The Second-order in Time Continuous Newton Method

H. Attouch and P. Redont

ACSIOM-CNRS EP 2066, Département de Mathématiques, case 51, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France, attouch@math.univ-montp2.fr, redont@math.univ-montp2.fr.

Abstract. Let H be a real Hilbert space and $\Phi : H \mapsto \mathbb{R}$ a twice continuously differentiable function, whose Hessian is Lipschitz continuous on bounded sets. We study the Newton-like second-order in time nonlinear dissipative dynamical system: $\ddot{x}(t) + \nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0$, plus Cauchy data, mainly in view of the unconstrained minimization of the function Φ. The main result is the gradient vanishing along any bounded trajectory as time goes to infinity. Results concerning the convergence of every bounded solution to a critical point are given in peculiar situations: when Φ is convex (with only one minimum) or is a Morse function.

Keywords: dissipative dynamical system, optimization, local minima, convex minimization, asymptotic behaviour, Newton method.

AMS classification: 34A12, 34Dxx, 49Mxx.

1 Introduction

When dealing numerically with the minimization of a function $\Phi : H \to \mathbb{R}$, or more generally with the calculation of the critical points of Φ, one usually uses some process generating a sequence (x_i) with properties like: $\lim_{t\to\infty} \nabla\Phi(x_i) = 0$ or, still better, $x_i \to \bar{x}, i \to \infty$ where $\bar{x}$ is a critical point of Φ. If the discrete dependence of the sequence (x_i) on step i can, at least formally, be turned into the continuous dependence on some parameter t, interpreted as the time, then the discrete process may become a continuous dynamical system with trajectories $t \to x(t)$; and the question now is the asymptotic behaviour of $\nabla\Phi(x(t))$ or $x(t)$ in relation with the critical points of Φ.

This passage from the discrete to the continuous is best illustrated by the steepest descent method, also known as the gradient method:

$$x_{i+1} - x_i + h\nabla\Phi(x_i) = 0, \ h > 0, \ x_0 \text{ given in } H,$$

whose continuous version is:

$$\dot{x}(t) + \nabla\Phi(x(t)) = 0, \ x(0) = x_0 \text{ given in } H. \tag{1}$$

A lot of work has been devoted to the continuous gradient equation, to quote a few: Attouch-Cominetti [3], Baillon [6], Brézis [8], Bruck [9], Haraux [10], Łojasiewicz [12,13], Palis-de Melo [14].

Following the same idea, Newton's method:

$$\nabla^2\Phi(x_i)(x_{i+1}-x_i)+\nabla\Phi(x_i)=0,\ x_0 \text{ given in } H,$$

can easily be transformed into a continuous system:

$$\nabla^2\Phi(x(t))\dot{x}(t)+\nabla\Phi(x(t))=0,\ x(0)=x_0 \text{ given in } H. \tag{2}$$

Aubin-Cellina [5], and Alvarez-Perez [2] have already studied this system; however, satisfying convergence results are impeded, as for the discrete version, by the possible ill-conditioning of the Hessian $\nabla^2\Phi$.

To cope with that problem, one is tempted to introduce a perturbation to the system, acting as a regularization in fact, and write the second-order in time continuous Newton method:

$$\ddot{x}(t)+\nabla^2\Phi(x(t))\dot{x}(t)+\nabla\Phi(x(t))=0,\ (x(0),\dot{x}(0))=(x_0,\dot{x}_0) \text{ given in } H. \tag{3}$$

If we keep in mind that our first objective is to get at the minima, or the critical points of Φ, introducing a second-order term may impulse dynamics to the solutions of (3) and confer exploration properties on them. That is what we already observed in Attouch-Goudou-Redont [4] when passing from the continuous gradient system (1) to the HBF (Heavy Ball with Friction) system:

$$\ddot{x}(t)+\lambda\dot{x}(t)+\nabla\Phi(x(t))=0,\ \lambda>0,\ (x(0),\dot{x}(0))=(x_0,\dot{x}_0) \text{ given in } H. \tag{4}$$

We also remarked that choosing the right friction coefficient $\lambda > 0$ is no easy task, all the more because it ought to depend on point $x(t)$ at least. Ideally it should first prevent the trajectory from zigzagging, that is damp the rapidly varying components of x (see Alvarez [1] for an illustration). Certainly, this cannot be achieved without taking into account some second order information about Φ: enter the Hessian. This is another reason for considering equation (3).

Thus our hope is to get from one critical point to another by following different trajectories of a dynamical system, an idea common to various optimization methods (ascent-descent methods for example, see Jongen-Ruiz Jhones [11]).

2 Global Existence

Let H be a real Hilbert space. Let us consider a mapping $\Phi : H \to \mathbb{R}$ which satisfies the following conditions :

$$(\mathcal{H})\ \begin{cases} \Phi \text{ is twice continuously differentiable on } H \\ \Phi \text{ is bounded from below on } H \\ \nabla^2\Phi \text{ is Lipschitz continuous on the bounded subsets of } H. \end{cases}$$

The second order system in H :

$$\ddot{x} + \nabla^2 \Phi(x)\dot{x} + \nabla \Phi(x) = 0 \tag{5}$$

can be written as a first order system in $H \times H$:

$$\dot{Y} = F(Y)$$

with

$$Y(t) = \begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix} \quad \text{and} \quad F(u,v) = \begin{pmatrix} v \\ -\nabla^2\Phi(u)v - \nabla\Phi(u) \end{pmatrix}. \tag{6}$$

For $Y_0 = \begin{pmatrix} x_0 \\ \dot{x}_0 \end{pmatrix}$ given in $H \times H$, the Cauchy-Lipschitz theorem and the hypothesis $(\mathcal{H})$ ensure the existence of a unique local solution to the problem:

$$\begin{cases} \dot{Y} = F(Y) \\ Y(0) = Y_0 \end{cases} \tag{7}$$

On the other hand, we can define along every trajectory of (5) an energy by:

$$E(t) = \frac{1}{2}|\dot{x}(t) + \nabla\Phi(x(t))|^2 + \Phi(x(t)).$$

The central result of this section is given by the following theorem.

Theorem 2.1 *Let us assume that $\Phi : H \to \mathbb{R}$ satisfies the assumptions $(\mathcal{H})$. Then, the following properties hold:*

(i) *For all $(x_0, \dot{x}_0)$ in $H \times H$, there exists a unique solution $x(t)$ of (5) defined on the whole interval $[0, +\infty[$, which is of class $\mathcal{C}^2$ on $[0, +\infty[$, and which satisfies the initial conditions $x(0) = x_0$ and $\dot{x}(0) = \dot{x}_0$.*

(ii) *For every trajectory $x(t)$ of (5), the energy $E(t)$ is decreasing on $[0, +\infty[$, bounded from below and hence converges to some real value E_∞. Moreover,*
- $\dot{x} + \nabla\Phi(x) \in L^\infty(0, +\infty; H)$,
- $\nabla\Phi(x) \in L^2(0, +\infty; H)$.

(iii) *Assuming moreover that x is in $L^\infty(0, +\infty; H)$, then we have*
- *$\nabla\Phi(x)$ and $\nabla^2\Phi(x)$ are bounded,*
- *$\dot{x}$ and $\ddot{x}$ belong to $L^\infty(0, +\infty; H)$,*
- $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$.
- $E_\infty = \lim_{t\to+\infty} \frac{1}{2}|\dot{x}(t)|^2 + \Phi(x(t))$.

Proof. i) For any choice of initial conditions $(x_0, \dot{x}_0) \in H \times H$, the existence and uniqueness of a local solution for (5), follows from the Cauchy-Lipschitz theorem. Let x denote the corresponding maximal solution, which is defined on some interval $[0, T_{max}[$ with $0 < T_{max} \leq +\infty$. In order to prove that $T_{max} = +\infty$, let us show that $\dot{x}$ and $\ddot{x}$ are bounded.

We first observe that equation (5) and the regularity assumptions on Φ automatically imply that $x(.)$ is $\mathcal{C}^2$ on $[0, T_{max}[$. By differentiation of $E(t)$, and using (5), we obtain:

$$\begin{aligned} \dot{E}(t) &= \langle \dot{x}(t) + \nabla\Phi(x(t)), \ddot{x}(t) + \nabla^2\Phi(x(t))\dot{x}(t)\rangle + \langle \dot{x}(t), \nabla\Phi(x(t))\rangle \\ &= -|\nabla\Phi(x(t))|^2. \end{aligned} \tag{8}$$

Thus, the function $E(.)$ is decreasing and for all $t \in [0, T_{max}[$:

$$E(t) \leq E(0).$$

Equivalently,

$$\frac{1}{2}|\dot{x}(t) + \nabla\Phi(x(t))|^2 + \Phi(x(t)) \leq \frac{1}{2}|\dot{x}_0 + \nabla\Phi(x_0))|^2 + \Phi(x_0). \tag{9}$$

Since Φ is bounded from below, we obtain that $\dot{x}(t) + \nabla\Phi(x(t))$ is bounded on $[0, T_{max}[$.

Let us turn to equation (5), which we write:

$$\ddot{x} + \nabla^2\Phi(x)\dot{x} + \dot{x} + \nabla\Phi(x) = \dot{x},$$

and let us integrate it on $[0, T]$ for any $T \in [0, T_{max}[$:

$$\dot{x}(T) + \nabla\Phi(x(T)) + \int_0^T (\dot{x}(t) + \nabla\Phi(x(t)))dt - \dot{x}_0 - \nabla\Phi(x_0) = x(T) - x_0. \tag{10}$$

Let us now argue by contradiction, and assume that $T_{max} < +\infty$. Since $\dot{x} + \nabla\Phi(x)$ is bounded on $[0, T_{max}[$, so is x according to equation (10). Thus, after our hypothesis $(\mathcal{H})$, $\nabla^2\Phi(x)$ and therefore $\nabla\Phi(x)$ are bounded on $[0, T_{max}[$. Since $\dot{x} + \nabla\Phi(x)$ is bounded, we conclude that $\dot{x}$ is bounded on $[0, T_{max}[$. Turning once more to equation (5) shows that $\ddot{x}$ is bounded on $[0, T_{max}[$ too.

It is now a standard argument to derive from the boundedness of $\dot{x}$ and $\ddot{x}$ that $T_{max} = +\infty$. Indeed we have for some constant C:

$$\forall (t, t') \in [0, T_{max}[^2, \ \ |x(t) - x(t')| \leq C|t - t'|,$$

and since $T_{max} < +\infty$, $\lim_{t \to T_{max}} x(t) := x_\infty$ exists. The same argument applies to $\dot{x}$ and shows that $\lim_{t \to T_{max}} \dot{x}(t) := \dot{x}_\infty$ exists. But, applying again the local existence theorem with initial data $(x_\infty, \dot{x}_\infty)$, we can extend the maximal solution to a strictly larger interval, which is a clear contradiction. So, $T_{max} = +\infty$, which completes the proof of i).

ii) We already proved that $E(.)$ is decreasing. Since Φ is bounded from below, and since $E(t) \geq \Phi(x(t))$, we have that $E(.)$ is also bounded from below. As a consequence, $\lim_{t\to+\infty} E(t) = E_\infty$ exists, with $E_\infty \in \mathbb{R}$. Using (8), and the fact that Φ is bounded from below, we obtain that, for all $t \geq 0$

$$\frac{1}{2}|\dot{x}(t) + \nabla\Phi(x(t))|^2 \leq \frac{1}{2}|\dot{x}_0 + \nabla\Phi(x_0)|^2 + \Phi(x_0) - \inf \Phi.$$

Hence,

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

From (8), we derive that, for all $0 \leq t < +\infty$

$$\int_0^t |\nabla\Phi(x(s))|^2 ds = E_0 - E(t).$$

Since $E(t)$ decreases to E_∞ as t increases to $+\infty$, we obtain that

$$\int_0^{+\infty} |\nabla\Phi(x(s))|^2 ds = E_0 - E_\infty,$$

and $\nabla\Phi(x) \in L^2(0, +\infty; H)$.

iii) We now assume that x is in $L^\infty(0, +\infty; H)$. Then, owing to our hypothesis $(\mathcal{H})$, $\nabla^2\Phi(x)$ and $\nabla\Phi(x)$ belong to $L^\infty(0, +\infty; H)$.

Since $\dot{x} + \nabla\Phi(x)$ is in $L^\infty(0, +\infty; H)$, $\dot{x}$ belongs to $L^\infty(0, +\infty; H)$, and equation (5) shows that $\ddot{x}$ belongs to $L^\infty(0, +\infty; H)$ too.

Let us now observe that the function $h(t) := \frac{1}{2}|\nabla\Phi(x(t))|^2$ satisfies both:

$$h \in L^1(0, +\infty; \mathbb{R}) \quad \text{and} \quad \dot{h} = < \nabla^2\Phi(x)\dot{x}, \nabla\Phi(x) > \in L^\infty(0, +\infty; \mathbb{R}).$$

According to a classical result, these two properties imply: $\lim_{t\to+\infty} h(t) = 0$. (Indeed, arguing by contradiction and owing to h being Lipschitzian, there would exist $\varepsilon > 0$, $\eta > 0$ and a sequence of non-overlapping intervals $[t_n - \eta, t_n + \eta] \subseteq [0, +\infty[$ such that $|t - t_n| < \eta \Rightarrow h(t) > \varepsilon$; which is inconsistent with $h \in L^1(0, +\infty; H)$). Therefore, in our situation, we have $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$.

The last result, $E_\infty = \lim_{t\to+\infty} F(t)$ with $F(t) = \frac{1}{2}|\dot{x}(t)|^2 + \Phi(x(t))$, is a consequence of $\dot{x}$ being bounded and of the convergence of $\nabla\Phi(x(t))$ to 0. Note that F is the energy functional of the HBF equation ([4]). $\triangle$

Corollary 2.1 *Assume that $\Phi : H \to \mathbb{R}$ satisfies the assumptions $(\mathcal{H})$ and is coercive, i.e.* $\lim_{|x|\to+\infty} \Phi(x) = +\infty$, *then x is in $L^\infty(0, +\infty; H)$ and the conclusions of theorem 3.1 hold.*

Proof. It is enough to observe that the inequality (9) gives

$$\Phi(x(t)) \leq \Phi(x_0) + \frac{1}{2}|\dot{x}_0 + \nabla\Phi(x_0)|^2.$$

This majorization on $\Phi(x(t))$ and the coerciveness of Φ imply that the trajectory $x(.)$ remains bounded, i.e. $x \in L^\infty(0, +\infty; H)$. $\triangle$

3 Convergence of the Trajectories

3.1 Morse functions

We first recall the notion of ω-limit set related to the asymptotic behaviour of a trajectory.
For a given initial condition $y_0 = (x_0, \dot{x}_0) \in H \times H$, let $x_{y_0}(.)$ denote the unique maximal solution of (3): $\ddot{x}(t) + \nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0$ with initial data y_0; we will use $x(.)$ when there is no ambiguity on y_0. The ω-limit set ω_{y_0} of the trajectory x_{y_0} is defined by

$$\omega_{y_0} = \bigcap_{t>0} \overline{x_{y_0}([t, +\infty[)}.$$

The set ω_{y_0} can also be obtained as the set of the limit points of $x_{y_0}(.)$ as $t \to +\infty$

$$\omega_{y_0} = \{\xi \in H : \exists (t_n)_{n\in\mathbb{N}}, \quad t_n \to +\infty \quad \text{and } x(t_n) \overset{n\to+\infty}{\longrightarrow} \xi\}.$$

The set of the critical points of Φ is denoted by S

$$S = \{x \in H : \nabla\Phi(x) = 0\}.$$

In order to obtain convergence of the trajectories we need to make further assumptions on Φ and on the trajectories themselves (precompactness).

We recall that $\Phi : H \to \mathbb{R}$ is a Morse function if $\Phi \in C^2$ and its Hessian $\nabla^2\Phi(\bar{x})$ possesses a continuous inverse at every critical point $\bar{x}$. It is a trivial result that all the critical points of a Morse function are isolated. We can now state:

Theorem 3.1 *Let H be a Hilbert space, and $\Phi : H \to \mathbb{R}$ a Morse function, with $\nabla^2\Phi$ Lipschitz continuous on bounded sets. For $y_0 = (x_0, \dot{x}_0) \in H \times H$, let x_{y_0} be the solution of (3):*

$$\begin{cases} \ddot{x}(t) + \nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0 \\ x(0) = x_0, \dot{x}(0) = \dot{x}_0 \end{cases}$$

For any y_0 such that the trajectory x_{y_0} is precompact for the topology of the norm in H, then $x_{y_0}(t)$ converges as t goes to infinity to a critical point of Φ.

Proof. The set ω_{y_0} is non-void connected compact as the decreasing intersection of non-void connected compact sets. In view of $\lim_{t\to+\infty} \nabla\Phi(x_{y_0}(t)) = 0$ and of the continuity of $\nabla\Phi(x_{y_0})$, every point in ω_{y_0} is a critical point of Φ. By assumption, Φ is a Morse function, and all the elements of S are isolated. So ω_{y_0} is a connected set contained in a set whose elements are all isolated. This implies that ω_{y_0} is reduced to a singleton, $\omega_{y_0} = \{\bar{x}\}$. The trajectory x_{y_0} which is contained in a compact set, and which has a unique limit point necessarily converges to this unique element $\bar{x} \in S$. $\triangle$

3.2 Convex functions

In the case where Φ is convex, the following corollary adds a little to the main theorem.

Corollary 3.1 *Assume that $\Phi : H \to \mathbb{R}$ satisfies the assumptions $(\mathcal{H})$ and is convex. Then for every bounded trajectory x, we have:*

- $\lim_{t\to+\infty} \Phi(x(t)) = \inf \Phi$,
- *every weak cluster point of x is a minimum of Φ,*
- $\lim_{t\to+\infty} |\dot{x}(t)|$ *exists.*

Proof. The first two points are merely an application of the gradient inequality for a convex function.

Let y be an arbitrary point in H; we have:

$$\Phi(y) \geq \Phi(x(t)) + < \nabla\Phi(x(t)), y - x(t) > .$$

Since $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$, after the theorem (2.1), and since x is bounded, passing to the limit and using the lower semicontinuity of Φ, we get:

$$\Phi(y) \geq \limsup_{t\to+\infty} \Phi(x(t)) \geq \liminf_{t\to+\infty} \Phi(x(t)) \geq \inf \Phi.$$

Hence:

$$\inf \Phi = \lim_{t\to+\infty} \Phi(x(t)).$$

Now, let $\bar{x}$ be a weak cluster point of Φ. For some sequence t_n increasing towards $+\infty$ we have:

$$\Phi(\bar{x}) \leq \liminf_{n\to+\infty} \Phi(x(t_n)) = \inf \Phi.$$

Hence, $\bar{x}$ is a minimum of Φ.

To prove the existence of $\lim_{t\to+\infty} |\dot{x}(t)|$, note that $\dot{x}$ is bounded and that the following limits exist: $\lim_{t\to+\infty} E(t) = \lim_{t\to+\infty} \frac{1}{2}|\dot{x}(t) + \nabla\Phi(x(t))|^2 + \Phi(x(t)) = E_\infty$, $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$, $\lim_{t\to+\infty} \Phi(x(t)) = \inf \Phi$. $\triangle$

It is remarkable that the second-order Newton equation enjoys another Liapunov functional:

$$F(t) = \frac{1}{2}|\dot{x}(t)|^2 + \Phi(x(t))$$

with decreasing rate:

$$\dot{F}(t) = - < \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) >,$$

a non-positive number since Φ is convex. Nevertheless, only under stringent hypotheses are we able to specify the asymptotic behaviour of the trajectories.

Corollary 3.2 *Assume that $\Phi : H \to \mathbb{R}$ is convex and has a third derivative which is bounded on bounded sets. Then for every bounded trajectory x:* $< \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) > \to 0, t \to \infty$.

Proof. Note first that the hypotheses of theorem (2.1) are verified indeed.

Define $h(t) = < \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) >$, and remember that h is non-negative. After the energy equation for F:

$$F(t) + \int_0^t < \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) > dt = F(0),$$

h is in $L^1([0,\infty[, H)$. Further: $\dot{h} = < \dot{x}, \nabla^3\Phi(x), \dot{x} > +2 < \nabla^2\Phi(x)\ddot{x}, \dot{x} >$ is bounded; hence $< \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) > \to 0, t \to \infty$. △

Corollary 3.3 *Assume that $\Phi : H \to \mathbb{R}$ satisfies the assumptions $(\mathcal{H})$, is convex and has only one minimum point $\bar{x}$. Then for any bounded trajectory we have:*

$$x(t) \to \bar{x} \text{ weakly in } H, t \to \infty,$$

$$\dot{x}(t) \to 0 \text{ weakly in } H, t \to \infty.$$

Proof. The trajectory x does have weak cluster-points which are minima of Φ; hence: $x(t) \to \bar{x}$ weakly in H, $t \to \infty$.

Now, for any fixed $a \in H$, define $h = < x, a >$; the functions $\dot{h} = < \dot{x}, a >$ and $\ddot{h} = < \ddot{x}, a >$ are bounded. Suppose that $\dot{h}$ does not tend to 0 as t goes to ∞. Then, owing to $\dot{h}$ being Lipschitzian, there exist some $\eta > 0$, some $\delta > 0$ and an infinite family of non-overlapping intervals $[t_i - \delta, t_i + \delta]$ such that: $|t - t_i| < \delta \Rightarrow |\dot{h}(t)| > \eta$; and we may even suppose $\dot{h}(t) > \eta$, in which case we have: $h(t_i + \delta) - h(t_i - \delta) > 2\delta\eta$. But this is inconsistent with the Cauchy property that h has to comply with since $\lim_{t\to\infty} h(t)$ exists. Hence: $\dot{x}(t) \to 0$ weakly in H, $t \to \infty$. △

The following result is the strong version of the preceding corollary.

Corollary 3.4 *Assume that $\Phi : H \to \mathbb{R}$ satisfies the assumptions $(\mathcal{H})$ and is strongly convex, that is:*

$$\exists k > 0 \;/\; \forall (u,v) \in H^2,\; \langle \nabla\Phi(v) - \nabla\Phi(u), v - u \rangle \;\geq\; k|v - u|^2.$$

Then for any bounded trajectory we have:

$$x(t) \to \bar{x} \text{ strongly in } H, t \to \infty,$$

$$\dot{x}(t) \to 0 \text{ strongly in } H, t \to \infty,$$

where $\bar{x}$ is the only minimum point of Φ.

Proof. The strong convexit

$k|x(t)$

Hence the strong converge

Now the consideration

$L^1([0,\infty[, H)$ and further

$\int_0^\infty < \nabla^2\Phi$

Since $|\dot{x}|^2$ is in $L^1([0,\infty[$

$|\dot{x}(t)|^2$ tends to 0 as t t

towards 0 as t tends to o

Observe that the co

convex case, under the h

function their convergenc

4 A Few Remark

In order to make the foll

about Newton's continu

Alvarez-Perez' [2].

With some adaptati

states:

Theorem 4.1 *Let Φ be*

to $\mathbb{R}$; let $\nabla\Phi$ be bounde

(*) $\exists c > 0$ such tha

$|u| \leq c|\nabla\Phi(x)|$.

Then the implicit differ

$\nabla^2\Phi(x(t))\dot{x}(t) + \nabla$

has for all $x_0 \in H$ traj

Furthermore, for all t

(i) $\nabla\Phi(x(t)) = e^{-t}$

(ii) $dist(x(t), \{u \in$

Every cluster point of

Note that a traject

Now in [2], corollar

Theorem 4.2 *Let Φ*

exists a unique $x : [0,$

$\nabla^2\Phi(x(t))\dot{x}(t) + \nabla$

which is a descent tra

$|x(t) - \bar{x}| \leq Ce^{-t}$

where C is a constant

Proof. The strong convexity inequality applied at points $x(t)$ and $\bar{x}$ gives:

$$k|x(t)-\bar{x}|^2 \leq \langle x(t)-\bar{x}, \nabla\Phi(x(t))\rangle.$$

Hence the strong convergence of $x(t)$ towards $\bar{x}$.

Now the consideration of the energy F shows that $< \nabla^2\Phi(x)\dot{x}, \dot{x} >$ is in $L^1([0,\infty[, H)$ and further:

$$\int_0^{\infty} < \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) > dt \geq k\int_0^{\infty} |\dot{x}(t)|^2 dt.$$

Since $|\dot{x}|^2$ is in $L^1([0,\infty[, H)$ and since its derivative $2 < x, \dot{x} >$ is bounded, $|\dot{x}(t)|^2$ tends to 0 as t tends to ∞; hence the strong convergence of $\dot{x}(t)$ towards 0 as t tends to ∞. $\triangle$

Observe that the convergence of the trajectories is established, in the convex case, under the hypothesis that they are bounded; while for a Morse function their convergence is established under a precompactness hypothesis.

4 A Few Remarks

In order to make the following remarks more clear, let us excerpt some results about Newton's continuous method from Aubin-Cellina's work [5], and from Alvarez-Perez' [2].

With some adaptation to our case and notations, theorem 4. p. 197 of [5] states:

Theorem 4.1 *Let Φ be a twice continuously differentiable map from $H = \mathbb{R}^n$ to $\mathbb{R}$; let $\nabla\Phi$ be bounded. It is assumed that:*

(*) $\exists c > 0$ such that $\forall x \in H, \exists u \in H$ satisfying $\nabla^2\Phi(x)u = -\nabla\Phi(x)$ and $|u| \leq c|\nabla\Phi(x)|$.

Then the implicit differential equation:

$\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0,\ x(0) = x_0 \in H,$

has for all $x_0 \in H$ trajectories $x(.)$ on $[0,\infty[$ satisfying $|\dot{x}(t)| \leq c|\nabla\Phi(x(t))|$.

Furthermore, for all $t \geq 0$, one has:

(i) $\nabla\Phi(x(t)) = e^{-t}\nabla\Phi(x_0)$,

(ii) $dist(x(t), \{u \in H / \nabla\Phi(u) = 0\}) \leq ce^{-t}|\nabla\Phi(x_0)|$.

Every cluster point of such trajectories is a critical point of Φ.

Note that a trajectory need not be unique.

Now in [2], corollary 3.1 states:

Theorem 4.2 *Let $\Phi \in C^2(H, \mathbb{R})$ be strongly convex. Then for $x_0 \in H$ there exists a unique $x : [0,\infty[\rightarrow H$ solution trajectory of*

$\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0,\ x(0) = x_0 \in H,$

which is a descent trajectory for Φ and satisfies:

$|x(t) - \bar{x}| \leq Ce^{-t}$

where C is a constant and $\bar{x}$ is the unique minimizer of Φ.

Let us also recall the global existence theorem for the HBF method ([4]):

Theorem 4.3 *Let us assume that $\Phi : H \to \mathbb{R}$ satisfies the following assumptions:*

$$\begin{cases} \Phi \text{ is continuously differentiable on } H \\ \Phi \text{ is bounded from below on } H \\ \nabla\Phi \text{ is Lipschitz continuous on the bounded subsets of } H, \end{cases}$$

and that the friction parameter λ is positive ($\lambda > 0$). Then, the following properties hold :

(i) *For all $(x_0, \dot{x}_0)$ in $H \times H$, there exists a unique solution $x(t)$ of (4) defined on the whole interval $[0, +\infty[$, which is of class C^2 on $[0, +\infty[$, and which satisfies the initial conditions $x(0) = x_0$ and $\dot{x}(0) = \dot{x}_0$.*

(ii) *For every trajectory $x(t)$ of (4), the energy $F(t) = \frac{1}{2}|\dot{x}(t))|^2 + \Phi(x(t))$ is decreasing on $[0, +\infty[$ and bounded from below, and hence converges to some real value F_∞. 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 $\lim_{t\to+\infty} \Phi(x(t)) = F_\infty$.*

Global existence. The first two propositions yield the existence of a trajectory under some invertibility of the Hessian; that is obvious in Alvarez-Perez' theorem, where strong convexity is required for Φ, and Aubin-Cellina's condition (*) is a sort of pseudo-invertibility. There is no such hypothesis in the study of the second-order Newton equation (3), where existence is proved under the mere Lipschitz continuity of the Hessian.

Critical points. The second-order Newton equation is devised with the hope that it bears some relationship with the minimization of Φ. Often one has to content oneself with the critical points of Φ, and more often one is happy to grasp a point where $\nabla\Phi$ is small (cf. [7]). Under the sole hypotheses $(\mathcal{H})$, theorem (2.1) tells us that $\nabla\Phi(x) \in L^2(0, +\infty; H)$, which implies $\liminf_{t\to+\infty} |\nabla\Phi(x(t))| = 0$. So along a trajectory there *are* points where $\nabla\Phi$ is arbitrarily small. Furthermore if the trajectory is bounded then $\nabla\Phi(x)$ is arbitrarily small for every point from some time t onwards ($\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$). In comparison with the HBF method, note that the latter gives information on $\nabla\Phi(x)$ only if x is supposed to be bounded, and then $\lim_{t\to+\infty} \nabla\Phi(x(t)) = 0$ indeed.

However the HBF method and the first-order continuous Newton method, for a convex functional at least for the latter, have minimization properties (cf. the above theorem of Alvarez-Perez, and prop. 3.1 and th. 4.2 in [4]). This is not so clear for the second-order Newton equation with a non-convex potential and deserves further studies.

Asymptotic behaviour of $\dot{x}$. In contrast to the HBF method note that, for a second-order Newton trajectory x with a non strongly convex functional, it is still an open question to know if $\lim_{t\to\infty} \dot{x}(t) = 0$. On the whole, convergence results are poorer for the second-order Newton equation, a fortiori we do not have any rate of convergence.

Energy. The proof of our global existence theorem (2.1) runs along the same lines as the HBF method; but it is much simpler. This may be due to the energy functionals that are used:

$$E(t) = \frac{1}{2}|\dot{x}(t) + \nabla\Phi(x(t))|^2 + \Phi(x(t)),$$

for the second-order Newton equation, and:

$$F(t) = \frac{1}{2}|\dot{x}(t)|^2 + \Phi(x(t)),$$

for the HBF equation. Remark that the dissipation rates are :

$$\dot{E}(t) = -|\nabla\Phi(x(t))|^2,$$

$$\dot{F}(t) = -|\dot{x}(t)|^2,$$

and may account for the properties of the methods that we have alluded to above.

Now the functional F is still meaningful along a second-order Newton trajectory; and its derivative is *in that case*: $\dot{F}(t) = -< \nabla^2\Phi(x(t))\dot{x}(t), \dot{x}(t) >$. So it is an energy functional, along with E, for a convex potential. Oddly enough taking this energy into account in the convex case does not help much, when the Hessian happens to be singular.

References

1. F. Alvarez, On the minimizing property of a second order dissipative system in Hilbert space, preprint 98-05, Département de Mathématiques, Université Montpellier II, to appear in SIAM J. of Control and Optimization.
2. F. Alvarez, J.M. Perez, Asymptotic analysis of evolution equations associated with Newton's method for parametric approximations of convex minimization problems, Appl. Math. Optim. 38 (1998), 193-217.
3. H. Attouch and R. Cominetti, A dynamical approach to convex minimization coupling approximation with the steepest descent method, J. Differential Equations, 128 (2), (1996), 519-540.

4. H. Attouch, X. Goudou and P. Redont, The heavy ball with friction method: I the continuous dynamical system, preprint 1998/11, Département de Mathématiques, Université de Montpellier II; to appear in Communications in Contemporary Mathematics.
5. J.-P. Aubin, A. Cellina, Differential Inclusions, Springer, 1984.
6. J.-B. Baillon, Un exemple concernant le comportement asymptotique de la solution du problème $du/dt + \partial\phi(u) = 0$, Journal of Functional Analysis 28, 369-376 (1978).
7. J. F. Bonnans, J.-Ch. Gilbert, C. Lemaréchal, C. Sagastizábal, Méthodes Numériques d'optimisation, Springer, 1998.
8. H. Brézis, Opérateurs maximaux monotones, Mathematics Studies 5, North-Holland-American Elsevier, 1973.
9. R.E. Bruck, Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, Journal of Functional Analysis, 18, (1975), 15-26.
10. A. Haraux, Systèmes dynamiques dissipatifs et applications, RMA 17, Masson, Paris, (1991).
11. H. Th. Jongen, A. Ruiz Jhones, Nonlinear optimization: on the min-max digraph and global smoothing, Calculus of variations and differential equations (Haifa, 1998), 119-135, Chapman & Hall/CRC Res. Notes Math., Chapman & Hall/CRC, Boca Raton, FL, 2000.
12. S. Lojasiewicz, Une propriété topologique des sous-ensembles analytiques réels. Colloques internationaux du C.N.R.S, n°117. Les Equations aux dérivées partielles (1963).
13. S. Lojasiewicz, Ensembles semi-analytiques, notes I.H.E.S. (1965).
14. J. Palis and W. de Melo, Geometric theory of dynamical systems, Springer, 1982.

Polynomial Dens
and Representati
Generate a Deter
Problem

Andrew G. Bakan

Institute of Mathematics, Na
Ukraine; email: andrew@bak

Abstract. For a positive B
bounded support it has bee
space $L_p(\mathbb{R}, d\mu)$, $1 \le p < \infty$,
following form: $d\mu(x) = w(x$
on $\mathbb{R}$ and $w : \mathbb{R} \to [0,1]$ is
braic polynomials are dense
representation of all measur
problem.

Keywords: moment proble

AMS classification: 42C0

1 Introduction

Consider an arbitrary up
isfying $\|x^n\|_w < \infty$ $\forall n$
$\sup_{x\in\mathbb{R}} w(x)|f(x)| =: \|w\cdot$
C^0_w consists of all function
$\lim_{|x|\to\infty} w(x)f(x) = 0$. L
$\mathcal{E}_0$ and Λ_f denote the set
and the set of all zeros of
the so-called Hamburger
only real and simple zero
bers $|f'(\lambda)|$, $\lambda \in \Lambda_f$, ten
λ as $|\lambda| \to \infty$; b) it is p
of the simple fractions: 1

In 1924, S. Bernstein
polynomials (with compl
more explicit survey see
obtained by S. Isumi an
if a function $w(x)$ is po
whole real line, then the

MARC LASSONDE
Editor

Approximation, Optimization and Mathematical Economics

Physica-Verlag
A Springer-Verlag Company