *If Φ is an even function, that is*

$$\forall x \in H, \quad \Phi(-x) = \Phi(x)$$

the convergence is strong.

In [1] Alvarez established statements (i), (v) and (vi) for the dynamical system (4) in case of convex potential Φ. Under this convexity hypothesis, the statements in the two last items were introduced first in [14,15], respectively, for a first order differential inclusion generalizing (3). Afterwards Baillon in [12] exhibits a counterexample showing that, in the first order case, no more than weak convergence could be expected. It is worth noticing that Baillon's counterexample has been used in [26] (Sect. 5) in order to construct a counterexample for the strong convergence of the Proximal Point Algorithm. Then Jendoubi and Polàcik proposed in [29] an elegant method to obtain counterexample of convergence of (4) from any counterexample of convergence of (3).

3 Convergence in the convex second order case

In order to enlighten the difficulties raised by the quasiconvex hypothesis, let us at first recall the main ideas behind the proof of the weak convergence of trajectories of (4) in the Convex case. Although the ideas are essentially the same, the following demonstration is differently structured than the one proposed in [1].

The mean tool is a lemma due to Opial (see [34] or [6] and reference therein), which can be formulated as follows:

Lemma 1 (Opial) *In a Hilbert space H, let $\mathcal{A}$ be the set of limit points of some trajectory $x : [0, +\infty) \to H$. Let us assume*:

(1) $\mathcal{A} \neq \emptyset$
(2) $\forall \tilde{x} \in \mathcal{A}, \ \lim_{t \to +\infty} |x(t) - \tilde{x}|$ *exists*

Then there exists x_∞ in $\mathcal{A}$, such that $w\!-\!\lim_{t\to+\infty} x(t) = x_\infty$.

The proof of the convergence of the trajectories of the second order system (4) consists in proving that:

(a) $\forall \hat{x} \in S, \ \lim_{t \to +\infty} |x(t) - \hat{x}|$ exists;
(b) $\mathcal{A} \subset S$

The property (a) together with the nonemptiness of S show that every trajectory $x : [0, +\infty) \to H$ is bounded, which implies Assumption 1 of Opial's lemma.

In order to establish (a), consider for some $\hat{x} \in S$ the function $h(t) := \frac{1}{2}|x(t) - \hat{x}|^2$. Let $\hat{w}(t) = \langle \hat{x} - x(t), \nabla\Phi(x(t))\rangle$. An easy calculation leads to:

$$\ddot{h}(t) + \lambda \dot{h}(t) = |\dot{x}(t)|^2 + \hat{w}(t)$$

By monotonicity of $\nabla\Phi$, the function $\hat{w}(t)$ is non negative, and then, as $|\dot{x}|$ belongs to $L^2(0,+\infty;H)$, we conclude that $(\dot{h})^+ \in L^1(0,+\infty;H)$. Since h is bounded from below, property (a) follows.

In order to establish property (b), remark that property (a) leads to the boundedness of the trajectory $x(\cdot)$, which implies from a general result concerning the steepest descent dynamic (3) that $\lim_{t\to+\infty}\nabla\Phi(x(t)) = 0$. Since, by convexity, the gradient $G := \nabla\Phi(\cdot)$ has a closed graph in $w-H\times s-H$ topology, we conclude that $\nabla\Phi(\tilde{x}) = 0$ for every $\tilde{x}$ in $\mathcal{A}$, which together with the classical first order characterization of infima of a convex smooth function, leads to property (b).

The difficulty in attempting to extend this convergence result to the quasiconvex case is due to the lack of closedness of G and of monotonicity of $\nabla\Phi(\cdot)$, and to the fact that for general quasiconvex potential, the property (b) may fail.

4 Convergence in the first order quasiconvex case

We consider in this section the system (3) with a smooth potential Φ. It is well known that the steepest descent system admits for every initial condition x_0 in H a unique global solution x on $[0,+\infty)$ (cf [27]). Yet the following result seems not to appear in the literature:

Theorem 2 *Under Assumption 7, each trajectory $x(\cdot)$ of the first order system (3) weakly converges towards some point x_∞ such that $\nabla\Phi(x_\infty) = 0$. Moreover if at least one of the following condition is satisfied*:

(a) *$x_\infty \notin$ argmin Φ*
(b) *S as a nonempty interior*
(c) *Φ is even*

then the convergence is strong.

Lemma 2 *Let $x(\cdot)$ be a trajectory of (3). Suppose there exist $\tilde{x} \in H$ and $T \geq 0$ such that*

$$\forall t \geq T,\ \ (x(t) \in S \text{ or } \Phi(\tilde{x}) < \Phi(x(t))).$$

Then the function $\tilde{w}(t) := \langle \tilde{x} - x(t), \nabla\Phi(x(t))\rangle$ is nonpositive on $[T,+\infty)$ and belongs to $L^1(0,+\infty;\mathbb{R})$, the function $\Phi(x(t))$ is nonincreasing and converges towards some $E_\infty \geq \inf\Phi$, the function $h(t) := \frac{1}{2}|x(t)-\tilde{x}|^2$ satisfies $\dot{h} \in L^1(0,+\infty;\mathbb{R})$ and thus $h(t)$ converges when $t \to +\infty$.

Proof The nonnegative function $h(t) := \frac{1}{2}|x(t)-\tilde{x}|^2$ verifies $\dot{h}(t) = \langle \tilde{x} - x(t), \nabla\Phi(x(t))\rangle$, which shows that $\dot{h}(t) = \tilde{w}(t)$, so that by characterization (2) of quasiconvexity, we conclude that $\tilde{w}(t) \leq 0$ for every $t \geq T$ and then that h is nonincreasing on $[T,+\infty)$. The nonpositivity of $\dot{h}$ on $[T,+\infty)$, together with nonnegativity of h implies that $\dot{h}$ belongs to $L^1(0,+\infty;\mathbb{R})$, which leads to the existence of $\lim_{t\to+\infty} h(t)$.

Moreover, the steepest descent dynamic implies that function $\Phi(x(\cdot))$ is nonincreasing. The potential Φ being minorized, $E_\infty := \lim_{t\to+\infty}\Phi(x(t))$ exists.

So Lemma 2 results. □

We are now able to prove Theorem 2:

If in Lemma 2 we take $\tilde{x} \in \operatorname{argmin} \Phi$, we conclude that trajectory $x(\cdot)$ is bounded, so that the set $\mathcal{A}$ of its weak limit points is nonempty, which is Assumption 1 of the preceding Opial's lemma.

Moreover, in view of function $\Phi(x(t))$ being decreasing, we have:

$$(\alpha)\ \ \forall \tau \geq 0, \quad \lim_{t \to +\infty} \Phi(x(t)) \leq \Phi(x(\tau))$$

Since Φ is continuous and quasiconvex, Φ is weakly continuous, and so:

$$(\beta)\ \ \forall \tilde{x} \in \mathcal{A}, \quad \Phi(\tilde{x}) \leq \lim_{t \to +\infty} \Phi(x(t)) =: E_\infty$$

If $E_\infty = \inf \Phi$, we conclude from (β) that for each $\tilde{x} \in \mathcal{A}$ we have $\Phi(\tilde{x}) \leq E_\infty = \inf \Phi$, and so $\mathcal{A} \in \operatorname{argmin} \Phi$. Moreover, comparison of (α) and (β) shows that the condition of Lemma 2 is fulfilled with $\tilde{x} \in \operatorname{argmin} \Phi$ and $T = 0$. Thus from Lemma 2 the function $t \mapsto |x(t) - \tilde{x}|$ converges for each $\tilde{x} \in \mathcal{A}$, which is Assumption 2 of Opial's lemma. We then conclude that the trajectory $x(\cdot)$ weakly converges. Let $x_\infty \in \mathcal{A}$ be the limit. Since by hypothesis $\mathcal{A} = \operatorname{argmin} \Phi$, the first order optimality condition leads to $\nabla \Phi(x_\infty) = 0$.

If $E_\infty > \inf \Phi$, let us first establish the following lemma:

Lemma 3 *Let $x : [0, +\infty) \to H$ be a C^1 function such that there is some point $\tilde{x}$ in H and a number $r > 0$ and $T \geq 0$ satisfying*:

$$\forall t \geq T \quad \forall y \in \bar{B}(\tilde{x}, r) \quad \Phi(y) \leq \Phi(x(t))$$

Then for all $t \geq T$ we have: $r|\nabla \Phi(x(t))| \leq -\tilde{w}(t)$, *where* $\tilde{w}(t) = \langle \tilde{x} - x(t), \nabla \Phi(x(t)) \rangle$

Proof For $d \in H$ with $|d| = 1$, the point $y := \tilde{x} + rd$ lies in $\bar{B}(\tilde{x}, r)$. As $\Phi(y) \leq \Phi(x(t))$ for all $t \geq T$, we conclude by quasiconvexity of Φ that: $\langle y - x(t), \nabla \Phi(x(t)) \rangle \leq 0$. This writes

$$\tilde{w}(t) := \langle \tilde{x} - x(t), \nabla \Phi(x(t)) \rangle \leq -r \langle d, \nabla \Phi(x(t)) \rangle.$$

If $\nabla \Phi(x(t)) \neq 0$, choose $d = \frac{\nabla \Phi(x(t))}{|\nabla \Phi(x(t))|}$, which gives

$$r|\nabla \Phi(x(t))| \leq -\tilde{w}(t),$$

this inequality being obviously true if $\nabla \Phi(x(t)) = 0$. □

We are now able to prove:

Lemma 4 *If $x(\cdot)$ is a trajectory of the steepest descent system* (3) *satisfying the assumption of Lemma* 3, *then $x(\cdot)$ strongly converges, and the limit x_∞ verifies* $\nabla \Phi(x_\infty) = 0$.

Proof Lemma 2 ensures that $\tilde{w}$ lies in $L^1(0, +\infty; \mathbb{R})$. Together with Lemma 3 this prove that $|\nabla\Phi(x(t))|$ is in $L^1(0, +\infty; \mathbb{R})$. The steepest descent dynamic then implies that $\dot{x}$ is in $L^1(0, +\infty; \mathbb{R})$, so that the trajectories x strongly converges. Let denote x_∞ the limit. From proposition (1) (iii) and the strong continuity of $\nabla\Phi$ we conclude that $\nabla\Phi(x_\infty) = 0$. □

Each assumptions $E_\infty > \inf \Phi$ or $\text{int}(\text{argmin}\, \Phi) \neq \emptyset$ implies the condition of Lemma 3 which by Lemma 4 achieves the proof of first part and points (a) and (b) of Theorem 2.

Let us now prove part (c) of this theorem: we first consider, for some $0 \leq t \leq t_0$, the function: $\mu(t) := |x(t) - x(t_0)|^2 - 2|x(t)|^2$.
An easy calculation show that: $\dot{\mu}(t) = 2\langle \dot{x}(t), -x(t) - x(t_0)\rangle$, which by steepest descent dynamic leads to: $\dot{\mu}(t) = -2\langle \nabla\Phi(x(t)), -x(t) - x(t_0)\rangle$. But as $\Phi(x(\cdot))$ is nonincreasing, we have $\Phi(x(t_0)) \leq \Phi(x(t))$. The symmetry of Φ then gives $\Phi(-x(t_0)) \leq \Phi(x(t))$, and, by quasiconvexity: $\langle \nabla\Phi(x(t)), -x(t_0) - x(t)\rangle \leq 0$. So, the function μ is nonincreasing, and $\mu(t_0) \geq \mu(t)$. Hence, $|x(t) - x(t_0)|^2 \leq 2(|x(t)|^2 - |x(t_0)|^2)$. Finally, since $0 \in \text{argmin}\, \Phi$, lemma 2 shows that $t \mapsto |x(t)|^2$ converges when $t \to +\infty$, which implies by preceding inequality that for any sequence $(\tau_n)_n$ with $\tau_n \to +\infty$, the sequence $x(\tau_n)$ is a Cauchy sequence. Since $x(\cdot)$ weakly converges, the weak limit is strong, which achieve the proof of Theorem 2.

5 The second order quasiconvex case: proof of Theorem 1

The proof is divided into four parts. The first concerns the first item of the theorem, the second deals with the three following parts of Theorem 1, while the third and the fourth ones prove items (v) and (vi), respectively.

5.1 Proof of (i)

The following lemma is an extended analogue of Lemma 2 in the second order case:

Lemma 5 *Let $x(\cdot)$ be a trajectory of (4). Suppose there exist $\tilde{x} \in H$ and $T \geq 0$ such that*

$$\forall t \geq T, \ (x(t) \in S \quad or \ \ \Phi(\tilde{x}) < \Phi(x(t))).$$

Then we have

- *The function $\tilde{w}(t) = \langle \tilde{x} - x(t), \nabla\Phi(x(t))\rangle$ is nonpositive on $[T, +\infty)$ and belongs to $L^1(0, +\infty; \mathbb{R})$.*
- *The function $h(t) = \frac{1}{2}|x(t) - \tilde{x}|^2$ satisfies $\dot{h} \in L^1(0, +\infty; \mathbb{R})$. Then $h(t)$ converges when $t \to +\infty$.*
- *The following convergence results hold*

$$\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$$
$$\lim_{t\to+\infty} \Phi(x(t)) = E_\infty.$$

Proof • The fact that $\tilde{w}(t) \leq 0$ for $t \geq T$ comes directly from characterization (2) of quasiconvexity when $\Phi(\tilde{x}) < \Phi(x(t))$, and is obvious when $x(t) \in S$. The function $\tilde{w}$ is locally integrable. We prove the other items before proving the integrability over all $(0, +\infty)$.

- As in the convex case, see [1], the scalar function h satisfies a useful differential inequality. Indeed, computing the derivatives of h: $\dot{h}(t) = \langle x(t) - \tilde{x}, \dot{x}(t)\rangle$ and $\ddot{h}(t) = |\dot{x}(t)|^2 + \langle x(t) - \tilde{x}, \ddot{x}(t)\rangle$, we observe that

$$\ddot{h}(t) + \lambda \dot{h}(t) = |\dot{x}(t)|^2 + \tilde{w}(t).$$

  According to Proposition 1, $|\dot{x}|^2 \in L^1(0, +\infty; \mathbb{R})$. On the other hand, $\tilde{w}$ is locally integrable, and nonpositive for $t \geq T$. We can then construct a function $g \in L^1(0, +\infty; \mathbb{R})$ such that $\ddot{h}(t) + \lambda \dot{h}(t) \leq g(t)$. According to Lemma 4.2 in [6] it implies that $\dot{h}^+$, the positive part of $\dot{h}$, is integrable. Since h is nonnegative, $\dot{h}^-$ is integrable too, and the conclusions follow.
- We just proved that the trajectory is bounded, thus the last part follows from Proposition 1 (iii).
- Let us finish with proving that $\tilde{w} \in L^1(0, +\infty; \mathbb{R})$: As $\tilde{w}$ is of constant sign, it suffices to show that

$$\lim_{t \to +\infty} \int_T^t \tilde{w}(s) ds$$

  exists. But

$$\int_T^t \tilde{w} = \int_T^t \ddot{h} + \lambda \int_T^t \dot{h} - \int_T^t |\dot{x}|^2.$$

The last two terms clearly converge, and the first one is $\dot{h}(t) - \dot{h}(T)$, which converges too because $\dot{h}(t) = \langle x(t) - \tilde{x}, \dot{x}(t)\rangle$ tends to 0. □

Remark 2 Under the assumption $S :=$argmin Φ nonempty we made, each trajectory of (4) satisfies assumptions of Lemma 5. It suffices indeed to pick some $\tilde{x}$ in S and take $T = 0$. This proves part (i) of Theorem 1.

5.2 Proof of (ii), (iii) and (iv)

As in the first order case, let us distinguish two situations, depending on the limit $E_\infty := \lim_{t \to +\infty} \Phi(x(t))$:

First case $E_\infty = \inf \Phi$.

We again apply the Opial's lemma.That method is quite classical since [15] to obtain weak convergence results.

We then consider the set $\mathcal{A}$ of the weak-limit points of the trajectory $x(\cdot)$. The first condition of Opial's lemma is true since S is nonempty, and so each $\tilde{x}$ in S satisfies the assumptions of Lemma 5, together with $T = 0$, so that the trajectory $x(\cdot)$ is bounded.

For the second condition of Opial's lemma, let us observe that $\Phi(x(\cdot))$ converges to E_∞ and Φ is lower semicontinuous for the weak topology (because quasiconvex and strongly l.s.c.)

Hence, for every $\tilde{x} \in \mathcal{A}$ and every real sequence $(t_n)_n$ such that $\lim_{n\to\infty} t_n = +\infty$ and $\lim_{n\to\infty} x(t_n) = \tilde{x}$, we have:

$$E_\infty = \lim_{n\to\infty} \Phi(x(t_n)) \geq \Phi(\tilde{x})$$

which asserts, as $E_\infty = \inf \Phi$, that $\mathcal{A} \subset \mathcal{S}$. From Remark 2 it follows that Assumption 2 in Opial's lemma is fulfilled. So, the trajectory $x(\cdot)$ weakly converges towards some point x_∞. As $x_\infty \in \mathcal{A} \subset S$, it implies of course from first order optimality condition that $\nabla\Phi(x_\infty) = 0$.

Second case $E_\infty > \inf \Phi$.

The Lemma 6 below applies and $x(\cdot)$ *strongly* converges to some $x_\infty \in H$. Finally, $\nabla\Phi(x_\infty) = 0$ because the function $\nabla\Phi(x(\cdot))$ strongly converges to 0 and $\nabla\Phi$ is continuous.

Lemma 6 *Let $x(\cdot)$ be a trajectory of 4. If there exists $\tilde{x} \in H$, $r > 0$ and $T \geq 0$ such that*:

$$\forall y \in \bar{B}(\tilde{x}, r), \quad \forall t \leq T, \quad \Phi(y) \geq \Phi(x(t))$$

then $\nabla\Phi(x(\cdot))$ belongs to $L^1(0, +\infty;\ H)$ and the trajectory strongly converges in H.

Proof The conditions of Lemma 3 are fulfilled with $\tilde{x}$ picked in S. So, for some reals $r > 0$ and $T > 0$ we have, for all $t \geq T$: $r|\nabla\Phi(x(t)| \leq -\dot{\tilde{w}}(t)$. According to Lemma 5, the right hand side is integrable. Thus so is $\nabla\Phi(x(\cdot))$.

Integrate the differential equation satisfied by x between 0 and t to obtain

$$x(t) = x_0 - \frac{1}{\lambda}\left(\dot{x}(t) - x_1 + \int_0^t \nabla\Phi(x(s))ds\right).$$

Since $\dot{x}(t)$ and the integral term strongly converge when $t \to +\infty$, so does $x(t)$. □

5.3 Proof of (v)

If the interior of argmin Φ is nonempty, there exist $\hat{x} \in H$ and $r > 0$ such that $B(\hat{x}, r) \subset \text{argmin}\ \Phi$. Thus

$$\forall y \in B(\hat{x}, r), \forall t \geq 0, \Phi(y) \leq \Phi(x(t))$$

and Lemma 6 leads to conclusion.

5.4 Proof of (vi)

Lemma 7 *Let* $\alpha : [0, +\infty) \to \mathbb{R}$ *be a continuous function such that*

$$\lim_{t\to+\infty} \alpha(t) = \inf_{\mathbb{R}^+} \alpha := \hat{\alpha}.$$

Then there exists a nondecreasing sequence $(t_n)_{n\in\mathbb{N}}$ *that converges to* $+\infty$ *such that*

$$\alpha(t_n) = \inf_{t\in[t_0,t_n]} \alpha(t).$$

Proof If there exists $t_n \to +\infty$ such that $\forall n, \alpha(t_n) = \hat{\alpha}$ it is immediate. Else there exists some $T \geq 0, \forall t \geq T, \alpha(t) > \hat{\alpha}$.

For each $k \in \mathbb{N}, k \geq T$, there exists $s_k \in [k, k+1]$ such that $\alpha(s_k) = \inf_{[k,k+1]} \alpha$. Choose $k_0 > T$, and if k_n is fixed, define

$$k_{n+1} = \inf\{p \geq k_n, \alpha(s_p) < \alpha(s_{k_n})\}.$$

Such k_{n+1} exists (and is greater than k_n) unless $\alpha(s_{k_n}) \leq \alpha(s_p) \leq \alpha(p)$ for all $p \geq k_n$. Letting p tend to $+\infty$, it implies that $\alpha(s_{k_n}) = \hat{\alpha}$ which contradicts the definition of T. This construction yields

$$\alpha(s_{k_{n+1}}) < \alpha(s_{k_n}) \leq \alpha(s_p), \forall k_n \leq p < k_{n+1}$$

which together with the definition of the sequence $(s_k)_{k\geq T}$ implies

$$\alpha(s_{k_{n+1}}) \leq \inf_{[s_{k_n}, s_{k_{n+1}}]} \alpha.$$

Thus the sequence $t_n = s_{k_n}$ satisfies the desired conclusion. Indeed $t_n \to +\infty$ as a subsequence of (s_k). On the other hand, $\alpha(t_n) \leq \inf_{[t_0,t_n]} \alpha$ by induction. □

We have already seen that if $E_\infty > \inf_H \Phi$ the convergence is strong, so we can assume that $E_\infty = \inf_H \Phi$. Thus $\Phi(x(t))$ satisfies condition of Lemma 7, and there exists a nondecreasing sequence t_n that converges to $+\infty$ such that

$$\forall n \in \mathbb{N}, \Phi(x(t_n)) = \inf_{t\in[t_0,t_n]} \Phi(x(t)).$$

As in [1], we define g_n by

$$g_n(t) = |x(t)|^2 - |x(t_n)|^2 - \frac{1}{2}|x(t) - x(t_n)|^2.$$

Direct computations give for any $p \geq 1$

$$\ddot{g}_{n+p}(t) + \lambda \dot{g}_{n+p}(t) = \langle -\nabla\Phi(x(t)), x(t) + x(t_{n+p})\rangle + |\dot{x}(t)|^2.$$

But the scalar product is nonpositive for $t \in [t_0, t_{n+p}]$. Indeed, as Φ is even, it is equal to $\langle \nabla\Phi(-x(t)), x(t_{n+p}) - (-x(t))\rangle$ and $\Phi(x(t_{n+p})) \le \Phi(x(t)) = \Phi(-x(t))$ on the cited interval. Solving the differential inequality yields

$$\exp(\lambda(t-t_n))\dot{g}_{n+p}(t) - \dot{g}_{n+p}(t_n) \le \int_{t_n}^{t} \exp(\lambda(s-t_n))|\dot{x}(s)|^2 ds.$$

This provides an expression for $\dot{g}_{n+p}(t)$ which we integrate on $[t_{n+1}, t_{n+p}]$. It gives

$$g_{n+p}(t_{n+p}) - g_{n+p}(t_{n+1}) \le \dot{g}_{n+p}(t_n) \int_{t_{n+1}}^{t_{n+p}} \exp(-\lambda(t-t_n))dt + \int_{t_{n+1}}^{t_{n+p}} \underbrace{\int_{t_n}^{t} \exp(\lambda(s-t))|\dot{x}(s)|^2 ds\, dt}_{q_n(t)}.$$

As $g_{n+p}(t_{n+p}) = 0$, the following estimate holds

$$-g_{n+p}(t_{n+1}) \le \frac{2}{\lambda}\dot{g}_{n+p}(t_n) + \int_{t_{n+1}}^{t_{n+p}} q_n(t)dt.$$

This means

$$\frac{1}{2}|x(t_{n+1}) - x(t_{n+p})|^2 \le |x(t_{n+1})|^2 - |x(t_{n+p})|^2 + \frac{2}{\lambda}\dot{g}_{n+p}(t_n) + \int_{t_{n+1}}^{t_{n+p}} q_n(t)dt.$$

First, 0 belongs to argmin Φ because Φ is even. So, with help of Lemma 5, we know that $t \mapsto |x(t)|^2$ has a limit as $t \to +\infty$. Then

$$\lim_{n\to\infty} \sup_{p\ge 1}(|x(t_{n+1})|^2 - |x(t_{n+p})|^2) = 0.$$

Secondly, $|\dot{g}_{n+p}(t_n)| \le 2|\dot{x}(t_n)| \sup_{t\ge 0}|x(t)|$, so that

$$\lim_{n\to\infty} \sup_{p\ge 1} \frac{2}{\lambda}\dot{g}_{n+p}(t_n) = 0.$$

Thirdly, $0 \le \int_{t_{n+1}}^{t_{n+p}} q_n(t)dt \le \frac{1}{\lambda}\int_{t_n}^{+\infty} |\dot{x}(s)|^2 ds$ and as Proposition 1 ensures that $\dot{x}(s) \in L^2(0, +\infty; H)$, it implies that

$$\lim_{n\to\infty} \sup_{p\ge 1} \int_{t_{n+1}}^{t_{n+p}} q_n(t)dt = 0.$$

Combining these informations, we obtain that $\left(x(t_n)\right)_{n\in\mathbb{N}}$ is a Cauchy sequence in H. Then x has at least one strong cluster point, which is necessarily x_∞ the weak limit of the trajectory. But

$$h : t \mapsto |x(t) - x_\infty|^2 = |x(t)|^2 + |x_\infty|^2 - 2\langle x(t), x_\infty\rangle$$

converges as $t \to +\infty$. Use the sequence (t_n) to conclude that the limit is necessarily 0, which means that x strongly converges to x_∞.

6 The proximal inertial algorithm

In this part we consider the proximal algorithm (5) which is an implicit discretization of Eq. (4). Like for the continuous method, we assume that the Assumption 7 hold. Note that the condition $\Phi \in C^{1,1}$ can be weakened to $\Phi \in C^1$. The result is then

Theorem 3 *Assume that the following hold*

- *The sequence* $(h_k)_{k\in\mathbb{N}}$ *is increasing and* $\lim_{k\to\infty} h_k = +\infty$.
- *There exists* $\alpha \in]0, 1[$ *such that for every* $k \in \mathbb{N}$, $0 \le \alpha_k \le \alpha$.

Under the condition

$$\sum_{k=1}^{\infty} \alpha_k |x_k - x_{k-1}|^2 < +\infty \tag{8}$$

we have

(i) *The sequence* $(x_k)_{k\in\mathbb{N}}$ *weakly converges to some* x_∞ *in* H.
(ii) *The limit satisfies* $\nabla\Phi(x_\infty) = 0$.
(iii) *If* x_∞ *does not belong to* $S = \text{argmin}\Phi$, *the convergence is strong.*

Remark 3 Coefficient α_k can be chosen once x_k and x_{k-1} have been computed, so that α_k can always iteratively be computed in order to have (8).

Remark 4 The class of quasiconvex functions is not closed by addition. So, in the proximal method (5) applied to the quasiconvex potential Φ, the function to be minimized may not be quasiconvex. In fact, for quasiconvex functions, operation $\vee$ (max) seems to be better adapted than operation +. This leads to the consideration of sublevels sums in place of epigraphical sums (cf [38]). The price to be paid is that regularity is lost.

Remark 5 The case with $\alpha_k = 0$ for every $k \in \mathbb{N}$ leads to the classical proximal method, that is the implicit discretization of the continuous equation (3). The convergence of that method for a quasiconvex function Φ can thus be shown with a proof of the same type. In that case the condition (8) is obviously true. As far as we know, such a result does not appear in the literature. The closest ones are obtained for an explicit discretization in [23,30] and for a proximal like method in the case of Lotka–Volterra dynamical system (cf. [8]).

7 Proof of Theorem 3

As for the continuous equation, we will separate the proof into two cases.
First case $\liminf_{k\to\infty} \Phi(x_k) = \inf_H \Phi$
We will apply the following version of Opial's lemma corresponding to the discrete case, see [34]

Lemma 8 *In a Hilbert space H, let $(x_k)_{k\in\mathbb{N}}$ be a sequence such that there exists a set $\emptyset \neq A \subset H$ satisfying*

- *if $x(k_j) \overset{w}{\rightharpoonup} \bar{x}$ for a sequence $k_j \to \infty$, then $\bar{x} \in A$,*
- *$\forall z \in A,\ \lim_{k\to\infty} |x_k - z|$ exists.*

Then there exists x_∞ in A, such that x_k weakly converges to x_∞ when $k \to \infty$.

We apply this lemma with the set $A = \operatorname{argmin} \Phi$. The first condition holds since Φ is lower semicontinuous for the weak topology. For the proof of the second condition, we use the same arguments as those in [2]. Take some $z \in \operatorname{argmin} \Phi$. The sequence $\varphi_k = \frac{1}{2}|x_k - z|^2$ converges. This proves (i). Results (ii) and (iii) follow from $x_\infty \in \operatorname{argmin} \Phi$.
Second case $\liminf_{k\to\infty} \Phi(x_k) > \inf_H \Phi$
In that case, there exists a closed ball $B(z,r)$ and a rank $K \in \mathbb{N}$ such that

$$\forall k \geq K,\ \forall y \in B(z,r),\quad \Phi(y) \leq \Phi(x_k). \tag{9}$$

Adapting the proof in [2], it shows that $\varphi_k = \frac{1}{2}|x_k - z|^2$ has a limit when $k \to \infty$. But (5) is equivalent to

$$\frac{x_{k+1} - x_k}{h_k} - \frac{x_k - x_{k-1}}{h_{k-1}} + \lambda(x_{k+1} - x_k) + h_k \nabla\Phi(x_{k+1}) = 0$$

and the sum of those equalities for $1 \leq k \leq n$ gives

$$x_{n+1} = x_1 - \frac{1}{\lambda}\left(\frac{x_{n+1} - x_n}{h_n} - \frac{x_1 - x_0}{h_0} + \sum_{k=1}^{n} h_k \nabla\Phi(x_{k+1})\right). \tag{10}$$

Since the sequence x_n is bounded and $\lim_{n\to\infty} h_n = +\infty$ by assumption, x_{n+1} converges if and only if $\sum_{k=1}^{n} h_k|\nabla\Phi(x_{k+1})|$ is finite. But for $k \geq K$ we have

$$\begin{aligned} r|\nabla\Phi(x_{k+1})| &= \left\langle \nabla\Phi(x_{k+1}), r\frac{\nabla\Phi(x_{k+1})}{|\nabla\Phi(x_{k+1})|}\right\rangle \\ &= \langle\nabla\Phi(x_{k+1}), y - x_{k+1}\rangle - \langle\nabla\Phi(x_{k+1}), z - x_{k+1}\rangle \\ &\leq -\langle\nabla\Phi(x_{k+1}), z - x_{k+1}\rangle \end{aligned}$$

where $y = z + r\frac{\nabla\Phi(x_{k+1})}{|\nabla\Phi(x_{k+1})|}$ belongs to $B(z, r)$. Then

$$\sum_{k=K}^{n} h_k|\nabla\Phi(x_{k+1})| \leq \frac{-1}{r}\sum_{k=K}^{n}\frac{h_k}{\beta_k}\langle\beta_k\nabla\Phi(x_{k+1}), z - x_{k+1}\rangle. \tag{11}$$

But on one hand $\frac{h_k}{\beta_k} \leq \lambda^{-1}$ and on the other hand

$$\begin{aligned} -\langle\beta_k\nabla\Phi(x_{k+1}), z - x_{k+1}\rangle &= -\langle x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}), x_{k+1} - z\rangle \\ &= -\theta_{k+1} + \alpha_k\theta_k - \frac{1}{2}|v_{k+1}|^2 + \frac{1}{2}(\alpha_k + \alpha_k^2)|x_k - x_{k-1}|^2 \\ &\leq -\theta_{k+1} + \alpha_k\theta_k - \frac{1}{2}|v_{k+1}|^2 + \delta_k \end{aligned}$$

The notations and the arguments are the same as in [2]: $\theta_k = \varphi_k - \varphi_{k-1}$, $v_{k+1} = x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})$ and $\delta_k = \alpha_k|x_k - x_{k-1}|^2$. The series in (11) is then convergent since

- $\sum_{k=1}^{\infty}\delta_k < +\infty$ by Assumption 8,
- $\sum_{k=1}^{N}\theta_k = \varphi_N - \varphi_0$ converges, thus $\sum_{k=1}^{\infty}|v_{k+1}|^2 < +\infty$, see [2],
- $\alpha_k|\theta_k| \leq \alpha(2[\theta_k]_+ - \theta_k)$, where $[\theta]_+ = \max\{\theta, 0\}$, and it is also shown in [2] that the series $\sum_{k=1}^{\infty}[\theta_k]_+$ is convergent. Then $\sum_{k=1}^{\infty}\alpha_k|\theta_k| < +\infty$.

This achieves to prove both (i) and (iii). The point (ii) follows from the convergence of the series $\sum_{k=K}^{n} h_k|\nabla\Phi(x_{k+1})|$ and the continuity of $\nabla\Phi$.

Acknowledgements The authors thank the anonymous referees for their helpful comments.

Appendix: Genericity result for a Morse function in finite dimension case

We now consider the case where $H = \mathbb{R}^N$ for some $N \geq 1$ and $\Phi : \mathbb{R}^N \to \mathbb{R}$ is a coercive C^2 Morse function. So, for each critical point $\hat{x}$ of Φ, the hessian $H_\Phi(\hat{x})$ of Φ at $\hat{x}$ is nonsingular, and each critical points of Φ is isolated, so that denoting $\mathcal{C}_\Phi$ the set of these points, there exists some $\emptyset \neq\subset I \subset \mathbb{N}$ such that $\mathcal{C}_\Phi = \{\hat{x}_k,\ \ k \in I\}$.

Our purpose is to prove the following:

Theorem 4 *Under the former assumptions, the heavy ball system (4) converges towards some local minimum of Φ, generically with respect to the initial conditions $(x_0, \dot{x}_0)$.*

Let us notice that such genericity result has been established fort more general systems, even in infinite dimensional spaces (cf. [16]). Our purpose here is only to propose a self contained proof of Theorem 4.

In order to prove this theorem, let us first recall the stable manifold Theorem (cf. [35, p. 223]): Let $F : \mathbb{R}^p \to \mathbb{R}^p$ be a C map and consider the dynamical system:

$$\dot{z}(t) = F(z(t)) \tag{12}$$

Let us denote $\psi(z_0, t)$ the value at t of the solution of (12) with initial condition z_0.

Assume that $\hat{z}$ is a hyperbolic equilibrium point of F, meaning that $F(\hat{z}) = 0$ and that no (complex) eigenvalue of $\nabla F(\hat{z})$ has zero real part. Consider the invariant set:

$$W^s(\hat{z}) = \left\{ y_0 \in \mathbb{R}^N \;/\; \lim_{t\to+\infty} \psi(z_0, t) = \hat{z} \right\}$$

The global stable manifold theorem claims that $W^s(\hat{z})$ is an immersed submanifold of $\mathbb{R}^p$, whose dimension equals the number of (complex) eigenvalues of $\nabla F(\hat{z})$ with negative real parts.

The dynamical system (4) can classically be written under the form (12) in which, for all $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$, $F(x, y) = (y, -\lambda y - \nabla\Phi(x))$. It is easily seen that the set $\mathcal{C}_F$ of the critical points of F is:

$$\mathcal{C}_F = \left\{ (\hat{x}, 0) \in \mathbb{R}^N \times \mathbb{R}^N \text{ such that } \hat{x} \in \mathcal{C}_\Phi \right\}.$$

If $\hat{z} = (\hat{x}, 0) \in \mathcal{C}_F$, an easy calculation shows that $\nu \in \mathbb{C}$ is a proper value of $\nabla F(\hat{z})$ if and only if $-\nu(\lambda + \nu)$ is a proper value of $H_\Phi(\hat{x})$, i.e., if and only if there exists some $\alpha \in H_\Phi(\hat{x})$ such that $\nu^2 + \lambda\nu - \alpha = 0$. Since $H_\Phi(\hat{x})$ is nonsingular, no proper value of $\nabla F(\hat{z})$) as zero real part, and then:

Lemma 9 *Each critical points of F is hyperbolic.*

From [6], we know that, in the coercice C^2 Morse function case, each trajectory $x(\cdot)$ of (4) converges towards some singular points x_∞ of $\mathcal{C}_\Phi$, with $\dot{x}(t) \to 0$. So, the following partition holds:

$$\mathbb{R}^N \times \mathbb{R}^N = \cup_{k\in I} W^s(\hat{z}_k) \tag{13}$$

Let us define

$$I^- = \{k \in I \text{ such that each proper value of } \nabla F(\hat{z}_k) \text{ has a negative real part}\}.$$

Denoting by J the complementary subset of I^- in I, the stable manifold theorem ensures that for all $k \in I$, the subset $W^s(\hat{z}_k)$ is an immersed manifold in $\mathbb{R}^N \times \mathbb{R}^N$, whose dimension is $2N$ when $k \in I^-$ and at most $2N-1$ when $k \in J$. Hence, $W^s(\hat{z}_k)$ is an open subset of $\mathbb{R}^N \times \mathbb{R}^N$ for $k \in I^-$ and has Lebesgue measure zero for $k \in J$.

It follows that $\cup_{k\in I^-} W^s(\hat{z}_k)$ is an open subset of $\mathbb{R}^N \times \mathbb{R}^N$, and that $\cup_{k\in J} W^s(\hat{z}_k)$ has Lebesgue measure zero, and hence en empty interior.

Moreover, let us consider some critical point $\hat{z}_k$ of F for $k \in I^-$. We have $\hat{z}_k = (\hat{x}_k, 0)$, with $\hat{x}_k$ is a critical point of Φ. We claim that $H_\Phi(\hat{x}_k)$ has only negative proper value: By contradiction, let α be some nonnegative proper value of $H_\Phi(\hat{x}_k)$. As Φ is a Morse function, we then have $\alpha > 0$. Moreover, each solution μ of $\mu + \lambda\mu - \alpha$ is a proper value of $\nabla F(\hat{z}_k)$. But one of these solution is $\frac{-\lambda+\sqrt{\lambda^2+4\alpha}}{2}$, which is positive, contradicting the assumption $k \in I^-$. Then, for all $k \in I^-$, the critical point $\hat{x}_k$ is a local minimum of Φ.

In conclusion, the set of initial conditions $z_0 = (x_0, \dot{x}_0)$ such that $\psi(z_0, t)$ converges to a local minimum of Φ is the subset $\cup_{k\in I^-} W^s(\hat{z}_k)$, which is open, ant the set of initial conditions such that $\psi(z_0, t)$ converges towards a nonlocal minimum point of Φ is $\cup_{k\in J} W^s(\hat{z}_k)$, which has Lebesgue measure zero, and so, has an empty interior. Together with (13), this result achieves the proof of Theorem 4.

References

1. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control Optim. **38**(4), 1102–1119 (2000)
2. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via dicretization of a nonlinear oscillator with damping. Wellposedness in optimization and related topics. Set Valued Anal. **9**(1–2), 3–11 (2001)
3. Arrow, K., Debreu, G.: Existence of an equilibrium for a competitive economy. Econometrica **22**, 265–290 (1954)
4. Attouch, H., Cabot, A., Redont, P.: The dynamics of elastic shocks via epigraphical regularization of a differential inclusion. Barrier and penalty approximations. Adv. Math. Sci. Appl. **12**(1), 273–306 (2002)
5. Attouch, H., Czarnecki, M.: Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria. J. Differ. Equ. **179**, 278–310 (2002)
6. Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of real-valued function by asymptotic analysis of a dissipative dynamical system. Commun. Contemp. Math. **2**(1), 1–34 (2000)
7. Attouch, H., Soubeyran, A.: Inertia and reactivity in decision making as cognitive variational inequalities. J. Convex Anal. **13**(2), 207–224 (2006, to appear)
8. Attouch, H., Teboule, M.: Regularized Lotka–Volterra dynamical system as continuous proximal-like method in optimization. J. Optim. Theory Appl. **121**(3), 77–106 (2004)
9. Aussel, D.: Subdifferential properties of quasiconvex and pseudoconvex function, unified approach. J. Optim. Theory Appl. **97**, 29–45 (1998)
10. Aussel, D.: Contributions en analyse multivoque et en optimisation, thèse hdr. Ph.D. thesis, Université Montpellier 2 (2005)
11. Aussel, D., Daniilidis, A.: Normal characterization of the main classes of quasiconvex analysis. Set Valued Anal. **8**, 219–236 (2000)
12. Baillon, J.: Un exemple concernant le comportement asymptotique de la solution du problème $du/dt + \partial\varphi(u) \ni 0$. J. Funct. Anal. **28**, 369–376 (1978)
13. Barron, N., Liu, W.: Calculus of variation in L^∞. Appl. Math. Optim. **35**, 237–263 (1997)
14. Brézis, H.: Opérateurs maximaux monotones et semi-groupes de contraction. Lectures Notes No. 5, North Holland (1973)
15. Bruck, R.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. J. Funct. Anal. **18**, 15–26 (1975)
16. Brunovsky, P., Polàcik, P.: The morse-smale structure of generic reaction-diffusion equation in higher space dimension. J. Differ. Equ. **135**(1), 129–181 (1997)

17. Crouzeix, J.: A review of continuity and differentiability properties of quasiconvex functions on $\mathbb{R}^n$. Research Notes in Mathematics 57, Aubin and Vinter Ed, pp. 18–34 (1982)
18. Crouzeix, J.P.: Contribution à l'étude des fonctions quasiconvexes. Ph.D. Thesis, Université de Clermont-Ferrand Thèse d'Etat (1977)
19. Crouzeix, J.-P.: La convexité généralisée en économie mathématique. In: Penot, J. (ed.) Proceedings of 2003 MODE-SMAI Conference, vol. 13, pp. 31–40 (2003)
20. Dacorogna, B.: Direct methods in the calculus of variation. Applied Mathematical Sciences, vol. 78. Springer, Heidelberg (1989)
21. Debreu, G.: Theory of Value. Wiley, New York (1959)
22. Debreu, G.: Mathematical Economics: Twenty papers. Cambridge University Press, Cambridge (1983)
23. Dussault, J.P.: Convergence of implementable descent algorithms for unconstrained optimization. J. Optim. Theory Appl. **104**(3), 739–745 (2000)
24. Ginsberg, W.: Concavity and quasiconcavity in economics. J. Econ. Theory **6**, 596–605 (1973)
25. Goudou, X., Munier, J.: Asymptotic behavior of solutions of a gradient-like integrodifferential Volterra inclusion. Adv. Math. Sci. Appl. **15**, 509–525 (2005)
26. Guler, O.: On the convergence of the proximal point algorithm for convex optimization. SIAM J. Control Appl. **29**(2), 403–419 (1991)
27. Haraux, A.: Systèmes dynamiques dissipatifs et applications. Masson, Paris (1990)
28. Haraux, A., Jendoubi, M.: Convergence of solutions of second-order gradient-like systems with analytic nonlinearities. J. Differ. Equ. **144**(2), 313–320 (1998)
29. Jendoubi, M., Polacik, P.: Non-stabilizing solutions of semilinear hyperbolic and elliptic equations with damping. In: Proceedings Section A: Mathematics, Royal Society of Edinburgh, pp. 1137–1153(17) (2003). http://www.ingentaconnect.com/content/rse/proca/2003/00000133/00000005/art00007
30. Kiwiel, K., Murty, K.: Convergence of the steepest descent method for minimizing quasiconvex functions. J. Optim. Theory Appl. **89**(1), 221–226 (1996)
31. Lemaire, B.: New methods in optimization and their industrial uses. In: Penot, J. (ed.) Inter. Series of Numerical Math., vol. 87. Birkhauser-Verlag, Basel, pp. 73–87. Symp. Pau and Paris (1989)
32. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Française d'Inform. Recherche Opérationnelle, Série R-3 **4**, 154–158 (1970)
33. Martinez-Legaz, J.E.: Generalized convex duality and its economic applications, handbook of generalized convexity and generalized monotonicity. Nonconvex Optim. Appl. **76**, 237–292 (2005)
34. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. **73**, 591–597 (1967)
35. Perko, L.: Differential equations and dynamical systems. Springer, Heidelberg (1996)
36. Polyack, B.: Some methods of speeding up the convergence of iterative methods. Z. Vylist Math. Fiz. **4**, 1–17 (1964)
37. Rockafellar, R.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
38. Seeger, A., Volle, M.: On a convolution operation obtained by adding level sets: classical and new results. Oper. Res. **29**(2), 131–154 (1995)
39. Takayama, A.: Mathematical Economics, 2nd edn. Cambridge University Press, Cambridge (1995)