JOURNAL OF DIFFERENTIAL EQUATIONS **128**, 519–540 (1996)
ARTICLE NO. 0104

A Dynamical Approach to Convex Minimization Coupling Approximation with the Steepest Descent Method*

H. ATTOUCH

Université de Montpellier, Place Eugène Bataillon, Montpellier, France

AND

R. COMINETTI

Universidad de Chile, Casilla 170/3 Correo 3, Santiago, Chile

We study the asymptotic behavior of the solutions to evolution equations of the form $0 \in \dot{u}(t) + \partial f(u(t), \varepsilon(t))$; $u(0) = u_0$, where $\{f(\cdot, \varepsilon): \varepsilon > 0\}$ is a family of strictly convex functions whose minimum is attained at a unique point $x(\varepsilon)$. Assuming that $x(\varepsilon)$ converges to a point x^* as ε tends to 0, and depending on the behavior of the optimal trajectory $x(\varepsilon)$, we derive sufficient conditions on the parametrization $\varepsilon(t)$ which ensure that the solution $u(t)$ of the evolution equation also converges to x^* when $t \to +\infty$. The results are illustrated on three different penalty and viscosity-approximation methods for convex minimization.

1. INTRODUCTION

Let H be a real Hilbert space with scalar product and norm denoted by $\langle \cdot, \cdot \rangle$ and $|\cdot|$ respectively. Given a convex, lower semicontinuous (lsc), proper function $f: H \to \mathbb{R} \cup \{+\infty\}$, we consider the minimization problem

$$\min\{f(x): x \in H\} \tag{P}$$

and we assume that the (closed convex) set of optimal solutions is non-empty

$$S(P) := \{x \in H: f(x) = \inf_H f\} \neq \varnothing.$$

* This work was done while the second author was visiting Université de Montpellier, January 1994. Partially supported by ECOS.

0022-0396/96 $18.00

As a matter of fact, we are specially interested in cases where (P) has multiple optimal solutions (the function f is not assumed to be strictly convex). This corresponds to some important situations as linear programming and semi-coercive minimization problems.

For each $\varepsilon > 0$, let us consider the approximate minimization problem

$$\min\{f(x, \varepsilon) : x \in H\} \qquad (P_\varepsilon)$$

where $f(\cdot, \varepsilon)$ is strictly convex, lsc and proper, and let us assume that

$$\begin{cases} \textit{There exists a unique solution } x(\varepsilon) \textit{ of } (P_\varepsilon). \\ \textit{The filtered sequence } \{x(\varepsilon); \varepsilon \to 0\} \textit{ norm converges to some } x^* \in S(P). \end{cases}$$

This is a common feature of all the viscosity approximation methods, see Tikhonov and Arsenine [23] and Attouch [3] for a recent survey on the "viscosity selection principle". In linear mathematical programming, this property is satisfied by most of the barrier methods, as well as the exponential penalty method recently studied by Cominetti and San Martín [12].

Our goal is to construct a general dynamical method which allows to compute the particular solution x^*. The computation of x^* is of interest since the solution obtained by the regularization-approximation method usually enjoys nice geometrical or variational properties, see [1], [3], [12].

When considering the initial problem (P), it is a classical result that the trajectory defined by the (generalized) steepest descent method

$$(SD) \begin{cases} \dot{u}(t) + \partial f(u(t)) \ni 0 \\ u(0) = u_0 \end{cases}$$

weakly converges to some element $u_\infty \in S(P)$ as t goes to $+\infty$, see Brézis [8, 9], Brück [10]. The limit u_∞ depends on u_0 and it is quite difficult to characterize, see Lemaire [17] for recent results in this direction. Moreover, the computation of the trajectory $t \to u(t)$ may become involved because of the operator ∂f. For instance, when f is an extended real-valued function (taking the value $+\infty$ because of the constraints), ∂f is a multi-valued operator which is not defined everywhere.

The idea to be developed in this paper is to take advantage of the regularizing properties of the approximations (P_ε), and to consider the "Descent and Approximation Dynamical Asymptotical" method

$$(DADA) \begin{cases} \dot{u}(t) + \partial f(u(t), \varepsilon(t)) \ni 0 \\ u(0) = u_0 \end{cases}$$

where $\varepsilon: [0, +\infty) \to \mathbb{R}_+$ is some strictly positive function decreasing to 0 with $t \to +\infty$. Of particular interest are approximations (P_ε) defined by sufficiently smooth functions $f(\cdot, \varepsilon)$, in which case $(DADA)$ becomes a (nonautonomous) differential equation.

Our main concern is the asymptotic behavior of the solutions of $(DADA)$ as t goes to $+\infty$, which depends critically on the rate at which $\varepsilon(t)$ goes to zero. We are specially interested in conditions ensuring the convergence of $u(t)$ towards the particular solution x^*.

Let us illustrate this on the following elementary example:

Example. Take $H = \mathbb{R}$, $f \equiv 0$ and $f(x, \varepsilon) = \varepsilon/2\,|x|^2$ (viscosity method). Then $S(P) = \mathbb{R}$, the approximate problem (P_ε) achieves its unique minimum at $x(\varepsilon) = 0$, and $x^* = \lim_{\varepsilon \to 0} x(\varepsilon) = 0$. The solution $u(\cdot)$ of the dynamical approximation method

$$\dot{u}(t) + \varepsilon(t)\, u(t) = 0;\ u(0) = u_0$$

is given by

$$u(t) = u_0 \exp\left[-\int_0^t \varepsilon(s)\, ds \right].$$

Clearly, $\lim_{t \to +\infty} u(t) = x^*$ iff $\int_0^{+\infty} \varepsilon(s)\, ds = +\infty$, in which case the limiting behavior does not depend on the initial data u_0. Otherwise, $\lim_{t \to +\infty} u(t)$ exists but differs from x^*. Note that, depending on the choice of the time parametrization $\varepsilon(t)$, one can obtain as a limit of the $(DADA)$ trajectory any point on the interval $[0, u_0]$.

This example suggests that when $\varepsilon(t)$ goes to zero "slowly", then $\lim_{t \to +\infty} u(t) = x^*$ and the $(DADA)$ trajectory asymptotically approaches the trajectory $\varepsilon \to x(\varepsilon)$ of approximate solutions. We shall refer to this situation as case (A). On the other extreme case, if $\varepsilon(t)$ goes to zero "fastly" with t, the $(DADA)$ trajectory stays close to the steepest descent trajectory. This is case (B).

As we already said, in this paper we shall be mostly interested in case (A), where $\varepsilon(t)$ decays "slowly". We shall find, in a fairly general setting, precise estimates on the decay of $\varepsilon(t)$ ensuring that $\lim u(t) = x^*$. As far as we know, this is a new approach[1] for studying the asymptotics of the

[1] After this paper was accepted for publication, the reference *A new approach to the investigation of evolution differential equations in Banach spaces* by Ya. I. Alber, appeared in *Nonlinear Analysis, Theory, Methods and Applications*, Vol. **23**, No. 9, 1994, pp. 1115–1134. This reference studies the asymptotic convergence of solutions to non-autonomous differential equations governed by strongly maximal monotone operators (and not only subdifferentials as we do), but assuming strong monotonicity and uniqueness of the solution for the limit problem (in our context, uniqueness and strong convexity of (P)). In this sense the results are complementary, while one may find some similarities between both approaches.

trajectories defined by $(DADA)$. There is certainly some analogy with numerical discretizations of evolution equations where one has to control the respective size of the time and space discretizations.

In contrast, case (B) is in some sense more natural: the error made when replacing f by $f(\cdot, \varepsilon(t))$ is controlled so that one may adapt to this situation the standard proof of the asymptotic behavior for the steepest descent. In this direction, let us mention the work of Furuya, Miyashiba and Kenmochi [14], and for the discretized implicit version (proximal method) the works of Lemaire [16], Tossings [22], Moudafi [19].

Let us finally notice that it is an open problem to know if, for an arbitrary parametrization $t \to \varepsilon(t)$ (decreasing to zero as t increases to $+\infty$), the limit of $u(t)$ as t goes to $+\infty$, exists or not.

2. Basic Assumptions

We consider a Hilbert space H and a family of strictly convex, proper, lsc functions $\{f(\cdot, \varepsilon): \varepsilon > 0\}$ such that the minimum

$$\min\{f(x, \varepsilon): x \in H\} \tag{P_ε}$$

is attained at a unique point $x(\varepsilon)$. We assume the optimal trajectory $x(\varepsilon)$ is absolutely continuous on every compact interval $[\varepsilon_1, \varepsilon_2]$ with $0 < \varepsilon_1 \leqslant \varepsilon_2 < +\infty$, and that it converges towards some $x^* \in H$ when $\varepsilon \to 0$. Notice that we have excluded the case $\varepsilon_1 = 0$ so that we allow $|dx/d\varepsilon\,(\varepsilon)|$ to diverge when ε goes to 0.

Let us mention that this absolute continuity property is satisfied for a number of specific approximations schemes as illustrated in the examples presented later on. Roughly speaking, such a property may be established by using variants of the implicit function theorem. In the simplest case, when the mapping $(x, \varepsilon) \to f(x, \varepsilon)$ is of class C^2 and the Hessian of $f(\cdot, \varepsilon)$ is non-singular at $x(\varepsilon)$, the trajectory $\varepsilon \to x(\varepsilon)$ turns out to be of class C^1 for $\varepsilon > 0$, hence absolutely continuous on each compact interval of the form $[\varepsilon_1, \varepsilon_2]$. Situations with weaker differentiability assumptions on the data, as Tikhonov regularization and viscosity-penalization schemes, will be discussed in the examples.

We also consider a continuous decreasing function $\varepsilon: [0, +\infty) \to \mathbb{R}_+$ with strictly positive values and

$$\lim_{t \to +\infty} \varepsilon(t) = 0,$$

and we assume that the evolution equation

$$(DADA)\ \begin{cases} \dot{u}(t) + \partial f(u(t), \varepsilon(t)) \ni 0 \\ u(0) = u_0 \end{cases}$$

has a (necessarily unique) strong global solution, that is, there exists an absolutely continuous function u: $[0, +\infty) \to H$ such that $u(0) = u_0$ and

$$0 \in \dot{u}(t) + \partial f(u(t), \varepsilon(t)) \qquad \text{a.e. in} \quad (0, +\infty).$$

The existence of global solutions for differential inclusions like the previous one, has been studied in a number of papers. Some relevant references in this respect are Attouch and Damlamian [4], Auchmuty [7], Kenmochi [15], Tataru [20].

3. The "Slow" Parametrization

A basic property of the approximation problem (P_ε) which ensures the uniqueness of the solution $x(\varepsilon)$, as well as its good numerical conditioning, is the strong convexity of the function $f(\cdot, \varepsilon)$. Equivalently, in terms of subdifferential operators, we consider the strong monotonicity property: for every $\varepsilon > 0$ there exists a constant $\beta(\varepsilon) > 0$ such that

$$\langle u - v, x - y \rangle \geqslant \beta(\varepsilon)\, |x - y|^2 \tag{H}$$

for all x, $y \in H$ and $u \in \partial f(x, \varepsilon)$, $v \in \partial f(y, \varepsilon)$.

Since nonuniqueness of problem (P) is a central issue in this paper, the function $\beta(\varepsilon)$ is allowed to go to zero as ε tends to zero. The rate at which $\beta(\varepsilon)$ goes to zero plays a crucial role in the following results, since it measures the attraction of the curve of approximate solutions over the $(DADA)$ trajectory.

Some approximation schemes $f(\cdot, \varepsilon)$ may fail to satisfy the strong monotonicity condition (H) over the whole space H, but they may satisfy this assumption over bounded sets, namely, for every $\varepsilon > 0$, $K > 0$ there exists a constant $\beta_K(\varepsilon) > 0$ such that

$$\langle u - v, x - y \rangle \geqslant \beta_K(\varepsilon)\, |x - y|^2 \tag{H_K}$$

for all x, $y \in H$ with $|x| \leqslant K$, $|y| \leqslant K$ and $u \in \partial f(x, \varepsilon)$, $v \in \partial f(y, \varepsilon)$.

Before using this local form of strong monotonicity for the asymptotic analysis of the $(DADA)$ trajectory $u(t)$, one must somehow ensure that $u(t)$ stays bounded, that is, we must find an *a priori* bound $K > 0$ such that $|u(t)| \leqslant K$ for all $t > 0$. In the next subsection 3.1, we will show that this is the case when the optimal path $\varepsilon \to x(\varepsilon)$ has a finite length, giving rise to our first result on the asymptotic convergence of $u(t)$, Theorem 3.2.

In Sect. 3.2 we consider the case of a *long* optimal path $\varepsilon \to x(\varepsilon)$, where no *a priori* bound on $u(t)$ is at our disposal and convergence of $u(t)$ towards x^* will be derived on the basis of the global strong monotonicity property (H) (which implies *a fortiori* that $u(t)$ stays bounded).

3.1. *The case of a "short" optimal trajectory.*

For the reader's convenience we present the following simple lemma which shall be used several times in the forthcoming proofs.

LEMMA 3.1. *Let* $m, \theta\colon [t_0, t_1] \to [0, +\infty[$ *with* m *integrable and* θ *absolutely continuous, such that*

$$\dot{\theta}(t) \leqslant m(t)\sqrt{\theta(t)} \qquad \text{a.e. on} \quad [t_0, t_1]. \tag{1}$$

Then, for all $t \in [t_0, t_1]$ *we have*

$$\sqrt{\theta(t)} \leqslant \sqrt{\theta(t_0)} + \tfrac{1}{2}\int_{t_0}^{t} m(s)\, ds.$$

Proof. For each $\alpha > 0$ the function $\theta_\alpha(t) = \sqrt{\alpha + \theta(t)}$ is absolutely continuous and from (1) we have $\dot{\theta}_\alpha(t) \leqslant m(t)/2$ almost everywhere on $[t_0, t_1]$. After integration we get for all $t \in [t_0, t_1]$

$$\theta_\alpha(t) \leqslant \theta_\alpha(t_0) + \tfrac{1}{2}\int_{t_0}^{t} m(s)\, ds,$$

and the conclusion follows by letting $\alpha \to 0$. ∎

The following is our first result on the asymptotics of $(DADA)$.

THEOREM 3.2. *In addition to the basic assumptions in* Sect. 2, *let us suppose that the optimal trajectory* $x(\varepsilon)$ *has finite length, that is to say, for each* $\varepsilon_0 > 0$ *we have*

$$\int_0^{\varepsilon_0} \left| \frac{dx}{d\varepsilon} \right| d\varepsilon < +\infty.$$

Suppose that for all $K > 0$ *condition* (H_K) *holds, the function* $\beta_K(\cdot)$ *is measurable, and the function* $\varepsilon(t)$ *satisfies*

$$\int_0^{+\infty} \beta_K(\varepsilon(s))\, ds = +\infty.$$

Then the solution $u(t)$ *of* $(DADA)$ *norm converges to* x^* *as* $t \to +\infty$.

Proof. Since $x(\varepsilon(t))$ converges to x^*, it suffices to show that the function $\varphi(t) := \frac{1}{2}|u(t) - x(\varepsilon(t))|^2$ converges to 0 as t goes to $+\infty$. This function is absolutely continuous and for almost all $t \in (0, +\infty)$ we have

$$\dot\varphi(t) = \left\langle \dot u(t) - \dot\varepsilon(t) \frac{dx}{d\varepsilon}(\varepsilon(t)), u(t) - x(\varepsilon(t)) \right\rangle$$
$$\leqslant \langle \dot u(t), u(t) - x(\varepsilon(t)) \rangle - \dot\varepsilon(t) \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| |u(t) - x(\varepsilon(t))|. \quad (2)$$

Since we have $-\dot u(t) \in \partial f(u(t), \varepsilon(t))$ a.e. and $0 \in \partial f(x(\varepsilon(t)), \varepsilon(t))$, the monotonicity of the subdifferential implies

$$\dot\varphi(t) \leqslant -\dot\varepsilon(t) \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| |u(t) - x(\varepsilon(t))| = -\dot\varepsilon(t) \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| \sqrt{2\varphi(t)}.$$

Using Lemma 3.1. we deduce

$$\sqrt{\varphi(t)} \leqslant \sqrt{\varphi(0)} - \frac{1}{\sqrt{2}} \int_0^t \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| \dot\varepsilon(t)\, dt$$
$$= \sqrt{\varphi(0)} + \frac{1}{\sqrt{2}} \int_{\varepsilon(t)}^{\varepsilon(0)} \left| \frac{dx}{d\varepsilon} \right| d\varepsilon$$
$$\leqslant \sqrt{\varphi(0)} + \frac{1}{\sqrt{2}} \int_0^{\varepsilon(0)} \left| \frac{dx}{d\varepsilon} \right| d\varepsilon$$

from which it follows that the $(DADA)$ trajectory $u(t)$ stays bounded.

Let $K > 0$ be such that $|u(t)| \leqslant K$ and $|x(\varepsilon(t))| \leqslant K$ for all $t \in [0, +\infty)$. Invoking assumption (H_K) and using (2) we get

$$\dot\varphi(t) \leqslant -\beta_K(\varepsilon(t))\, |u(t) - x(\varepsilon(t))|^2 - \dot\varepsilon(t) \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| |u(t) - x(\varepsilon(t))|$$

that is to say

$$\dot\varphi(t) + 2\beta_K(\varepsilon(t))\, \varphi(t) \leqslant -\sqrt{2}\dot\varepsilon(t) \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| \sqrt{\varphi(t)}.$$

Denoting $E(t) = \int_0^t \beta_K(\varepsilon(s))\, ds$ (we may always assume that β_K is bounded from above so that the integral is finite) and multiplying the above inequality by $\exp(2E(t))$ we get

$$\frac{d}{dt}[e^{2E(t)}\varphi(t)] \leqslant -\sqrt{2} e^{E(t)} \left| \frac{dx}{d\varepsilon}(\varepsilon(t)) \right| \dot\varepsilon(t) \sqrt{e^{2E(t)}\, \varphi(t)}$$

so we may use Lemma 3.1 with $\theta(t)=e^{2E(t)}\varphi(t)$ in order to deduce for all $0 \leqslant t_0 < t < +\infty$

$$e^{E(t)}\sqrt{\varphi(t)} \leqslant e^{E(t_0)}\sqrt{\varphi(t_0)} - \frac{1}{\sqrt{2}}\int_{t_0}^{t} e^{E(s)} \left|\frac{dx}{d\varepsilon}(\varepsilon(s))\right| \dot{\varepsilon}(s)\, ds. \tag{3}$$

Multiplying by $\exp(-E(t))$ and noting that $E(\cdot)$ is increasing we get

$$\begin{aligned}
\sqrt{\varphi(t)} &\leqslant e^{-(E(t)-E(t_0))}\sqrt{\varphi(t_0)} - \frac{1}{\sqrt{2}}\int_{t_0}^{t} e^{-(E(t)-E(s))} \left|\frac{dx}{d\varepsilon}(\varepsilon(s))\right| \dot{\varepsilon}(s)\, ds \\
&\leqslant e^{-(E(t)-E(t_0))}\sqrt{\varphi(t_0)} - \frac{1}{\sqrt{2}}\int_{t_0}^{t} \left|\frac{dx}{d\varepsilon}(\varepsilon(s))\right| \dot{\varepsilon}(s)\, ds \\
&\leqslant e^{-(E(t)-E(t_0))}\sqrt{\varphi(t_0)} + \frac{1}{\sqrt{2}}\int_{\varepsilon(t)}^{\varepsilon(t_0)} \left|\frac{dx}{d\varepsilon}\right| d\varepsilon.
\end{aligned} \tag{4}$$

Using the fact that $E(t) \to +\infty$ we obtain

$$\limsup_{t\to+\infty} \sqrt{\varphi(t)} \leqslant \frac{1}{\sqrt{2}}\int_{0}^{\varepsilon(t_0)} \left|\frac{dx}{d\varepsilon}\right| d\varepsilon$$

so that letting $t_0 \to +\infty$ and using the finite length assumption we conclude

$$\limsup_{t\to+\infty} \sqrt{\varphi(t)} = 0$$

as was to be proved. ∎

Example (The Exponential Penalty Method). Consider the linear programming problem

$$\min_{x\in\mathbb{R}^n} \{c'x\colon Ax \leqslant b\} \tag{P}$$

and its nonlinear unconstrained approximation

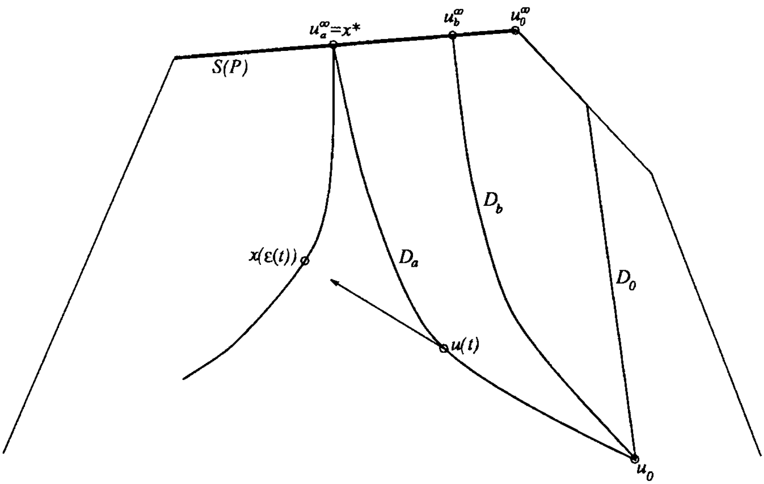
$$\min_{x\in\mathbb{R}^n} \left\{c'x + \varepsilon \sum_{i=1}^{m} \exp[(A_i x - b_i)/\varepsilon]\right\}. \tag{P_ε}$$

Under very mild assumptions (boundedness of the optimal set) it is known [12] that the trajectory $x(\varepsilon)$ is Lipschitz continuous on $[0, \varepsilon_0]$ and that

$\lim_{\varepsilon \to 0} x(\varepsilon) = x^*$ exists, x^* being characterized as the "centroid" of the optimal face of (P). Moreover, a direct calculation of the Hessian of $f(\cdot, \varepsilon)$ yields that for each $K > 0$ there exist constants $\alpha > 0$, $M > 0$ such that $\beta_K(\varepsilon) \geqslant \alpha/\varepsilon\, e^{-M/\varepsilon}$ (notice that no "global" $\beta(\varepsilon)$ may be found in this case).

Thus, the length of the trajectory $x(\varepsilon)$ is obviously finite and the condition $\int_0^{+\infty} \beta_K(\varepsilon(s))\, ds = +\infty$ is satisfied if $\int_0^{+\infty} 1/\varepsilon(s)\, e^{-M/\varepsilon(s)}\, ds = +\infty$. This holds for instance if $\varepsilon(t) = M/\ln(1+t)$. Thus, in order to reach the centroid x^*, one has to choose a very slow parametrization $\varepsilon(t)$.

The following picture illustrates the above results ($(DADA)$ method for linear programming and exponential penalization).



Let us examine the different trajectories:

1. D_0: It is the steepest descent trajectory ("free" trajectory which is orthogonal to the level sets, that is, hyperplans parallel to the optimal face $S(P)$, until it hits the boundary of the feasible region...). It converges to an optimal solution u_∞^0 of the linear program.

Trajectories D_a and D_b correspond to $(DADA)$ solutions respectively in

2. D_a: "slow" parametrization and as a consequence "strong" attraction of the curve $t \to x(\varepsilon(t))$ on the curve $t \to u(t)$. At time t, the curve $u(t)$ behaves like the steepest descent trajectory for the function $f(\cdot, \varepsilon(t))$ and is therefore attracted by its unique minimizer $x(\varepsilon(t))$. In this case, the limit u_∞^a is equal to x^*.

3. D_b: "fast" parametrization and the attraction is "weak". The curve D_b ends up at a point u^b_∞ which, a priori, may be any point of $S(P)$. For a proof of convergence in case (B) for the exponential penalty we refer to [11].

3.2. *The case of a "long" optimal trajectory.*

Let us now consider the case where the trajectory $\varepsilon \to x(\varepsilon)$ may have an infinite length.

THEOREM 3.3. *In addition to the basic assumptions of* Sect. 2 *let us assume that the trajectory of approximate solutions* $x(\varepsilon)$ *satisfies*

$$\left| \frac{dx}{d\varepsilon}(\varepsilon) \right| \leqslant \frac{1}{\gamma(\varepsilon)} \qquad \textit{a.e. on} \quad]0, +\infty[$$

for some function γ *possibly tending to zero as* $\varepsilon \to 0$. *Suppose that* (H) *is satisfied with* $\beta(\cdot)$ *measurable, and let us take a parametrization* $\varepsilon(t)$ *such that*

$$\text{(i)} \qquad \int_0^{+\infty} \beta(\varepsilon(s))\, ds = +\infty,$$

$$\text{(ii)} \qquad \lim_{t \to +\infty} \frac{\dot{\varepsilon}(t)}{\beta(\varepsilon(t))\, \gamma(\varepsilon(t))} = 0.$$

Then, the $(DADA)$ *trajectory* $u(t)$ *norm converges to* x^* *as* $t \to +\infty$.

Proof. Proceeding as in the proof of Theorem 3.2 with $\beta(\cdot)$ instead of $\beta_K(\cdot)$ we obtain (see (4))

$$\sqrt{\varphi(t)} \leqslant e^{-(E(t)-E(t_0))} \sqrt{\varphi(t_0)} - \frac{1}{\sqrt{2}} e^{-E(t)} \int_{t_0}^{t} e^{E(s)} \left| \frac{dx}{d\varepsilon}(\varepsilon(s)) \right| \dot{\varepsilon}(s)\, ds$$

where $E(t) = \int_0^t \beta(\varepsilon(s))\, ds$. Defining

$$h(t_0) = \sup_{s \geqslant t_0} \frac{|\dot{\varepsilon}(s)|}{\beta(\varepsilon(s))\, \gamma(\varepsilon(s))}$$

and since $\dot{E}(s) = \beta(\varepsilon(s))$, it follows that

$$\begin{aligned} \sqrt{\varphi(t)} &\leqslant e^{-(E(t)-E(t_0))} \sqrt{\varphi(t_0)} + \frac{1}{\sqrt{2}} e^{-E(t)}\, h(t_0) \int_{t_0}^{t} e^{E(s)} \beta(\varepsilon(s))\, ds \\ &= e^{-(E(t)-E(t_0))} \sqrt{\varphi(t_0)} + \frac{1}{\sqrt{2}} e^{-E(t)}\, h(t_0) [e^{E(t)} - e^{E(t_0)}] \\ &\leqslant e^{-(E(t)-E(t_0))} \sqrt{\varphi(t_0)} + \frac{1}{\sqrt{2}} h(t_0). \end{aligned}$$

Using (i) we obtain

$$\limsup_{t \to +\infty} \sqrt{\varphi(t)} \leqslant \frac{1}{\sqrt{2}} h(t_0)$$

and since by (ii) we have $h(t_0) \to 0$ when $t_0 \to +\infty$, we conclude

$$\limsup_{t \to +\infty} \sqrt{\varphi(t)} \leqslant 0$$

achieving the proof. ∎

Example (The Viscosity Method). We consider for simplicity the classical Tikhonov regularization: f is a closed proper convex function with Argmin $f \neq \emptyset$, and $f(x, \varepsilon) = f(x) + \varepsilon/2\, |x|^2$, so that $x(\varepsilon)$ is uniquely determined by

$$\partial f(x(\varepsilon)) + \varepsilon x(\varepsilon) \ni 0. \tag{5}$$

It is a classical result that $x(\varepsilon)$ is bounded iff Argmin $f \neq \emptyset$ which is our basic assumption, and in that case

$$\lim_{\varepsilon \to 0} x(\varepsilon) = \text{proj}_{S(P)} 0 = x^*.$$

In other words, x^* is the element of minimal norm in $S(P) = \text{Argmin}\, f$.

In order to study the differentiability properties of the mapping $\varepsilon \to x(\varepsilon)$, we give two proofs of independent interest.

(a) The first one, which is the simplest, relies on the implicit function theorem for equations governed by possibly multivalued operators, see Aubin [6]. We do not enter into the details, just say that a formal calculation gives a bound for $|\dot{x}(\varepsilon)|$ which can be further justified. So, let us assume that f is smooth ($\mathscr{C}^2$ for example) and let us differentiate (5) to obtain:

$$\nabla_x^2 f(x(\varepsilon))\, \dot{x}(\varepsilon) + \varepsilon \dot{x}(\varepsilon) + x(\varepsilon) = 0.$$

Multiplying this equation by $\dot{x}(\varepsilon)$ we get

$$\langle \nabla_x^2 f(x(\varepsilon))\, \dot{x}(\varepsilon), \dot{x}(\varepsilon) \rangle + \varepsilon\, |\dot{x}(\varepsilon)|^2 + \langle x(\varepsilon), \dot{x}(\varepsilon) \rangle = 0.$$

By convexity of f the first term above is nonnegative, so that

$$\varepsilon\, |\dot{x}(\varepsilon)|^2 \leqslant -\langle x(\varepsilon), \dot{x}(\varepsilon) \rangle$$

and then

$$|\dot{x}(\varepsilon)| \leqslant \frac{1}{\varepsilon} |x(\varepsilon)|.$$

Finally, since the trajectory $x(\varepsilon)$ is bounded we obtain that for some constant $C > 0$

$$\left| \frac{dx}{d\varepsilon}(\varepsilon) \right| \leqslant \frac{C}{\varepsilon}.$$

(b) The second proof relies on the resolvent equation. Note that

$$x(\varepsilon) = \left(I + \frac{1}{\varepsilon} \partial f \right)^{-1} (0) = J_{1/\varepsilon}(0)$$

where $J_\lambda x = (I + \lambda A)^{-1} x$ is the resolvent of index $\lambda > 0$ of the maximal monotone operator A. Here $A = \partial f$ is the subdifferential of the convex lsc proper function f.

From the resolvent equation,

$$\forall \lambda, \mu > 0 \quad J_\lambda x = J_\mu \left(\frac{\mu}{\lambda} x + \left(1 - \frac{\mu}{\lambda} \right) J_\lambda x \right)$$

we infer

$$J_\lambda x - J_\mu x = J_\mu \left(\frac{\mu}{\lambda} x + \left(1 - \frac{\mu}{\lambda} \right) J_\lambda x \right) - J_\mu x.$$

Since $x \to J_\mu x$ is a contraction we have

$$|J_\lambda x - J_\mu x| \leqslant \left| 1 - \frac{\mu}{\lambda} \right| |x - J_\lambda x| = |\mu - \lambda| \left| \frac{x - J_\lambda x}{\lambda} \right|.$$

Going back to $x(\varepsilon) = J_{1/\varepsilon}(0)$ we obtain

$$|x(\varepsilon) - x(\varepsilon')| = |J_{1/\varepsilon}(0) - J_{1/\varepsilon'}(0)| \leqslant \left| \frac{1}{\varepsilon} - \frac{1}{\varepsilon'} \right| \varepsilon |x(\varepsilon)|$$

so that there exists a constant $C > 0$ such that

$$|x(\varepsilon) - x(\varepsilon')| \leqslant \frac{C}{\varepsilon'} |\varepsilon - \varepsilon'|.$$

Hence $x(\varepsilon)$ is locally Lipschitz on $]0, +\infty[$ and

$$\left|\frac{dx}{d\varepsilon}(\varepsilon)\right| \leqslant \frac{C}{\varepsilon}$$

almost everywhere.

Let us now return to the corresponding conditions for the parametrization $\varepsilon(t)$ in order to have $\lim u(t) = x^*$. We can take $\beta(\varepsilon) = \varepsilon$ so that the conditions in Theorem 3.3 turn into

$$\text{(i)} \qquad \int_0^{+\infty} \varepsilon(s)\, ds = +\infty$$

$$\text{(ii)} \qquad \lim_{t \to +\infty} \dot{\varepsilon}(t)/\varepsilon(t)^2 = 0.$$

If we take for instance a function of the form $\varepsilon(t) = 1/t^\alpha$, these conditions are fulfilled as far as $0 < \alpha < 1$. We summarize the previous discussion in the following statement.

PROPOSITION 3.4. *For every $u_0 \in \operatorname{dom} f$ and each $0 < \alpha < 1$, the solution of*

$$\begin{cases} \dot{u}(t) + \partial f(u(t)) + 1/t^\alpha\, u(t) = 0 \\ u(0) = u_0 \end{cases}$$

tends to x^* the element of minimal norm in $\operatorname{Argmin} f$ *as $t \to +\infty$.*

REMARK. One can take in fact $u_0 \in \overline{\operatorname{dom} f}$ (cf. regularization effect, Brézis [8]).

EXAMPLE (The Penalization-Viscosity Approximation). Let us start with the constrained convex minimization problem

$$\min\{f_0(x) \colon x \in C\} \qquad (P)$$

where $C \neq \varnothing$ is a closed convex subset of H, and f_0 is a convex Lipschitz continuous function on H. Let us assume that the set of solutions $S(P) = \operatorname{Argmin}_C f_0$ is non empty and consider, for each $\varepsilon > 0$, the approximate minimization problem ($\theta > 0$ is a positive parameter)

$$\min\left\{f_0(x) + \frac{1}{2\varepsilon^\theta} \operatorname{dist}(x, C)^2 + \frac{\varepsilon}{2}|x|^2 \colon x \in H\right\}. \qquad (P_\varepsilon)$$

This approximation procedure combines the exterior penalty method and the viscosity method, see Torralba [21] for more details on this method (note the analogy with the exponential penalty method).

(a) Let us examine the existence and convergence of an optimal trajectory $x(\varepsilon)$. For each $\varepsilon > 0$, the function $x \to f_0(x) + 1/2\varepsilon^\theta \operatorname{dist}(x, C)^2 + \varepsilon/2\, |x|^2$ is strictly convex, continuous and coercive, so there exists a unique solution $x(\varepsilon)$ of (P_ε). Let us prove that

$$\lim_{\varepsilon \to 0} x(\varepsilon) = \operatorname{proj}_{S(P)}(0) = x^*$$

the element of minimal norm in $S(P)$. By definition of $x(\varepsilon)$, for all $\varepsilon > 0$ we have

$$f_0(x(\varepsilon)) + \frac{1}{2\varepsilon^\theta} \operatorname{dist}(x(\varepsilon), C)^2 + \frac{\varepsilon}{2} |x(\varepsilon)|^2 \leqslant f_0(x^*) + \frac{\varepsilon}{2} |x^*|^2. \tag{6}$$

If k is a Lipschitz constant for f_0 then

$$f_0(x(\varepsilon)) \geqslant f_0(\operatorname{proj}_C x(\varepsilon)) - k \operatorname{dist}(x(\varepsilon), C) \geqslant f_0(x^*) - k \operatorname{dist}(x(\varepsilon), C)$$

from which we deduce

$$\frac{1}{2\varepsilon^\theta} \operatorname{dist}(x(\varepsilon), C)^2 + \frac{\varepsilon}{2} |x(\varepsilon)|^2 \leqslant k \operatorname{dist}(x(\varepsilon), C) + \frac{\varepsilon}{2} |x^*|^2.$$

Hence

$$\operatorname{dist}(x(\varepsilon), C)^2 - 2\varepsilon^\theta k \operatorname{dist}(x(\varepsilon), C) + \varepsilon^{\theta+1} |x(\varepsilon)|^2 \leqslant \varepsilon^{\theta+1} |x^*|^2.$$

Since $s^2 - 2\varepsilon^\theta ks \geqslant -(\varepsilon^\theta k)^2$ holds true for any $s \geqslant 0$, we obtain

$$\varepsilon^{\theta+1} |x(\varepsilon)|^2 \leqslant \varepsilon^{\theta+1} |x^*|^2 + k^2 \varepsilon^{2\theta}$$

so that

$$|x(\varepsilon)|^2 \leqslant |x^*|^2 + k^2 \varepsilon^{\theta-1}. \tag{7}$$

We conclude that for $\theta \geqslant 1$ the sequence $x(\varepsilon)$ is bounded and then, using (6), any weak limit point minimizes f_0 on C. Moreover, when $\theta > 1$ we obtain from (7) that the only possible weak limit point is x^* and also

$$\limsup_{\varepsilon \to 0} |x(\varepsilon)| \leqslant |x^*|$$

so that convergence occurs in the strong sense.

(b) Let us now study the differentiability properties of the trajectory $x(\varepsilon)$. Let us denote $f_1(x) = 1/2\ \mathrm{dist}(x, C)^2$ which is a convex $\mathscr{C}^{1,1}$ function. The optimality condition for (P_ε) is

$$\nabla f_0(x(\varepsilon)) + \frac{1}{\varepsilon^\theta} \nabla f_1(x(\varepsilon)) + \varepsilon x(\varepsilon) = 0$$

which can be rewritten as

$$\varepsilon^\theta \nabla f_0(x(\varepsilon)) + \nabla f_1(x(\varepsilon)) + \varepsilon^{\theta+1} x(\varepsilon) = 0.$$

Let us take $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ and denote $x_1 = x(\varepsilon_1)$, $x_2 = x(\varepsilon_2)$. We obtain

$$(\varepsilon_1^\theta - \varepsilon_2^\theta)\, \nabla f_0(x_1) + \varepsilon_2^\theta(\nabla f_0(x_1) - \nabla f_0(x_2)) + (\nabla f_1(x_1) - \nabla f_1(x_2)) \\ + \varepsilon_1^{\theta+1} x_1 - \varepsilon_2^{\theta+1} x_2 = 0.$$

Multiplying by $(x_1 - x_2)$ and using the monotonicity of ∇f_0 and ∇f_1 we get

$$(\varepsilon_1^\theta - \varepsilon_2^\theta)\langle \nabla f_0(x_1), x_1 - x_2\rangle + \langle \varepsilon_1^{\theta+1} x_1 - \varepsilon_2^{\theta+1} x_2, x_1 - x_2\rangle \leqslant 0$$

from which it follows that

$$\varepsilon_1^{\theta+1}\, |x_1 - x_2|^2 \leqslant |\varepsilon_1^\theta - \varepsilon_2^\theta|\ |\nabla f_0(x_1)|\ |x_1 - x_2| + |\varepsilon_1^{\theta+1} - \varepsilon_2^{\theta+1}|\ |x_2|\ |x_1 - x_2|.$$

Since the trajectory $x(\varepsilon)$ is bounded we may find a constant $C > 0$ such that

$$|x(\varepsilon_1) - x(\varepsilon_2)| \leqslant \frac{C}{\varepsilon_1^{\theta+1}} (|\varepsilon_1^\theta - \varepsilon_2^\theta| + |\varepsilon_1^{\theta+1} - \varepsilon_2^{\theta+1}|).$$

It follows that $x(\varepsilon)$ is absolutely continuous, and we obtain a bound of the form

$$\left|\frac{dx}{d\varepsilon}(\varepsilon)\right| \leqslant \frac{C}{\varepsilon^{\theta+1}} [\theta \varepsilon^{\theta-1} + (\theta+1)\, \varepsilon^\theta] \sim \frac{1}{\varepsilon^2}.$$

(c) Let us examine the corresponding conditions on the parametrization $\varepsilon(t)$. Note that the length of $\varepsilon \to x(\varepsilon)$ may be infinite, so we are in the situation covered in Theorem 3.3.

We may take $\beta(\varepsilon) = \varepsilon$ and $\gamma(\varepsilon) = \varepsilon^2$. Therefore, in order to obtain that $\lim_{t\to+\infty} u(t) = x^* = \mathrm{proj}_{S(P)}\, 0$, the parametrization $t \to \varepsilon(t)$ has to satisfy the following conditions:

$$\text{(i)} \qquad \int_0^{+\infty} \varepsilon(s)\, ds = +\infty$$

$$\text{(ii)} \qquad \lim_{t\to+\infty} \dot{\varepsilon}(t)/\varepsilon(t)^3 = 0.$$

For instance, when taking $\varepsilon(t)=1/t^\alpha$ these conditions hold iff $0<\alpha<1/2$. We summarize the above results in the following statement.

PROPOSITION 3.5. *Let $f_0\colon H\to\mathbb{R}$ be a Lipschitz continuous function, and $C\subset H$ a closed convex subset of H. For any $u_0\in C$, for all $\theta>1$ and $0<\alpha<1/2$, the solution of*

$$\begin{cases} \dot u(t)+\nabla f_0(u(t))+t^{\theta\alpha}(u(t)-\mathrm{proj}_C\, u(t))+1/t^\alpha\, u(t)=0 \\ u(0)=u_0 \end{cases}$$

converges as $t\to+\infty$ to x^* the element of minimal norm of $\mathrm{Argmin}_C f_0$.

REMARK. One can replace the term $\mathrm{dist}(x,C)^2$ in the previous example by any function $g(x)^2$ as far as a global estimate of the form $\mathrm{dist}(x,C)\leqslant M\,g(x)$ holds for some constant M. For instance, when $C=\{x\colon Ax\leqslant b\}$ one may take $g(x)^2=\sum_{i=1}^m[(A_ix-b_i)^+]^2$ which has the advantage of being a computable expression.

4. An Alternative Approach: Epi-Convergence and Scaling

Let us now make the connection between the "epiconvergence and scaling method" and the preceding results. A major feature of the epiconvergence and scaling method, as developed in [3], see also Anzellotti and Baldo [1], is that it allows to characterize which point $x^*\in\mathrm{Argmin}\,f$ is obtained as a limit of the sequence $\{x(\varepsilon);\varepsilon\to 0\}$ of solutions of the approximate problems. Let us recall the main lines of this method.

The first step is to prove that the Mosco-epilimit

$$f=\operatorname*{epi-lim}_{\varepsilon\to 0} f(\cdot,\varepsilon)$$

exists. This is a natural assumption ensuring the convergence of the corresponding minimization problems but, in general, besides the convergence of the infimal values it only gives that every limit point x^* of the sequence $\{x(\varepsilon);\varepsilon\to 0\}$ minimizes f.

Let us now take advantage of the strong monotonicity assumption

$$\langle\partial f(x,\varepsilon)-\partial f(y,\varepsilon),x-y\rangle\geqslant\beta(\varepsilon)\,|x-y|^2$$

and rescale the minimization problem (P_ε) as follows: let

$$h(x,\varepsilon):=\frac{1}{\beta(\varepsilon)}[f(x,\varepsilon)-\inf f]$$

so that the point $x(\varepsilon)$ satisfies

$$h(x(\varepsilon), \varepsilon) = \min\{h(x, \varepsilon) : x \in H\}. \qquad (P_\varepsilon)$$

The second step is to prove that the Mosco-epilimit

$$h := \text{epi-lim}_{\varepsilon \to 0} h(\cdot, \varepsilon)$$

exists and is proper. In this case x^* is precisely the unique minimizer of h on $S(P) = \operatorname{Argmin} f$. Indeed, since

$$\langle \partial h(x, \varepsilon) - \partial h(y, \varepsilon), x - y \rangle \geqslant |x - y|^2$$

and since the Mosco-epiconvergence of a family of convex functions implies the graph-convergence of their subdifferentials (see [2]) it follows that

$$\langle \partial h(x) - \partial h(y), x - y \rangle \geqslant |x - y|^2.$$

Hence h has a unique minimizer, which belongs to $S(P)$ since $h \equiv +\infty$ outside of $S(P) = \operatorname{Argmin} f$.

Let us assume from now on that we have been able to prove the Epi-Scaling property $h = \text{epi-lim}_{\varepsilon \to 0} h(\cdot, \varepsilon)$, and show how the assumptions of Theorems 3.2 and 3.3 can be naturally interpreted in this setting.

We rescale the dynamical system

$$\begin{cases} \dot{u}(t) + \partial f(u(t), \varepsilon(t)) \ni 0 \\ u(0) = u_0 \end{cases}$$

as follows

$$\begin{cases} \dfrac{1}{\beta(\varepsilon(t))} \dot{u}(t) + \partial \left[\dfrac{1}{\beta(\varepsilon(t))} f(\cdot, \varepsilon(t)) \right] (u(t)) \ni 0 \\ u(0) = u_0. \end{cases} \qquad (8)$$

Introducing a change of variables $t = \tau(s)$, the new function $v(s) := u(\tau(s))$ satisfies $\dot{v}(s) = \dot{u}(\tau(s))\, \dot{\tau}(s)$, so that we are naturally led to choose $\tau(s)$ such that

$$\dot{\tau}(s) = \frac{1}{\beta(\varepsilon(\tau(s)))} \qquad (9)$$

in such a way that (8) is transformed into

$$\begin{cases} \dot{v}(s) + \partial h(v(s), \varepsilon(\tau(s))) \ni 0 \\ v(0) = u_0. \end{cases}$$

To find the appropriate change of variables let $E(t) = \int_0^t \beta(\varepsilon(s))\, ds$ so that (9) is equivalent to

$$\dot{E}(\tau(s))\, \dot{\tau}(s) = 1$$

so that we obtain $E(\tau(s)) = s$. Clearly E is a strictly increasing function from $[0, +\infty[$ onto $[0, E_\infty[$ where $E_\infty = \int_0^{+\infty} \beta(\varepsilon(s))\, ds$. In order that E realize a bijection from $\mathbb{R}^+$ onto $\mathbb{R}^+$, we need

$$\int_0^{+\infty} \beta(\varepsilon(s))\, ds = +\infty$$

which is precisely the basic assumption of Theorems 3.2 and 3.3. In such a case we have $\tau(s) = E^{-1}(s)$ with $\lim_{s \to +\infty} \tau(s) = +\infty$, and then

$$\lim_{t \to +\infty} u(t) \text{ exists} \Leftrightarrow \lim_{s \to +\infty} v(s) \text{ exists}$$

in which case both limits coincide.

So we are reduced to the following "standard" situation: denoting for simplicity $h(\cdot, s) = h(\cdot, \varepsilon(\tau(s)))$, we introduce the "Renormalized parabolic system"

$$\begin{cases} \dot{v}(s) + \partial h(v(s), s) \ni 0 \\ v(0) = u_0 \end{cases} \tag{10}$$

with

$$\begin{cases} h(\cdot, s) \to h \text{ Mosco-epiconvergence as } s \to +\infty \\ \langle \partial h(x, s) - \partial h(y, s), x - y \rangle \geqslant |x - y|^2 \text{ for all } s \geqslant 0, \text{ all } x, y \in H. \end{cases}$$

When $h(\cdot, s) \equiv h$ this is the evolution governed by a strongly monotone maximal monotone operator. It is a classical result that $\lim_{s \to +\infty} v(s) = \partial h^{-1}(0)$. Moreover, this asymptotic behavior is stable with respect to perturbation of the dynamics (see Brézis [8]). It is then natural to obtain that for the perturbed problem (10) we still have

$$\lim_{t \to +\infty} u(t) = \lim_{s \to +\infty} v(s) = \partial h^{-1}(0) = x^*.$$

This is what we now consider and make precise. Let us denote for simplicity $x(s) = x(\varepsilon(\tau(s)))$ the unique minimizer of $h(\cdot, s)$ and let

$$\psi(s) := \tfrac{1}{2}\, |v(s) - x(s)|^2.$$

We have

$$\dot{\psi}(s) = \langle \dot{v}(s) - \dot{x}(s), v(s) - x(s) \rangle = \langle \dot{v}(s), v(s) - x(s) \rangle - \langle \dot{x}(s), v(s) - x(s) \rangle.$$

Since a.e. we have $-\dot{v}(s) \in \partial h(v(s), s)$ and $0 \in \partial h(x(s), s)$, we deduce

$$\dot{\psi}(s) \leqslant - |v(s) - x(s)|^2 + |\dot{x}(s)| \, |v(s) - x(s)|$$

hence

$$\dot{\psi}(s) + 2\psi(s) \leqslant |\dot{x}(s)| \sqrt{2\psi(s)}.$$

Multiplying by e^{2s} we get

$$\frac{d}{ds} [e^{2s}\psi(s)] \leqslant \sqrt{2} e^s \, |\dot{x}(s)| \sqrt{e^{2s}\psi(s)}$$

and from Lemma 3.1

$$e^s \sqrt{\psi(s)} \leqslant e^{s_0} \sqrt{\psi(s_0)} + \frac{1}{\sqrt{2}} \int_{s_0}^{s} e^v \, |\dot{x}(v)| \, dv$$

so that

$$\sqrt{\psi(s)} \leqslant e^{-(s - s_0)} \sqrt{\psi(s_0)} + \frac{1}{\sqrt{2}} \int_{s_0}^{s} e^{-(s - v)} \, |\dot{x}(v)| \, dv.$$

Let us analyze the conditions for convergence.

(a) If $\lim_{s \to +\infty} |\dot{x}(s)| = 0$, then

$$\sqrt{\psi(s)} \leqslant e^{-(s - s_0)} \sqrt{\psi(s_0)} + \frac{1}{\sqrt{2}} (\sup_{v \geqslant s_0} |\dot{x}(v)|)(1 - e^{-(s - s_0)}).$$

Hence, s_0 being fixed we get

$$\limsup_{s \to +\infty} \sqrt{\psi(s)} \leqslant \frac{1}{\sqrt{2}} \sup_{v \geqslant s_0} |\dot{x}(v)|.$$

Letting $s_0 \to +\infty$ and using the assumption $\lim_{s \to +\infty} |\dot{x}(s)| = 0$ we obtain the conclusion: $\lim_{s \to +\infty} \psi(s) = 0$ so that $\lim v(s) = \lim u(t) = x^*$.

Let us consider in more detail the assumption $\lim |\dot{x}(s)| = 0$. We have

$$\dot{x}(s) = \frac{dx}{d\varepsilon} (\varepsilon(\tau(s))) \, \dot{\varepsilon}(\tau(s)) \, \dot{\tau}(s) = \frac{dx}{d\varepsilon} (\varepsilon(\tau(s))) \, \frac{\dot{\varepsilon}(\tau(s))}{\beta(\varepsilon(\tau(s)))}.$$

If we assume $|dx/d\varepsilon(\varepsilon)| \leqslant 1/\gamma(\varepsilon)$ as in Theorem 3.3 the assumption will hold when

$$\lim_{s \to +\infty} \frac{\dot{\varepsilon}(s)}{\gamma(\varepsilon(s))\,\beta(\varepsilon(s))} = 0.$$

We recover exactly assumption (ii) of Theorem 3.3.

(b) If we assume $\int_0^\infty |\dot{x}(s)|\,ds < +\infty$ then

$$\sqrt{\psi(s)} \leqslant e^{-(s-s_0)}\sqrt{\psi(s_0)} + \int_{s_0}^{s} |\dot{x}(v)|\,dv$$

and from this we get $\lim_{s \to +\infty} \psi(s) = 0$. This is equivalent to Theorem 3.2.

We conclude that the two approaches presented are essentially equivalent.

5. Appendix

5.1. *Estimation of the rate of convergence of $u(t)$ to x^**

Example. Let us return to the setting of Theorem 3.2 by assuming the following stronger conditions. We suppose the trajectory $x(\varepsilon)$ is Lipschitz

$$\left|\frac{dx}{d\varepsilon}(\varepsilon)\right| \leqslant C$$

so that the finite length assumption holds, and we assume that condition (H) is satisfied with $\beta(\varepsilon) = \varepsilon$.

Let us analyze two cases where condition $\int_0^\infty \beta(\varepsilon(s))\,ds = +\infty$ holds.

Case 1. Taking $\varepsilon(t) = (1-\alpha)(1+t)^{-\alpha}$ with $0 < \alpha < 1$, we have $E(t) = (1+t)^{1-\alpha}$ and we get the estimate (see (4)),

$$\sqrt{\varphi(t)} \leqslant \exp\left[(1+t_0)^{1-\alpha} - (1+t)^{1-\alpha}\right]\sqrt{\varphi(t_0)} + \frac{(1-\alpha)C}{\sqrt{2}}\left[(1+t_0)^{-\alpha} - (1+t)^{-\alpha}\right].$$

With $t_0 = (t-1)/2$ we obtain

$$\sqrt{\varphi(t)} \leqslant \exp\left[-\left(1 - \frac{1}{2^{1-\alpha}}\right)(1+t)^{1-\alpha}\right]\sqrt{\varphi((t-1)/2)} + \frac{(1-\alpha)C(2^\alpha - 1)}{\sqrt{2}}\,\frac{1}{(1+t)^\alpha}.$$

Since $\sqrt{\varphi((t-1)/2)}$ tends to 0 and the exponential decays faster than $1/t^\alpha$, we get that asymptotically

$$\sqrt{\varphi(t)} \leqslant \frac{K}{t^\alpha}$$

for an appropriate constant K and t large.

Case 2. Taking $\varepsilon(t)=1/(1+t)$ we have $E(t)=\ln(1+t)$ and from (3) we obtain the estimate

$$(1+t)\sqrt{\varphi(t)} \leqslant \sqrt{\varphi(0)} + \frac{C}{\sqrt{2}} \int_0^t \frac{1+s}{(1+s)^2}\, ds = \sqrt{\varphi(0)} + \frac{C}{\sqrt{2}} \ln(1+t)$$

so that asymptotically we get

$$\sqrt{\varphi(t)} \leqslant K \frac{\ln(t)}{t}$$

for some constant K and t large.

5.2. Open Problems

The results presented in this paper open a number of questions, some of which are stated below:

1. Precise the asymptotic behavior of the $(DADA)$ trajectory.
   (a) Does $\dot{u}(t) \to 0$ as $t \to +\infty$?
   (b) Are the curves $t \to x(\varepsilon(t))$ and $t \to u(t)$ asymptote?
2. Does $\lim u(t)$ exist for an arbitrary parametrization $\varepsilon(t)$?
3. What is the structure of the set of all limit points $u_\infty = \lim u(t)$ corresponding to different parametrizations $\varepsilon(t)$?

References

1. G. Anzellotti and S. Baldo, Asymptotic development by Γ-convergence, *Appl. Math. Optim.* **27** (1993), 105–123.
2. H. Attouch, "Variational Convergence for Functions and Operators," Applicable Maths. Series, Pitman, London, 1984.
3. H. Attouch, "Viscosity Solutions of Minimization Problems," Séminaire d'Analyse Convexe, Université de Montpellier, exposé 8, 1992. To appear in SIAM Journal of Optimization.

4. H. Attouch and A. Damlamian, Strong solutions for parabolic variational inequalities, *J. Nonlinear Analysis* **2**(3) (1978), 329–353
5. H. Attouch and R. Wets, Quantitative stability of variational systems: I. The epigraphical distance, *Trans. Amer. Math. Soc.* **328**(2) (1991), 695–729.
6. J. P. Aubin, Lipschitz behavior of solutions to convex minimization problems, *Math. Operation Research* **9** (1984), 87–111.
7. G. Auchmuty, Saddle points and existence-uniqueness for evolution equations, *Diff. and Integral Eq.* **6**(5) (1993), 1161–1171.
8. H. Brézis, "Opérateurs Maximaux Monotones," North-Holland, Mathematics Studies 5, 1973.
9. H. Brézis, "Asymptotic Behavior of Some Evolution Systems," Nonlinear Evolution Equations, Academic Press, 1978.
10. R. E. Brück, Asymptotic convergence of nonlinear contraction semigroups in Hilbert space, *J. of Func. Anal.* **18** (1975), 15–26.
11. R. Cominetti, Asymptotic convergence of Euler's method for the exponential penalty in linear programming, *J. of Convex Analysis* **2**(1/2) (1995), 145–152.
12. R. Cominetti and J. San Martín, Trajectory analysis for the exponential penalty method in linear programming, *Math. Programming* **67** (1994), 169–187.
13. Dontchev and T. Zolezzi, "Well posed optimization problems," Lecture Notes in Mathematics, Vol. 1543, Springer-Verlag, 1993.
14. H. Furuya, K. Miyashiba, and N. Kenmochi, Asymptotic behavior of solutions to a class of nonlinear evolution equations, *J. Differential Equations* **62** (1986), 73–94.
15. N. Kenmochi, Solvability of nonlinear evolution equations with time-dependent constraints and applications, *Bull. Fac. Ed. Chiba Univ.* **30** (1981), 1–87.
16. B. Lemaire, "Coupling Optimization Methods and Variational Convergence," Trends in Mathematical Optimization, International series on numerical Mathematics, Vol. 84(C), Birkhäuser Verlag, 163–179, 1988.
17. B. Lemaire, "An Asymptotical Variational Principle Associated with some Differential Inclusions, Continuous and Discrete Case," Conference colloque franco-vietnamien, ICAA, Hanoi, 1993.
18. U. Mosco, Convergence of convex sets and solutions of variational inequalities, *Adv. in Math.* **3** (1969), 510–585.
19. A. Moudafi, Coupling proximal methods and variational convergence, *ZOR-Methods and Models of Operations Research* **38** (1993), 269–280.
20. D. Tataru, Time dependent *m*-accretive operators generating differential evolutions, *Differential and Integral Equations* **4**(1) (1991), 137–150.
21. D. Torralba, "Convergence épigraphique et changements d'échelle en analyse variationnelle et optimisation," Thesis, Université de Montpellier II, 1966.
22. P. Tossings, "Sur les Zéros des opérateurs maximaux monotones," Thèse, Université de Liège, 1990.
23. A. Tikhonov and V. Arsenine, "Méthodes de Résolution de Problèmes Mal Posés," MIR, 1974.