

# A DYNAMICAL SYSTEM ASSOCIATED WITH NEWTON'S METHOD FOR PARAMETRIC APPROXIMATIONS OF CONVEX MINIMIZATION PROBLEMS\*

F. ALVAREZ D.<sup>†</sup> AND J.M. PÉREZ C.<sup>‡</sup>

**Abstract.** We study the existence and asymptotic convergence when  $t \rightarrow +\infty$  for the trajectories generated by

$$\nabla^2 f(u(t), \epsilon(t))\dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \nabla f(u(t), \epsilon(t)) = 0$$

where  $\{f(\cdot, \epsilon)\}_{\epsilon > 0}$  is a parametric family of convex functions which approximates a given convex function  $f$  we want to minimize, and  $\epsilon(t)$  is a parametrization such that  $\epsilon(t) \rightarrow 0$  when  $t \rightarrow +\infty$ . This method is obtained from the following variational characterization of Newton's method

$$(P_t^\epsilon) \quad u(t) \in \operatorname{Argmin}\{f(x, \epsilon(t)) - e^{-t} \langle \nabla f(u_0, \epsilon_0), x \rangle : x \in H\}$$

where  $H$  is a real Hilbert space. We find conditions on the approximating family  $f(\cdot, \epsilon)$  and the parametrization  $\epsilon(t)$  to ensure the norm convergence of the solution trajectories  $u(t)$  towards a particular minimizer of  $f$ . The asymptotic estimates obtained allow us to study the rate of convergence as well. The results are illustrated through some applications to barrier and penalty methods for linear programming, and to viscosity methods for an abstract non-coercive variational problem. Comparisons with the steepest descent method are also provided.

**Key words.** Convex minimization, approximate methods, continuous methods, evolution equations, existence, optimal trajectory, asymptotic analysis.

**AMS subject classifications.** 34G20, 34A12, 34D05, 90C25, 90C31

**1. Introduction.** Newton's method for solving a smooth optimization problem on a real Hilbert space  $H$

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

is based on sequentially minimizing the quadratic expansion of  $f$ , that is to say, the method recursively computes

$$(1) \quad u_{k+1} = u_k - \nabla^2 f(u_k)^{-1} \nabla f(u_k).$$

It is well known that under suitable assumptions ( $f \in C^3$  and  $\nabla^2 f(\hat{x})$  positive definite at a local minimizer  $\hat{x}$ ) and starting sufficiently close to  $\hat{x}$ , the iterative scheme (1) is well defined and converges to  $\hat{x}$  quadratically:  $|u_{k+1} - \hat{x}| \leq C|u_k - \hat{x}|^2$ . However, when the method is far from the solution there is no guarantee of convergence and one must introduce an appropriate *step size*  $\lambda_k > 0$ , leading to

$$\frac{u_{k+1} - u_k}{\lambda_k} = -\nabla^2 f(u_k)^{-1} \nabla f(u_k),$$

which can be interpreted as an explicit discrete scheme for the *continuous Newton's method*,

$$(2) \quad \nabla^2 f(u(t))\dot{u}(t) + \nabla f(u(t)) = 0.$$

---

\*Partially supported by Fondecyt 1961131.

<sup>†</sup>Supported by the CEE grant CII-CT94-01115. Universidad de Chile, Casilla 170/3 Correo 3, Santiago, Chile. e-mail: falvarez@dim.uchile.cl

<sup>‡</sup>Supported by funds from Fundación Andes. Pedro Montt 147, Peñablanca, Villa Alemana, Chile. e-mail: jmperez@dim.uchile.cl

It is easy to check that every solution of (2) satisfies

$$\frac{d}{dt} [\nabla f(u(t))] = -\nabla f(u(t))$$

and therefore

$$(3) \quad \nabla f(u(t)) - e^{-t} \nabla f(u_0) = 0.$$

When the function  $f$  is convex, the latter is equivalent to

$$(4) \quad u(t) \in \operatorname{Argmin}\{f(x) - e^{-t} \langle \nabla f(u_0), x \rangle : x \in H\}$$

providing a variational characterization for the Newton trajectory  $u(t)$  which does not require  $f$  to be of class  $C^2$  nor to have a nonsingular Hessian, and which makes sense even for non smooth  $f$ 's (replace  $\nabla f(u_0)$  with any  $u_0^* \in \partial f(u_0)$ ). We consider this non-smooth extension of Newton's method in §3, proving that  $u(t)$  is a descent trajectory for  $f$  which is defined for all  $t \geq 0$  (an interesting result even in the smooth case). Moreover, when  $f$  is strongly convex, we establish an exponential rate of convergence for  $u(t)$  towards the unique minimizer of  $f$ .

When the function  $f$  is not strongly convex, and also when it presents an irregular behavior because of non-smoothness or when it takes the value  $+\infty$  because of implicit constraints, a classical approach is to replace  $f$  by a sequence of better behaved parametric approximations (penalty methods, Tikhonov regularization, viscosity methods, etc.). More precisely, consider the one-parameter family of problems

$$(P^\epsilon) \quad \min\{f(x, \epsilon) : x \in H\}$$

where for each  $\epsilon > 0$  the function  $f(\cdot, \epsilon)$  is a closed proper convex approximate converging to  $f$  when  $\epsilon \rightarrow 0$ . Numerical methods based on such parametric approximation schemes consist in solving  $(P^\epsilon)$  for an appropriate sequence  $\epsilon_k$  converging to 0. When  $f(\cdot, \epsilon)$  is sufficiently smooth, a common practice is to use Newton's method for solving  $(P^{\epsilon_k})$  up to a prescribed accuracy and then proceed to the next iteration with parameter  $\epsilon_{k+1}$ .

Because of the usual increasing ill-conditioning of the Hessian matrix of the approximating function  $f(\cdot, \epsilon)$ , the convergence of the overall method requires a careful selection of the starting point for the Newton's iterations as well as the sequence of parameters  $\epsilon_k$ , choices which are strongly interdependent. Unfortunately, no general theory is available in order to guide these choices. With the aim of contributing to this issue, in §4 we consider the question in the simpler setting of a continuous method coupling Newton's method with approximation schemes.

We observe that there is no standard way of coupling Newton's method with an approximation scheme. Possibly, the most straightforward alternative would be to consider the non-autonomous differential equation

$$\nabla^2 f(u(t), \epsilon(t)) \dot{u}(t) + \nabla f(u(t), \epsilon(t)) = 0$$

where the parameter function  $\epsilon(t)$  is *a priori* chosen so that it decreases to 0 when  $t \rightarrow +\infty$ . However, there is no reason for this method to be the most appropriate (not even the most natural) way of efficiently combining Newton's method with an approximation scheme. Taking into account the variational characterization (4) we study in §4 a different generalization, namely

$$(5) \quad u(t) \in \operatorname{Argmin}\{f(x, \epsilon(t)) - e^{-t} \langle \nabla f(u_0, \epsilon_0), x \rangle : x \in H\}.$$

In the smooth case, this approach corresponds to the differential equation

$$\nabla^2 f(u(t), \epsilon(t)) \dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \nabla f(u(t), \epsilon(t)) = 0$$

whose associated vector field combines a Newton direction with an extrapolation term that specifically takes into account the rate of change in the objective function  $t \rightarrow f(\cdot, \epsilon(t))$ .

We study the existence of solutions for (5) as well as their asymptotic behavior. In a general setting, we obtain a weak condition on the parametrization  $\epsilon(t)$  ensuring the asymptotic convergence of the solution trajectories  $u(t)$  towards a particular solution of the original problem (P), giving at the same time an estimate of the rate of convergence. We illustrate the general results through some examples in the setting of barrier and penalty methods for linear programming, as well as for Tikhonov regularization methods.

We compare our results with those obtained for the steepest descent method (see [3]). To this end we include a short discussion of this method in the preliminary section §2, for which we provide a new convergence proof for the non-parametric case.

In order to handle with the problem of the increasing ill-conditioning of Hessian operator of the parametric approximation, we shall see in §5 how in certain cases an appropriate rescaling of the approximate problem may be used to obtain an uniform strongly convex approximate scheme.

## 2. Preliminaries.

**2.1. Basic assumptions.** Throughout this paper we shall consider an abstract optimization problem of the form

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

where  $H$  is a real Hilbert space and  $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$  is a closed proper convex function. The set of optimal solutions of (P) is denoted  $\text{Argmin } f$  and we assume throughout that

$$(H_0) \quad \text{the optimal set } \text{Argmin } f \text{ is nonempty and bounded.}$$

This holds for instance when  $f$  is coercive (*i.e.*  $f$  has bounded level sets). We recall that according to Moreau's theorem, the latter is equivalent to finiteness and continuity of the Fenchel conjugate  $f^*$  at  $x^* = 0$ .

We say that  $f$  is strongly convex if there exists  $\beta > 0$  such that

$$f(x) + \langle x^*, y - x \rangle + \frac{\beta}{2} |y - x|^2 \leq f(y),$$

for all  $x, y \in H$  and  $x^* \in \partial f(x)$ . Equivalently,  $f$  is strongly convex when its sub-differential operator is strongly monotone, that is to say, there exists  $\beta > 0$  such that

$$(H_1) \quad \text{if } x^* \in \partial f(x) \text{ and } y^* \in \partial f(y) \text{ then } \langle x^* - y^*, x - y \rangle \geq \beta |x - y|^2$$

If  $f$  is strongly convex then  $f$  is strictly convex and coercive. In particular, there exists a unique point  $\hat{x}$  solution of (P). A weaker condition is the strong monotonicity over bounded sets,

$$(H_2) \quad \forall K > 0, \exists \beta_K > 0 \text{ s.t. } \forall x, y \in B(0, K), \text{ if } x^* \in \partial f(x) \text{ and } y^* \in \partial f(y) \\ \text{then } \langle x^* - y^*, x - y \rangle \geq \beta_K |x - y|^2$$

We remark that in this case  $f$  is also strictly convex but may fail to satisfy  $(H_0)$ . As a matter of fact, in this case there is no guarantee of existence of an optimal solution. In the finite dimensional case, if  $f \in C^2(\mathbb{R}^n; \mathbb{R})$  is such that the Hessian matrix  $\nabla^2 f(x)$  is positive definite for all  $x \in \mathbb{R}^n$ , then  $(H_2)$  is satisfied and the parameter  $\beta_K$  is a lower bound for the eigenvalues of the matrix  $\nabla^2 f(x)$  over  $B(0, K)$ .

Together with the original problem  $(P)$  we shall consider the parametric family

$$(P^\epsilon) \quad \min\{f(x, \epsilon) : x \in H\}$$

where for each  $\epsilon > 0$ ,  $f(\cdot, \epsilon)$  is a closed proper convex function approximating  $f$ . Very often the function  $f(\cdot, \epsilon)$  may be chosen so that it enjoys good properties such as coercivity and strong convexity and therefore the approximate problem  $(P^\epsilon)$  is well posed in existence and uniqueness. We shall assume throughout that  $(P^\epsilon)$  has a unique solution which we denote  $x(\epsilon)$ , and more precisely we suppose that

$$(H_3) \quad \begin{array}{l} \text{there exists a unique path } x(\epsilon) \text{ of optimal solutions of } (P^\epsilon) \text{ which} \\ \text{converge in norm towards an optimal point } \hat{x} \in \text{Argmin } f \text{ when } \epsilon \rightarrow 0. \end{array}$$

Variational notions of convergence for sequences of functions provide an appropriate setting for studying the convergence of the optimal solutions of approximate schemes as the previous one. For instance, assuming that the following Mosco-epi-limit holds

$$f = \text{epi} - \lim_{\epsilon \rightarrow 0} f(\cdot, \epsilon),$$

and under appropriate compactness assumptions, the optimal value  $\min f(\cdot, \epsilon)$  converges to  $\min f$  and every limit point of the optimal path  $\{x(\epsilon) : \epsilon \rightarrow 0\}$  is an optimal solution of  $(P)$ . The latter does not ensure the convergence of the optimal path unless  $(P)$  has a unique solution. However, with a more specific analysis, the validity of  $(H_3)$  has been established for many approximate schemes: for viscosity approximation methods see Tikhonov and Arsenine [16], Attouch [1] and references therein; for interior-point and penalty methods in convex and linear programming see Megiddo [14], Gonzaga [12], Cominetti and San Martin [9], Fiacco [10] and Auslender, Cominetti and Haddou [5]; for an abstract approach based on epi-convergence and scaling see Attouch [2] and Torralba [17], as well as §5. In all these cases, the limit point  $\hat{x}$  has remarkable variational properties which make its computation interesting by its own.

**2.2. Steepest descent method.** For the sake of comparison with the results to be presented for the Newton method, we briefly recall in this section the convergence properties of the steepest descent method for the problem  $(P)$  as well as its coupling with the approximate scheme  $(P^\epsilon)$ .

Let us begin with the steepest descent differential inclusion

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

An absolutely continuous function  $u : [0, +\infty) \rightarrow H$  is a solution of  $(SD; u_0)$  if  $u(0) = u_0$  and the above inclusion is satisfied almost everywhere in  $[0, +\infty)$ . It is a classical result that when the minimum of  $f$  is attained, every solution  $u(t)$  of  $(SD; u_0)$ , weakly converges to an element  $u_\infty \in \text{Argmin } f$  as  $t \rightarrow +\infty$  (see Brézis [6, 7], Bruck [8]). When  $(P)$  has multiple optimal solutions, the limit point may

depend on the initial condition  $u_0$  and it may be difficult to characterize (see Lemaire [13] for results along these lines).

When  $f$  is strongly convex, it is possible to estimate the rate of convergence of the trajectory  $u(t)$  (see Brézis [6, Theorem 3.9]). In the following proposition, we present a slight variant of this result which assumes only the strong convexity over bounded sets, and which also provides an estimate for the rate of convergence of the function values.

**PROPOSITION 2.1.** *Let  $f$  be a coercive convex function satisfying  $(H_2)$ , and let  $\hat{x}$  be the unique minimizer of  $f$ . If  $u : [0, +\infty) \rightarrow H$  is a solution of  $(SD; u_0)$ , then there exists  $\beta > 0$  such that for all  $t > 0$ ,*

$$f(u(t)) - \min f \leq (f(u_0) - \min f)e^{-\beta t},$$

and

$$|u(t) - \hat{x}| \leq \sqrt{\frac{2}{\beta}(f(u_0) - \min f)}e^{-\frac{\beta}{2}t}.$$

*Proof.* Let  $v(t) := f(u(t)) - \min f$ . It is known [6, Theorem 3.6 and lemma 3.3] that when

$$u_0 \in \text{dom}f = \{x \in H : f(x) < +\infty\}$$

then  $f(u(t))$  is an absolutely continuous function. Moreover, we have

$$\frac{d}{dt}f(u(t)) = \langle x^*, \dot{u}(t) \rangle$$

for all  $x^* \in \partial f(u(t))$  and almost everywhere in  $[0, +\infty)$ . In particular we obtain  $\dot{v}(t) = -|\dot{u}(t)|^2$ , so that  $v(t)$  is non-increasing and since  $f$  is coercive the trajectory  $u(t)$  stays bounded as  $t \rightarrow +\infty$ .

Let us assume that for all  $t \geq 0$  we have  $u(t) \neq \hat{x}$  (otherwise the result is obvious). Then, using the convexity inequality  $\langle x^*, \hat{x} - u(t) \rangle \leq f(\hat{x}) - f(u(t))$  together with

$$|x^*| \geq \frac{1}{|u(t) - \hat{x}|} \langle x^*, u(t) - \hat{x} \rangle,$$

it follows that

$$\dot{v}(t) \leq \frac{|x^*|}{|u(t) - \hat{x}|} [f(\hat{x}) - f(u(t))] = -\frac{|x^*|}{|u(t) - \hat{x}|} v(t),$$

when  $x^* \in \partial f(u(t))$ . Taking a suitable constant  $K$  such that  $|u(t)| \leq K$  for all  $t \geq 0$  and  $|\hat{x}| \leq K$ , using  $(H_2)$  and the facts that  $0 \in \partial f(\hat{x})$  and  $v(t) \geq 0$ , we obtain  $\dot{v}(t) \leq -\beta v(t)$  with  $\beta := \beta_K$ . The conclusions follow at once by integrating the inequality  $\dot{v}(t) \leq -\beta v(t)$ , and then using the strong convexity inequality  $f(\hat{x}) + \frac{\beta}{2}|u(t) - \hat{x}|^2 \leq f(u(t))$ .  $\square$

**REMARK.** Using a special notion of weak solution for  $(SD; u_0)$  (Brézis [6]), it is possible to take in the previous proposition  $u_0 \in \text{Adh}(\text{dom}f)$ .

This result shows that when  $f$  is strongly convex, the steepest descent method has a very good behavior, converging towards the solution of  $(P)$  at an exponential rate. However, these assumptions do not hold in many interesting situations, particularly

when the original problem ( $P$ ) has multiple solutions or when the function  $f$  is not regular or finite. This led Attouch and Cominetti [3] to consider the coupling of the steepest descent method with approximation schemes in the following way

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

where the function  $\epsilon : [0, +\infty) \rightarrow \mathbb{R}_+$  is strictly positive and decreasing to 0 with  $t \rightarrow +\infty$ .

Roughly speaking, if  $\epsilon(t)$  converges sufficiently *fast* to 0, one expects that the solution trajectories of ( $DADA$ ) behave like the solutions of ( $SD$ ) which can be considered as a limit equation for ( $DADA$ ) (with  $\epsilon(t) = 0$ ). As a matter of fact, under suitable hypothesis, it was proved by Furuya, Miyashiba and Kenmochi [11] that in this case the trajectory  $u(t)$  converges towards an optimal solution of ( $P$ ).

On the other hand, it has been proved by Attouch and Cominetti [3] that when  $\epsilon(t)$  converges to 0 sufficiently *slow*, then  $u(t)$  asymptotically approaches the optimal trajectory  $x(\epsilon(t))$  and therefore it is attracted towards  $\hat{x}$ . The speed of convergence of  $\epsilon(t)$  is measured in terms of the strong convexity parameter  $\beta(\epsilon)$  of  $f(\cdot, \epsilon)$ , namely, the condition

$$(H_4) \quad \langle \partial f(x, \epsilon) - \partial f(y, \epsilon), x - y \rangle \geq \beta(\epsilon)|x - y|^2$$

implies that (see [3] for details) for all  $t \geq t_0 \geq 0$

$$(6) \quad |u(t) - x(\epsilon(t))| \leq C e^{-(E(t) - E(t_0))} - e^{-E(t)} \int_{t_0}^t e^{E(s)} \left| \frac{dx}{d\epsilon}(\epsilon(s)) \right| \dot{\epsilon}(s) ds$$

where  $C := |u_0 - x(\epsilon(t_0))|$  and  $E(t) = \int_0^t \beta(\epsilon(s)) ds$ . It follows that  $u(t)$  converges strongly towards  $\hat{x}$  whenever

$$(7) \quad \int_0^{+\infty} \beta(\epsilon(s)) ds = +\infty$$

and either (a) or (b) below hold:

$$(a) \quad \int_0^{\epsilon_0} \left| \frac{dx}{d\epsilon} \right| ds < +\infty,$$

$$(b) \quad \lim_{t \rightarrow +\infty} \left| \frac{dx}{d\epsilon}(\epsilon(t)) \right| \frac{\dot{\epsilon}(t)}{\beta(\epsilon(t))} = 0.$$

REMARK. When  $f(\cdot, \epsilon) = f(\cdot)$  for all  $\epsilon > 0$  then ( $DADA$ ) correspond to ( $SD$ ) with  $f$  strongly convex. In this case, the estimate (6) implies

$$|u(t) - \hat{x}| \leq |x_0 - \hat{x}| e^{-\beta t}$$

which is in fact the estimate obtained by Brézis [6].

It is quite natural to investigate the coupling of approximation schemes with other descent methods, particularly a second order method such as Newton's. In the next section we study a variational characterization of the Newton trajectory which leads naturally to a dynamical method coupling Newton's method with approximation schemes.

**3. A variational characterization of the Newton trajectory.** Let us go back to the original problem  $(P)$  and consider the following family of problems

$$(P_t) \quad u(t) \in \operatorname{Argmin}\{f(x) - e^{-t}\langle x_0^*, x \rangle : x \in H\},$$

where  $x_0^* \in \partial f(u_0)$  is fixed. The optimality condition for  $(P_t)$  is

$$\partial f(u(t)) - e^{-t}x_0^* \ni 0$$

which is the non-smooth version of equation (3), so that  $(P_t)$  can be considered as a non-smooth extension of the Newton method. As a matter of fact, we will prove in §3.1 that under appropriate assumptions the trajectory  $u(t)$  coincides with the Newton trajectory.

**PROPOSITION 3.1.** *If  $f$  is coercive then the optimal set of  $(P_t)$  is non-empty and bounded for all  $t > 0$ .*

*Proof.* Consider the function  $g(x, t) := f(x) - e^{-t}\langle x_0^*, x \rangle$ . For all  $x^* \in H$  the Fenchel conjugate of  $g(\cdot, t)$  is given by

$$g^*(x^*, t) = \sup_{x \in H} \{\langle x^*, x \rangle - g(x, t)\} = f^*(x^* + e^{-t}x_0^*).$$

Since  $x_0^* \in \partial f(u_0)$  it follows that  $f^*$  is finite at  $x_0^*$ . On the other hand, our assumption on  $f$  and Moreau's theorem ensure that  $f^*$  is finite and continuous at 0, and therefore the same holds at each point in the segment  $[0, x_0^*]$ . Hence  $g^*(\cdot, t)$  is finite and continuous at 0 for all  $t > 0$  and the conclusion follows by using Moreau's theorem once again.  $\square$

Since  $u_0$  solves  $(P_0)$ , applying the last proposition one may select an optimal trajectory  $u : [0, +\infty) \rightarrow H$  with  $u(0) = u_0$ . If the function  $f$  is strictly convex, then this trajectory  $u(t)$  is unique. Having established conditions for the existence and uniqueness of  $u(t)$ , the question that comes out is its behavior when  $t \rightarrow +\infty$ .

**PROPOSITION 3.2.** *Let  $u : [0, +\infty) \rightarrow H$  be a trajectory of solutions of  $(P_t)$ . Then  $u(\cdot)$  is a descent trajectory for  $f$ , that is to say*

$$\forall t \geq 0, \forall h \geq 0 \quad f(u(t+h)) \leq f(u(t)).$$

*In particular, if  $f$  is coercive then  $u(t)$  stays bounded as  $t \rightarrow +\infty$ .*

*Proof.* Since  $u(t+h)$  is an optimal solution for  $(P_{t+h})$ , we have

$$f(u(t+h)) - e^{-(t+h)}\langle x_0^*, u(t+h) \rangle \leq f(u(t)) - e^{-(t+h)}\langle x_0^*, u(t) \rangle$$

and therefore

$$(8) \quad f(u(t+h)) - f(u(t)) \leq e^{-(t+h)}\langle x_0^*, u(t+h) - u(t) \rangle.$$

On the other hand, the optimality of  $u(t)$  for  $(P_t)$  implies that  $e^{-t}x_0^* \in \partial f(u(t))$  so that

$$e^{-t}\langle x_0^*, u(t+h) - u(t) \rangle \leq f(u(t+h)) - f(u(t))$$

which combined with (8) gives

$$f(u(t+h)) - f(u(t)) \leq e^{-h}[f(u(t+h)) - f(u(t))]$$

and therefore  $f(u(t+h)) - f(u(t)) \leq 0$ .  $\square$

With the previous results we may now state

**PROPOSITION 3.3.** *Assume that  $f$  is coercive and let  $u(t)$  be an optimal trajectory for  $(P_t)$ . Then*

$$\lim_{t \rightarrow +\infty} f(u(t)) = \min f$$

and every weak limit point of the trajectory  $\{u(t) : t \rightarrow +\infty\}$  belongs to the optimal set  $\text{Argmin } f$ .

Moreover, if  $f$  satisfies  $(H_2)$  then there exists a constant  $C > 0$  such that

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

where  $\hat{x}$  is the unique optimal solution of  $(P)$ .

*Proof.* The optimality condition  $\partial f(u(t)) - e^{-t}x_0^* \ni 0$  can be equivalently stated as

$$(9) \quad f(u(t)) + f^*(e^{-t}x_0^*) = e^{-t}\langle x_0^*, u(t) \rangle.$$

By Moreau's theorem,  $f^*$  is finite and continuous at 0, while the previous proposition and the coercivity of  $f$  implies that  $u(t)$  stays bounded as  $t \rightarrow +\infty$ . Passing to the limit in (9) we get

$$\lim_{t \rightarrow +\infty} f(u(t)) = -f^*(0) = \inf_{x \in H} f(x).$$

Since  $u(t)$  stays bounded as  $t \rightarrow +\infty$  it has weak limit points, and the weak lower semi-continuity of  $f$  together with the convergence of the optimal values, imply that these limit points minimize  $f$ .

When  $f$  satisfies  $(H_2)$ , recalling that  $e^{-t}x_0^* \in \partial f(u(t))$  and  $0 \in \partial f(\hat{x})$ , we obtain

$$\langle e^{-t}x_0^*, u(t) - \hat{x} \rangle \geq \beta_K |u(t) - \hat{x}|^2$$

for a suitable constant  $K > 0$ , and therefore

$$|u(t) - \hat{x}| \leq \frac{1}{\beta_K} e^{-t} |x_0^*|,$$

and the conclusion follows with  $C = |x_0^*|/\beta_K$ .  $\square$

**3.1. Equivalence with the continuous Newton's method.** When motivating the study of the family of the problems  $(P_t)_{t \geq 0}$  in §1, we started from the continuous Newton method

$$(CN; u_0) \quad \begin{cases} \nabla^2 f(u(t))\dot{u}(t) + \nabla f(u(t)) = 0 \\ u(0) = u_0, \end{cases}$$

and we observed that every solution was an optimal trajectory for  $(P_t)$ .

The following simple result establishes that the converse holds under suitable differentiability assumptions. Therefore, the previous results extend to the non-smooth setting those obtained for  $(CN; u_0)$  by Aubin and Cellina [4] using the viability theory for differential inclusions.



PROPOSITION 3.4. *Let  $f \in C^2(H; \mathbb{R})$  be strongly convex function. Then a function  $u : [0, +\infty) \rightarrow H$  is an optimal trajectory for  $(P_t)$  if and only if it is the global solution of the differential equation  $(CN; u_0)$ .*

*Proof.* We just prove the “only if” part. Let  $u : [0, +\infty) \rightarrow H$  be the optimal trajectory for  $(P_t)$ , characterized by the optimality condition

$$(10) \quad \nabla f(u(t)) - e^{-t} \nabla f(u_0) = 0.$$

Since  $f \in C^2(H; \mathbb{R})$  and  $\nabla^2 f(x)$  is invertible for all  $x \in H$  (Lax-Milgram theorem) we can apply the Implicit Function Theorem to (10) in order to conclude that  $u(\cdot)$  is differentiable for all  $t \geq 0$ . Differentiating (10) with respect to  $t$ , we obtain

$$\nabla^2 f(u(t)) \dot{u}(t) + e^{-t} \nabla f(u_0) = 0,$$

and using the optimality condition once again we get

$$\nabla^2 f(u(t)) \dot{u}(t) + \nabla f(u(t)) = 0,$$

so that  $u(\cdot)$  is a solution of  $(CN; u_0)$ .  $\square$

COROLLARY 3.5. *Let  $f \in C^2(H; \mathbb{R})$  be strongly convex. Then for all  $u_0 \in H$  there exists a unique  $u : [0, +\infty) \rightarrow H$  solution trajectory of  $(CN; u_0)$ , which is a descent trajectory for  $f$  and satisfies*

$$|u(t) - \hat{x}| \leq C e^{-t}$$

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

**4. Approximation and continuous Newton's method.** We have seen in the previous section that under strong convexity the convergence of Newton's method is very fast. Unfortunately, strong convexity is a drastic restriction which does not hold in many interesting optimization problems, particularly when there exist multiple optimal solutions. A typical and very important case is linear programming.

As mentioned in §1, the main idea to be developed hereafter is to combine the continuous Newton method with approximation schemes in which the original problem  $(P)$  is replaced by a sequence of well-posed strongly convex minimization problems  $\min\{f(x, \epsilon) : x \in H\}$ . Our previous analysis was mainly based on the variational characterization of the Newton trajectory. Hence we are naturally led to consider the following family of optimization problems

$$(P_t^\epsilon) \quad u(t) = \text{Argmin}\{f(x, \epsilon(t)) - e^{-t} \langle x_0^*, x \rangle : x \in H\}$$

where  $x_0^* \in \partial f(u_0, \epsilon_0)$  and the functions  $\epsilon(t)$  and  $f(\cdot, \epsilon)$  are chosen as in §2.1.

Although much of the subsequent analysis remains valid in a non-smooth setting, we are specially interested in obtaining a differential equation equivalent to  $(P_t^\epsilon)$ . Therefore, we shall assume that the approximating function  $f(x, \epsilon)$  is as smooth as required.

The optimality condition for  $(P_t^\epsilon)$  is

$$\nabla f(u(t), \epsilon(t)) - e^{-t} \nabla f(u_0, \epsilon_0) = 0.$$

If the solution trajectory  $u(t)$  exists and is smooth, we can differentiate this stationarity condition in order to obtain

$$\nabla^2 f(u(t), \epsilon(t)) \dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + e^{-t} \nabla f(u_0, \epsilon_0) = 0,$$

from which we derive the differential problem

$$(ACN) \quad \nabla^2 f(u(t), \epsilon(t)) \dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \nabla f(u(t), \epsilon(t)) = 0$$

to which we shall refer as the *Approximate Continuous Newton method*. This dynamical system combines a Newton correction term leading the trajectory  $u(t)$  towards the exact minimizer  $x(\epsilon(t))$  of  $f(\cdot, \epsilon(t))$ , with an extrapolation term anticipating changes in the target point  $x(\epsilon(t))$ .

REMARK. An alternative to  $(P_t^\epsilon)$  would be to introduce a one-to-one differentiable function  $\eta : [0, +\infty) \rightarrow [0, +\infty)$  as follows

$$u(t) = \operatorname{Argmin}\{f(x, \epsilon(t)) - e^{-\eta(t)} \langle \nabla f(u_0, \epsilon_0), x \rangle : x \in H\},$$

which leads to the differential equation

$$(11) \quad \nabla^2 f(u(t), \epsilon(t)) \dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \dot{\eta}(t) \nabla f(u(t), \epsilon(t)) = 0.$$

Introducing the change of variables  $t = \theta(s)$  with  $\theta = \eta^{-1}$ , and defining  $v(s) := u(\theta(s))$ , (11) is transformed into

$$\nabla^2 f(v(s), \epsilon(\theta(s))) \dot{v}(s) + \frac{d}{ds}(\epsilon(\theta(s))) \frac{\partial^2 f}{\partial \epsilon \partial x}(v(s), \epsilon(\theta(s))) + \nabla f(v(s), \epsilon(\theta(s))) = 0,$$

which is (ACN) with a reparametrized approximating scheme. We will use a reparametrization of this kind in §5.

**4.1. Existence and asymptotic behavior of the trajectories.** We begin by noting that the unique minimizer  $x(\epsilon)$  of  $f(\cdot, \epsilon)$  can be characterized as the unique solution of

$$\nabla f(x(\epsilon), \epsilon) = 0.$$

When  $\nabla^2 f(x(\epsilon), \epsilon)$  is positive definite, the implicit function theorem ensures that the curve  $\epsilon \rightarrow x(\epsilon)$  is differentiable and satisfies

$$\nabla^2 f(x(\epsilon), \epsilon) \frac{dx}{d\epsilon}(\epsilon) + \frac{\partial^2 f}{\partial \epsilon \partial x}(x(\epsilon), \epsilon) = 0.$$

Therefore, if the initial condition is  $u_0 = x(\epsilon_0)$  then  $u(t) = x(\epsilon(t))$  is the unique solution of (ACN).

We shall prove in the sequel that when the initial condition  $u_0$  does not lie on the optimal trajectory, the solution trajectory  $u(t)$  will nevertheless approach  $x(\epsilon(t))$ , from which we may deduce its convergence towards  $\hat{x}$ . As a by-product we obtain a useful estimate for the rate of convergence.

**4.1.1. The strongly convex case.** For simplicity, we shall assume in the following that the effective domain of the approximating function is a constant open set, that is to say, there exists an open set  $\Omega \subset H$  such that

$$\Omega = \operatorname{dom} f(\cdot, \epsilon) = \{x \in H : f(x, \epsilon) < +\infty\}$$

for all  $\epsilon > 0$ .

**THEOREM 4.1.** *Let us consider an approximate family  $\{f(\cdot, \epsilon)\}_{\epsilon > 0}$  such that  $f(\cdot, \cdot) \in C^2(\Omega \times (0, \epsilon_0]; \mathbb{R})$ , where  $\Omega = \text{dom}f(\cdot, \epsilon)$  for all  $\epsilon > 0$ . If  $f(\cdot, \epsilon)$  is strongly convex with parameter  $\beta(\epsilon)$  then*

(i) *For every initial condition  $u_0 \in \Omega$ , there exists a unique  $u : [0, +\infty) \rightarrow \Omega$  solution trajectory for (ACN) which satisfies*

$$(12) \quad |u(t) - x(\epsilon(t))| \leq |\nabla f(u_0, \epsilon_0)| \frac{e^{-t}}{\beta(\epsilon(t))}$$

where  $x(\epsilon)$  is the optimal solution of  $(P^\epsilon)$ .

(ii) *If  $(H_3)$  is satisfied and*

$$(13) \quad \lim_{t \rightarrow +\infty} e^t \beta(\epsilon(t)) = +\infty$$

then

$$\lim_{t \rightarrow +\infty} u(t) = \lim_{t \rightarrow +\infty} x(\epsilon(t)) = \hat{x}.$$

*Proof.* The proof uses the same techniques of the non-parametric case. To prove (i), we study the existence and uniqueness of an optimal trajectory for the variational characterization

$$(P_t^\epsilon) \quad u(t) = \text{Argmin}\{f(x, \epsilon(t)) - e^{-t} \langle \nabla f(u_0, \epsilon_0), x \rangle : x \in H\}.$$

Since  $f(\cdot, \epsilon)$  is strongly convex, then the function

$$g(x, t) := f(x, \epsilon(t)) - e^{-t} \langle \nabla f(u_0, \epsilon_0), x \rangle$$

is also strongly convex with the same parameter  $\beta(\epsilon)$ . So, there exists a unique  $u(t)$  optimal solution  $(P_t^\epsilon)$  which belongs to  $\Omega$ . Moreover, applying the Implicit Function Theorem to the optimality condition of  $(P_t^\epsilon)$ , we conclude that  $u : [0, +\infty) \rightarrow \Omega$  is differentiable and is a solution trajectory for (ACN). Noting that a solution of (ACN) is a optimal trajectory for  $(P_t^\epsilon)$ , the uniqueness for (ACN) follows.

For the estimate, it is enough to notice that from the strong monotonicity property it follows that

$$\langle \nabla f(u(t), \epsilon(t)), u(t) - x(\epsilon(t)) \rangle \geq \beta(\epsilon(t)) |u(t) - x(\epsilon(t))|^2,$$

so that

$$|u(t) - x(\epsilon(t))| \leq \frac{e^{-t}}{\beta(\epsilon(t))} |\nabla f(u_0, \epsilon_0)|.$$

Since (ii) is a direct consequence of (i) the proof is complete.  $\square$

Note that we have made no assumptions on the optimal path  $x(\epsilon)$  except its convergence towards  $\hat{x}$ . The latter is a distinguishing feature of the approximate Newton's method (ACN) with respect to known results for the steepest descent method (DADA) (see conditions (a) and (b) in §2.2). Moreover, equation (13) provides a relatively weak condition on the parameter function  $\epsilon(t)$  which ensures that the solution trajectories of (ACN) asymptotically approaches the optimal trajectory  $x(\epsilon(t))$ . We shall see in the following examples that condition (13) is much weaker than the basic hypothesis (7) needed for convergence in the steepest descent method.

Before proceeding with these examples it is illustrative to compare (12) with the corresponding estimate (6) for  $(DADA)$ . Assume that the strong convexity property holds with parameter  $\beta(\epsilon) = \epsilon$  and also that the optimal path satisfies

$$\left| \frac{dx}{d\epsilon}(\epsilon) \right| \leq C_0$$

for some constant  $C_0$ , which implies the finite length of the optimal path (see condition (a)). If we have

$$\int_0^{+\infty} \epsilon(s) ds = +\infty$$

then the solution  $u_D(t)$  of  $(DADA)$  norm converges to  $\hat{x}$ .

Take for instance  $\epsilon(t) = (1 - \alpha)(1 + t)^{-\alpha}$  with  $0 < \alpha < 1$ . After a careful analysis of (6), it is possible to prove that asymptotically,

$$(14) \quad |u_D(t) - x(\epsilon(t))| \leq \frac{C}{t^\alpha}$$

for a suitable constant  $C$  and  $t$  large enough (see [3]). Denote by  $u_N(t)$  the corresponding solution for  $(ACN)$ , then (12) directly becomes

$$|u_N(t) - x(\epsilon(t))| \leq \frac{|\nabla f(u_0, \epsilon_0)| (1 + t)^\alpha}{(1 - \alpha) e^t}$$

which is sharper than (14).

If we take  $\epsilon(t) = 1/(1 + t)$  then asymptotically (see [3]),

$$|u_D(t) - x(\epsilon(t))| \leq C \frac{\ln(t)}{t}$$

for certain  $C > 0$  and  $t$  large, while for the solution of  $(ACN)$  we have

$$|u_N(t) - x(\epsilon(t))| \leq |\nabla f(u_0, \epsilon_0)| \frac{1 + t}{e^t}$$

which converges to 0 much faster than  $\ln(t)/t$ .

The sharper estimates obtained with  $(ACN)$  must be balanced with the simpler form of  $(DADA)$  and with the extra effort involved in the computation of the inverse of the Hessian in  $(ACN)$ .

EXAMPLE 1. *Log-barrier in linear programming.*

With the linear program

$$(LP) \quad \min\{c'x : Ax \leq b\}$$

it is associated the function

$$f(x) = \begin{cases} c'x & \text{if } Ax \leq b \\ +\infty & \text{otherwise} \end{cases}$$

The log-barrier approximation of  $f$  is given by

$$(15) \quad f(x, \epsilon) = c'x - \epsilon \sum_{i=1}^m \ln(b_i - a'_i x).$$

where the vectors  $a_i$  denote the rows of the full rank matrix  $A$ .

If we assume that the interior of the feasible set  $F = \{x \in \mathbb{R}^n : Ax \leq b\}$  is non-empty and bounded, then the corresponding problem  $(P^\epsilon)$  has a unique solution  $x(\epsilon)$  which converges towards the analytic center  $\hat{x}$  of the optimal set  $S(LP)$  defined as the unique solution of

$$\max\left\{\sum_{i \notin I_0} \ln(b_i - a'_i x) : x \in S(LP)\right\}$$

where  $I_0 = \{i : a'_i x = b_i \text{ for all } x \in S(LP)\}$ .

For the log-barrier approximation we have  $\Omega = \text{dom}f(\cdot, \epsilon) = \{x \in \mathbb{R}^n : Ax < b\}$  which is an open subset of  $\mathbb{R}^n$ ,  $f(\cdot, \cdot) \in C^\infty(\Omega \times (0, +\infty); \mathbb{R})$  and the strong convexity hypothesis holds with  $\beta(\epsilon) = \mu\epsilon$  for a suitable constant  $\mu > 0$ . For  $x \in \Omega$  we have

$$\nabla f(x, \epsilon) = c + \epsilon \sum_{i=1}^m \frac{1}{(b_i - a'_i x)} a_i = c + \epsilon A' d(x),$$

$$\nabla^2 f(x, \epsilon) = \epsilon \sum_{i=1}^m \frac{1}{(b_i - a'_i x)^2} a_i a'_i = \epsilon A' D(x)^2 A$$

and

$$\frac{\partial^2 f}{\partial \epsilon \partial x}(x, \epsilon) = A' d(x)$$

where the vector valued function  $d : \Omega \rightarrow \mathbb{R}^m$  is defined by

$$d(x) := \left( \frac{1}{(b_i - a'_i x)} : i = 1, \dots, m \right),$$

and  $D(x)$  is the diagonal matrix  $D(x) := \text{diag}(d(x))$ .

Let  $\epsilon(\cdot) \in C^1([0, +\infty); \mathbb{R}_+)$  be decreasing to 0. In this case, condition (13) of Theorem 4.1 becomes

$$(16) \quad \lim_{t \rightarrow +\infty} e^t \epsilon(t) = +\infty.$$

For example, take  $\epsilon(t) = e^{-\alpha t}$  with  $0 < \alpha < 1$  or  $\epsilon(t) = (1+t)^{-r}$  with  $r > 0$ . Then we obtain as a corollary the following result.

**PROPOSITION 4.2.** *According to the previous definitions and assuming (16), for every interior point  $u_0 \in \{x \in \mathbb{R}^n : Ax < b\}$  the unique solution of*

$$(ACN) \quad \begin{cases} [\epsilon(t)A'D(u(t))^2A]\dot{u}(t) + (\dot{\epsilon}(t) + \epsilon(t))A'd(u(t)) + c = 0 \\ u(0) = u_0 \end{cases}$$

converges as  $t \rightarrow +\infty$  towards  $\hat{x}$  the analytic center of the optimal set  $S(LP)$ .

**REMARK.** If  $\epsilon(t) = e^{-\alpha t}$  with  $0 < \alpha < 1$  or  $\epsilon(t) = (1+t)^{-r}$  with  $r > 1$ , then  $\int_0^{+\infty} \beta(\epsilon(s))ds < +\infty$ , and we can not ensure the convergence to  $\hat{x}$  of the solution for the corresponding approximate steepest descent method (DADA).  $\square$

**4.1.2. The non-global strongly convex case.** Let us return to the variational characterization

$$(P_t^\epsilon) \quad u(t) = \operatorname{Argmin}\{f(x, \epsilon(t)) - e^{-t}\langle x_0^*, x \rangle : x \in H\}.$$

where  $x_0^* \in \partial f(u_0, \epsilon_0)$  and  $\epsilon(t)$  decreases to 0 as  $t \rightarrow +\infty$ . When the approximate function fails to satisfy the strong convexity property (for example, when it is just strongly convex over bounded sets), we can not apply directly Theorem 4.1. However, in certain cases we can generalize our analysis if we take into account the following steps: first prove that the problem  $(P_t^\epsilon)$  admits a unique solution  $u(t)$  for all  $t > 0$ , and then prove that the trajectory  $u(t)$  is bounded, hence the estimate

$$|u(t) - x(\epsilon(t))| \leq |\nabla f(u_0, \epsilon_0)| \frac{e^{-t}}{\beta_K(\epsilon(t))}$$

holds for a suitable constant  $K > 0$ .

In order to accomplish the first step, the following proposition may be a useful tool. We shall assume that the effective domain of the Fenchel conjugate of the approximate function is independent of  $\epsilon$ . More precisely, there exists a set  $F \subset H$  such that

$$F = \operatorname{dom} f^*(\cdot, \epsilon) = \{x^* \in H : f^*(x^*, \epsilon) < +\infty\}$$

for all  $\epsilon > 0$ , where

$$f^*(x^*, \epsilon) := \sup_{x^* \in H} \{\langle x^*, x \rangle - f(x, \epsilon)\}.$$

**PROPOSITION 4.3.** *Assume that for every  $\epsilon > 0$  the approximate function  $f(\cdot, \epsilon)$  is coercive and satisfies  $(H_2)$ . If there exists  $F \subset H$  such that  $F = \operatorname{dom} f^*(\cdot, \epsilon)$  for all  $\epsilon > 0$ , then problem  $(P_t^\epsilon)$  admits a unique solution  $u(t)$  for all  $t > 0$ .*

*Proof.* As in the proof of Proposition 3.1, let us consider the function  $g(x, t) := f(x, \epsilon(t)) - e^{-t}\langle x_0^*, x \rangle$ . Since  $f(\cdot, \epsilon)$  is strictly convex, the same holds for  $g(\cdot, t)$  so its minimum must be unique.

To establish the existence of  $u(t)$ , we note that  $f(\cdot, \epsilon_0)$  is strongly convex over bounded sets, so that the same holds for  $g(\cdot, 0)$ . Since the minimum of the latter is attained (at  $u_0$ ) we deduce that there exist constants  $K > 0$  and  $\beta_0 > 0$  such that

$$g(u_0, 0) + \frac{\beta_0}{2}|x - u_0|^2 \leq g(x, 0)$$

for all  $x \in B(u_0, K)$ . For simplicity assume that  $u_0 = 0$  and  $g(u_0, \cdot) = 0$ , so we have

$$(17) \quad \frac{\beta_0}{2}|x|^2 \leq g(x, 0)$$

when  $x \in B(0, K)$ . Suppose that for certain  $\alpha \in \mathbb{R}$  the corresponding level set is unbounded. Take a sequence  $x_k \in H$  such that  $|x_k| \rightarrow +\infty$  with  $g(x_k, 0) \leq \alpha$ . Using (17) with  $x := Kx_k/|x_k|$  and from the convexity of  $g(\cdot, 0)$  it follows that for  $k$  large enough such that  $K/|x_k| < 1$  then

$$\frac{\beta_0}{2}K^2 \leq g\left(K\frac{x_k}{|x_k|}, 0\right) \leq \frac{K}{|x_k|}g(x_k, 0) \leq \frac{k\alpha}{|x_k|}$$

which contradicts the fact that  $K > 0$  and  $\beta_0 > 0$ . We conclude that  $g(\cdot, 0)$  is a coercive function, thus  $g^*(\cdot, 0)$  is finite and continuous at 0.

On the other hand, we have

$$g^*(x^*, t) = f^*(x^* + e^{-t}x_0^*, \epsilon(t))$$

for all  $x^* \in H$ , and therefore  $f^*(\cdot, \epsilon_0)$  is finite and continuous at  $x_0^*$ . In particular  $x_0^*$  belongs to the interior of the effective domain of  $f^*(\cdot, \epsilon_0)$ , and since  $F = \text{dom}f^*(\cdot, \epsilon)$  for all  $\epsilon > 0$ , it follows that  $x_0^*$  belongs to the interior of  $\text{dom}f^*(\cdot, \epsilon(t))$  for all  $t \geq 0$ . Since  $f(\cdot, \epsilon(t))$  is coercive, we also have that 0 belongs to the interior of  $\text{dom}f^*(\cdot, \epsilon(t))$  for all  $t > 0$ , and therefore the same conclusion holds for  $e^{-t}x_0^*$ . It follows that  $f^*(\cdot, \epsilon(t))$  is finite and continuous at  $e^{-t}x_0^*$ , from which we get that  $g^*(\cdot, t)$  is finite and continuous at 0, that is to say  $g(\cdot, t)$  is coercive and its minimum must be attained.  $\square$

REMARK. In the finite dimensional case it is possible to replace  $(H_2)$  by just assuming  $f(\cdot, \epsilon)$  strictly convex, and the previous proposition still holds.

Notice that when  $f(\cdot, \epsilon)$  is strongly convex over the whole space, it is clear that

$$f^*(x^*, \epsilon) = \min_{x \in H} \{f(x, \epsilon) - \langle x^*, x \rangle\} < +\infty$$

so that  $\text{dom}f^*(\cdot, \epsilon) = H$  and we may apply Proposition 4.3. Nevertheless, in this case the proof for the existence of  $u(t)$  is direct. A more interesting situation where the latter result applies is the following.

EXAMPLE 2. *Exponential penalty in linear programming.*

Let us consider the linear program (LP) of Example 1 and the approximation given by

$$(18) \quad f(x, \epsilon) = c'x + \epsilon \sum_{i=1}^m \exp((a'_i x - b_i)/\epsilon).$$

If we assume that the optimal set  $S(LP)$  is non-empty and bounded, then there exists a unique optimal path  $x(\epsilon)$  of solutions of  $(P^\epsilon)$  which converges to a particular point  $\hat{x} \in S(LP)$  called the centroid (see Cominetti and San Martin [9]).

For every  $x \in \mathbb{R}^n$  we have

$$\nabla f(x, \epsilon) = c + \sum_{i=1}^m \exp((a'_i x - b_i)/\epsilon) a_i = c + A' \lambda(x, \epsilon),$$

$$\nabla^2 f(x, \epsilon) = \frac{1}{\epsilon} \sum_{i=1}^m \exp((a'_i x - b_i)/\epsilon) a_i a'_i = \frac{1}{\epsilon} A' D(x, \epsilon) A$$

and

$$\frac{\partial^2 f}{\partial \epsilon \partial x}(x, \epsilon) = -\frac{1}{\epsilon^2} \sum_{i=1}^m \exp((a'_i x - b_i)/\epsilon) (a'_i x - b_i) a_i = -\frac{1}{\epsilon} A' S(\lambda(x, \epsilon))$$

where the vector valued function  $\lambda : \mathbb{R}^n \times (0, +\infty) \rightarrow \mathbb{R}^m$  is defined by

$$(19) \quad \lambda(x, \epsilon) := \exp((Ax - b)/\epsilon) = (\exp((a'_i x - b_i)/\epsilon) : i = 1, \dots, m),$$

the diagonal matrix  $D(x, \epsilon) := \text{diag}(\lambda(x, \epsilon))$  and  $S : \mathbb{R}_+^m \rightarrow \mathbb{R}^m$  by

$$S(\lambda) := (\lambda_i \ln \lambda_i : i = 1, \dots, m),$$

with the convention  $0 \ln 0 = 0$ .

For this approximation, we have  $\Omega = \text{dom}f(\cdot, \epsilon) = \mathbb{R}^n$  and  $f(\cdot, \epsilon)$  is strongly convex over bounded sets with parameter

$$\beta_K(\epsilon) = \frac{\mu_K}{\epsilon} e^{-M_K/\epsilon}$$

for suitable strictly positive constants  $\mu_K$  and  $M_K$ . Note that if the feasible set is unbounded, the exponential penalty function fails to be strongly convex with global parameter (an analogous situation occurs for the log-barrier function) and we can not apply Theorem 4.1.

**PROPOSITION 4.4.** *Let  $\epsilon : \mathbb{R} \rightarrow \mathbb{R}_+$  be decreasing to 0 and consider the exponential penalty function  $f(\cdot, \epsilon)$  defined above. Then problem  $(P_t^\epsilon)$  admits a unique solution trajectory  $u(t)$  which stays bounded as  $t \rightarrow +\infty$ .*

*Proof.* In this case,  $f(\cdot, \epsilon)$  is strongly convex over bounded sets and it is not difficult to prove that

$$\text{dom}f^*(\cdot, \epsilon) = \{\lambda \in \mathbb{R}^m : A' \lambda = -c, \lambda \geq 0\}$$

which does not depend on  $\epsilon$ . Hence, the existence and uniqueness of the optimal trajectory  $u(t)$  follows from Proposition 4.3.

Let us consider once again the function

$$g(x, t) := f(x, \epsilon(t)) - e^{-t} \nabla f(u_0, \epsilon_0)' x.$$

Suppose that for a sequence  $t_k \rightarrow +\infty$  we have  $|u(t_k)| \rightarrow +\infty$ . Passing to a subsequence we can assume that  $\frac{u(t_k)}{|u(t_k)|} \rightarrow z \in \mathbb{R}^n$  for some  $z$  such that  $|z| = 1$ . Let  $x \in \mathbb{R}^n$  be fixed and such that  $Ax \leq b$ . From the optimality of  $u(t_k)$  and the definition of the exponential penalty we have

$$(20) \quad g(u(t_k), t_k) \leq g(x, t_k) \leq c'x - e^{-t_k} \nabla f(u_0, \epsilon_0)' x + \epsilon(t_k)m.$$

Hence

$$c'u(t_k) - e^{-t_k} \nabla f(u_0, \epsilon_0)' u(t_k) \leq c'x - e^{-t_k} \nabla f(u_0, \epsilon_0)' x + \epsilon(t_k)m,$$

and dividing by  $|u(t_k)|$  and passing to the limit we get

$$c'z \leq 0.$$

On the other hand, from (20) we also have

$$a'_i x_k - b_i \leq \epsilon(t_k) \ln[(L + c'(x - u(t_k)) - e^{-t_k} \nabla f(u_0, \epsilon_0)'(x - u(t_k)) + \epsilon(t_k)m)/\epsilon(t_k)]$$

for a suitable constant  $L$ . From this it follows that  $a'_i z \leq 0$  and we have found a vector  $z \neq 0$  such that  $c'z \leq 0$  and  $Az \leq 0$ , which contradicts the boundedness of the optimal set  $S(LP)$ .  $\square$

With this result we have the existence of a bounded solution trajectory  $u(t)$  for the associated dynamical system (ACN), and we obtain the estimate

$$(21) \quad |u(t) - x(\epsilon(t))| \leq |\nabla f(u_0, \epsilon_0)| \frac{e^{-t}}{\beta_K(\epsilon(t))} = C e^{-t+M/\epsilon(t)} \epsilon(t)$$



for suitable constants  $C > 0$  and  $M > 0$ . If we have

$$\lim_{t \rightarrow +\infty} \frac{e^{t-M/\epsilon(t)}}{\epsilon(t)} = +\infty$$

for all  $M > 0$ , then  $u(t)$  converges towards the centroid of  $S(LP)$ .

Estimates like (21) may also be useful in order to obtain dual convergence results. To be more precise we observe that by means of the Fenchel duality theory and in the case of the exponential penalty approximation, it is possible to associate to the corresponding problem  $(P^\epsilon)$  the following dual problem

$$(D^\epsilon) \quad \min\{b'\lambda + \epsilon \sum_{i=1}^m \lambda_i (\ln \lambda_i - 1) : A'\lambda = -c, \lambda \geq 0\}$$

which can be interpreted as a penalized version of the classical dual problem of  $(LP)$ , namely

$$(D) \quad \min\{b'\lambda : A'\lambda = -c, \lambda \geq 0\}$$

The previous dual was studied in [9] where it is shown that if  $x(\epsilon)$  denotes the optimal solution of the exponential penalized problem  $(P^\epsilon)$ , then  $\lambda(x(\epsilon), \epsilon)$  (with  $\lambda(x, \epsilon)$  defined as in (19)) is the unique optimal solution of  $(D^\epsilon)$ . Moreover,  $\lambda(x(\epsilon), \epsilon)$  converges towards a particular solution  $\widehat{\lambda}$  of the dual problem  $(D)$ , characterized as the unique solution of

$$(D_0) \quad \min\{\sum_{i \in I_0} \lambda_i (\ln \lambda_i - 1) : A'\lambda = -c, \lambda_i = 0 \ i \notin I_0, \lambda_i \geq 0 \ i \in I_0\}.$$

where  $I_0 = \{i : a'_i x = b_i \text{ for all } x \in S(LP)\}$ .

Denote for simplicity  $\lambda(x, t) := \lambda(x, \epsilon(t))$ . Since

$$\begin{aligned} \lambda_i(u(t), t) &= \exp(a'_i(u(t) - x(\epsilon(t)))/\epsilon(t)) \exp((a'_i x(\epsilon(t)) - b_i)/\epsilon(t)) \\ &= \exp(a'_i(u(t) - x(\epsilon(t)))/\epsilon(t)) \lambda_i(x(\epsilon(t)), t), \end{aligned}$$

for all  $i = 1, \dots, m$ , if we have

$$(22) \quad \lim_{t \rightarrow +\infty} \frac{|u(t) - x(\epsilon(t))|}{\epsilon(t)} = 0$$

then

$$\lim_{t \rightarrow +\infty} \lambda(u(t), t) = \lim_{t \rightarrow +\infty} \lambda(x(\epsilon(t)), t) = \widehat{\lambda}.$$

In order to have (22), using the estimate (21) we obtain

$$\frac{|u(t) - x(\epsilon(t))|}{\epsilon(t)} \leq |\nabla f(u_0, \epsilon_0)| \frac{e^{-t}}{\beta_K(\epsilon(t))\epsilon(t)} = C e^{-t+M/\epsilon(t)}.$$

Therefore, we must have

$$\lim_{t \rightarrow +\infty} e^{t-M/\epsilon(t)} = +\infty$$

or equivalently,

$$(23) \quad \lim_{t \rightarrow +\infty} t - \frac{M}{\epsilon(t)} = +\infty$$

for all  $M > 0$ . For example,  $\epsilon(t) = (1+t)^{-\alpha}$  with  $0 < \alpha < 1$ .

Finally, since (22) implies

$$\lim_{t \rightarrow +\infty} |u(t) - x(\epsilon(t))| = 0,$$

we have the following result.

**PROPOSITION 4.5.** *According to the previous definitions, if the parameter  $\epsilon(t)$  satisfies (23) then for every  $u_0 \in \mathbb{R}^n$  the unique solution of*

$$(ACN) \quad \begin{cases} [\frac{1}{\epsilon(t)} A' D(u(t), \epsilon(t)) A] \dot{u}(t) - \frac{\dot{\epsilon}(t)}{\epsilon(t)} A' S(\lambda(u(t), t)) + A' \lambda(u(t), t) + c = 0 \\ u(0) = u_0 \end{cases}$$

converges as  $t \rightarrow +\infty$  towards  $\hat{x}$  the centroid of the optimal set  $S(LP)$ . Moreover

$$\lambda(u(t), t) = \exp((Au(t) - b)/\epsilon(t))$$

converges to  $\hat{\lambda}$  a particular solution of the dual problem (D).

**REMARK.** In this case, the dual convergence for (ACN) comes from the sharp estimate (21) for the distance between  $u(t)$  and the optimal trajectory  $x(\epsilon(t))$ . A similar analysis for (DADA) trajectories may be involved since the corresponding estimate (6) is more complicated.

**5. Epi-convergence and scaling.** When the strong monotonicity parameter  $\beta(\epsilon)$  goes to 0, the increasing ill-conditioning of the Hessian operator of the approximating function represents an important problem from the numerical point of view. In this section, we shall see how in certain cases, an appropriate rescaling of the differential equation may be used to transform the original system into a renormalized system where  $\beta(\epsilon)$  is identically equal to 1.

In many situations (see [1, 17]) the epi-convergence and scaling method provides a variational characterization of the point  $\hat{x} \in \text{Argmin } f$  obtained as a limit of the optimal trajectory  $x(\epsilon) := \text{Argmin } f(\cdot, \epsilon)$ . In order to describe this method assume that

$$f = \text{epi} - \lim_{\epsilon \rightarrow 0} f(\cdot, \epsilon),$$

that is,  $f$  is the Mosco-epi-limit of the parametric family of strongly convex functions  $f(\cdot, \epsilon)$ .

Let us define

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

and consider the rescaled minimization problem

$$(R^\epsilon) \quad \min\{h(x, \epsilon) : x \in H\}$$

Note that  $(R^\epsilon)$  has also the point  $x(\epsilon)$  as unique solution. Hence, with this rescaling no changes are introduced in the optimal trajectory, and the parametric family

$\{h(\cdot, \epsilon)\}_{\epsilon > 0}$  has the same *regularizing* properties as  $\{f(\cdot, \epsilon)\}_{\epsilon > 0}$  (coercivity, smoothness, etc.). However, for the new family of functions, the strong monotonicity property holds with parameter identically equal 1,

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

Suppose now that in addition to the epi-convergence of  $f(\cdot, \epsilon)$  we assume that  $h(\cdot, \epsilon)$  epi-converges towards a proper function  $h$ ,

$$h = \text{epi} - \lim_{\epsilon \rightarrow 0} h(\cdot, \epsilon).$$

Then, as a consequence (see Attouch [1]) we have

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

so that  $h$  is a strongly convex function. Let  $\hat{x}$  be its unique minimizer. Since  $h(x) = +\infty$  for  $x \notin \text{Argmin } f$ , then  $\hat{x} \in \text{Argmin } f$ , and since from the Mosco-epi-convergence we have that every limit point of the optimal path  $\{x(\epsilon) : \epsilon \rightarrow 0\}$  minimizes  $h$ , we conclude

$$\lim_{\epsilon \rightarrow 0} x(\epsilon) = \hat{x}.$$

Let us assume smoothness for the parametric family  $f(\cdot, \epsilon)$  in order to study the dynamical system

$$\nabla^2 f(u(t), \epsilon(t)) \dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \nabla f(u(t), \epsilon(t)) = 0.$$

If we re-scale this equation multiplying by  $1/\beta(\epsilon(t))$  we get

$$\nabla^2 h(u(t), \epsilon(t)) \dot{u}(t) + \frac{\dot{\epsilon}(t)}{\beta(\epsilon(t))} \frac{\partial^2 f}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \nabla h(u(t), \epsilon(t)) = 0.$$

Assuming in addition that  $\beta(\epsilon)$  is differentiable, we have

$$\frac{\partial^2 f}{\partial \epsilon \partial x}(x, \epsilon) = \beta'(\epsilon) \nabla h(x, \epsilon) + \beta(\epsilon) \frac{\partial^2 h}{\partial \epsilon \partial x}(x, \epsilon),$$

and therefore we get

$$\nabla^2 h(u(t), \epsilon(t)) \dot{u}(t) + \dot{\epsilon}(t) \frac{\partial^2 h}{\partial \epsilon \partial x}(u(t), \epsilon(t)) + \left[1 + \frac{\beta'(\epsilon(t)) \cdot \epsilon(t)}{\beta(\epsilon(t))}\right] \nabla h(u(t), \epsilon(t)) = 0.$$

Let us define

$$\eta(t) := t + \ln(\beta(\epsilon(t)))$$

and introduce the change of variables  $s = \eta(t)$ . Denoting for simplicity  $v(s) := u(\eta^{-1}(s))$  and  $h(x, s) := h(x, \epsilon(\eta^{-1}(s)))$  we get the equivalent differential equation

$$\nabla^2 h(v(s), s) \dot{v}(s) + \frac{\partial^2 h}{\partial s \partial x}(v(s), s) + \nabla h(v(s), s) = 0$$

where now the functions  $h(\cdot, s)$  are strongly convex with a uniform parameter  $\beta = 1$ .

For the asymptotic analysis it is natural to require that the reparametrization satisfies  $\lim_{t \rightarrow +\infty} \eta(t) = +\infty$ . In the particular case above, this amounts to

$$\lim_{t \rightarrow +\infty} e^t \beta(\epsilon(t)) = +\infty,$$

which is precisely the hypothesis made in Theorem 4.1. Under this condition, the estimate

$$|v(s) - x(\epsilon(s))| \leq e^{-s} |\nabla h(v_0, \epsilon_0)|$$

is equivalent to the one given in Theorem 4.1.

EXAMPLE 3. *Viscosity method.*

Let  $f$  be a closed convex function with optimal set  $\text{Argmin } f$  nonempty. Let the approximation be

$$f(x, \epsilon) = f(x) + \frac{\epsilon}{2} |x|^2,$$

which regularizes problem  $(P)$  by adding a strongly convex term. In this case, we may take  $\beta(\epsilon) = \epsilon$ . We have that

$$f = \text{epi} - \lim_{\epsilon \rightarrow 0} f(\cdot, \epsilon),$$

and if we define

$$h(x, \epsilon) := \frac{1}{\epsilon} [f(x, \epsilon) - \min f] = \frac{f(x) - \min f}{\epsilon} + \frac{|x|^2}{2},$$

then it is easy to prove that

$$h = \text{epi} - \lim_{\epsilon \rightarrow 0} h(\cdot, \epsilon)$$

where

$$h := \begin{cases} \frac{1}{2} |x|^2 & \text{if } x \in \text{Argmin } f \\ +\infty & \text{otherwise} \end{cases}$$

Therefore, the unique solution  $x(\epsilon)$  of  $(P^\epsilon)$  converges towards the element of minimal norm in  $\text{Argmin } f$  (see Tikhonov and Arsenine [16]).

Assume that  $f \in C^2(H; \mathbb{R})$  and take the parameter function  $\epsilon(t) = e^{-\alpha t}$  with  $0 < \alpha < 1$ . Then

$$\eta(t) = t + \ln(\epsilon(t)) = (1 - \alpha)t$$

and the change of variables  $s = (1 - \alpha)t$  leads to the function  $\epsilon(s) = e^{-\frac{\alpha}{1-\alpha}s}$ . If we choose for instance  $\alpha = \frac{1}{2}$  then

$$h(x, s) := e^s (f(x) - \min f) + \frac{|x|^2}{2}$$

and we obtain

PROPOSITION 5.1. *For every  $v_0 \in H$ , the unique solution of*

$$\begin{cases} [e^s \nabla^2 f(v(s)) + I] \dot{v}(s) + 2e^s \nabla f(v(s)) + v(s) = 0 \\ v(0) = v_0 \end{cases}$$

*converges as  $s \rightarrow +\infty$  towards  $\hat{x}$  the element of minimal norm in  $\text{Argmin } f$ .  $\square$*

## ACKNOWLEDGMENT

We want to express our thanks to Prof. R. Cominetti for having suggested us the study of coupling approximating schemes with Newton's method and for his valuable comments and discussions.

## REFERENCES

- [1] Attouch H (1984) Variational convergence for functions and operators. Applicable Maths. Series, Pitman, London.
- [2] Attouch H (1996) Viscosity solutions of minimization problems. *SIAM J. Optimization*, Vol. 6, No. 3 : 769-806
- [3] Attouch H, Cominetti R (1996) A dynamical approach to convex minimization coupling approximation with the steepest descent method. *J. Diff. Equations* 128 : 519-540.
- [4] Aubin JP, Cellina A (1984) *Differential Inclusions : Set-valued maps and viability theory*. Springer-Verlag.
- [5] Auslender A, Cominetti R, Haddou M. Asymptotic analysis for penalty methods in convex and linear programming, to appear in *Math. of Operations Research*.
- [6] Brézis H (1973) *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. Mathematical Studies 5, North-Holland, Amsterdam.
- [7] Brézis H (1978) Asymptotic behavior of some evolution systems. *Nonlinear Evolution Equations*, Academic Press.
- [8] Bruck RE (1975) Asymptotic convergence of nonlinear contraction semi-groups in Hilbert spaces. *J. Functional Analysis* 18 : 15-26.
- [9] Cominetti R, San Martín J. (1994) Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Programming*. 67 : 169-187.
- [10] Fiacco AV (1990) Perturbed variations of penalty function methods. *Annals of Operations Research* 27 : 371-380.
- [11] Furuya H, Miyashiba K, Kenmochi N (1986) Asymptotic behavior of solutions to a class of nonlinear evolution equations. *J. Diff. Equations* 62 : 73-94.
- [12] Gonzaga CG (1992) Path-following methods for linear programming. *SIAM Review* 34 : 167-224.
- [13] Lemaire B (1993) An asymptotical variational principle associated with some differential inclusions, continuous and discrete case. Conference colloque franco-vietnamien, ICAA, Hanoi.
- [14] Megiddo N (1989) Pathways to the optimal set in linear programming. *Progress in Math. Prog.: interior-point and related methods*. Springer-Verlag 131-158.
- [15] Rockafellar RT (1970) *Convex Analysis*. Princeton University Press, Princeton, N.J.
- [16] Tikhonov A, Arsenine V (1974) *Méthodes de résolution de problèmes mal posés*. MIR.
- [17] Torralba D (1996) Développements asymptotiques pour les méthodes d'approximation par viscosité. *CRAS*, t. 322, Série I : 123-128.