Journal of Convex Analysis
Volume 3 (1996), No.1, 63–70

# An Asymptotical Variational Principle Associated with the Steepest Descent Method for a Convex Function

**B. Lemaire**
*Université Montpellier II, Place E. Bataillon,*
*34095 Montpellier Cedex 5, France.*
*e-mail:lemaire@math.univ-montp2.fr*

Received July 5, 1994
Revised manuscript received January 22, 1996

**Dedicated to R. T. Rockafellar on his 60th Birthday**

The asymptotical limit of the trajectory defined by the continuous steepest descent method for a proper closed convex function $f$ on a Hilbert space is characterized in the set of minimizers of $f$ via an asymptotical variational principle of Brezis-Ekeland type. The implicit discrete analogue (prox method) is also considered.

*Keywords* : Asymptotical, convex minimization, differential inclusion, prox method, steepest descent, variational principle.

*1991 Mathematics Subject Classification:* 65K10, 49M10, 90C25.

## 1. Introduction

Let $X$ be a real Hilbert space endowed with inner product $\langle .,. \rangle$ and associated norm $\|.\|$, and let $f$ be a proper closed convex function on $X$.
The paper considers the problem of minimizing $f$, that is, of finding $\inf_X f$ and some element in the optimal set $S := \operatorname{Argmin} f$, this set assumed being non empty.
Letting $\partial f$ denote the subdifferential operator associated with $f$, we focus on the continuous steepest descent method associated with $f$, *i.e.*, the differential inclusion

$$-\frac{du}{dt} \in \partial f(u), \quad t > 0$$

with initial condition

$$u(0) = u_0.$$

This method is known to yield convergence under broad conditions summarized in the following theorem. Let us denote by $\mathcal{A}$ the real vector space of continuous functions from $[0, +\infty[$ into $X$ that are absolutely continuous on $[\delta, +\infty[$ for all $\delta > 0$.

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

**Theorem 1.1.**

(i) ([2]) *For all $u_0$ in* $\mathrm{cl}\,\mathrm{dom}\, f$*, the closure of the effective domain of $f$, there exists a unique $u$ in $\mathcal{A}$ such that:*

*$\forall t > 0, u(t)$ belongs to the domain of $\partial f$;*

*he differential inclusion is verified for a.e. $t > 0$ (recall that $u$ is strongly derivable for a.e. $t > 0$);*

*$\forall \delta > 0$, $u$ is Lipschitz continuous on $[\delta, +\infty[$ i.e. $u'(:= \frac{du}{dt}) \in \mathcal{L}^\infty([\delta, +\infty[; X)$;*

*$f \circ u$ is convex, non increasing, Lipschitz continuous on $[\delta, +\infty[$ $\forall \delta > 0$;*

*$u$ has a derivative to the right for all $t > 0$ and, if $S := \mathrm{Argmin}\, f$ is non empty, then $\lim_{t \to +\infty} \| \frac{du}{dt}^+(t) \| = 0$.*

(ii) ([5]) *If $S \neq \emptyset$ then the weak limit $u_\infty := w - \lim_{t \to +\infty} u(t)$ exists and belongs to $S$, this asymptotic convergence being true for the norm topology if $f$ is even* ([5]) *or if* $\mathrm{Int}\, S \neq \emptyset$ ([2]).

Assuming henceforth that the optimal set $S$ is non empty, Theorem 1.1 motivates us to consider the following problem:

*For a given $u_0 \in \mathrm{cl}\,\mathrm{dom}\, f$, characterize the asymptotic limit $u_\infty$ as the unique solution to some convex optimization problem with feasible set $S$.*

This task amounts to show that, like regularization procedures *à la Tikhonov* ([15], [7], [1]), the steepest descent method enables us to select a particular optimal solution respectively to some auxiliary criterium. As we shall see, unlike regularization procedures, this auxiliary criterium is not at our disposal independently from the data $f$ but is defined from this data. Indeed, the result is based upon a variational principle of Brezis-Ekeland type for an infinite horizon. An analogue holds true for the Euler's implicit discretization of the differential inclusion *i.e.* the proximal method of Martinet-Rockafellar.

The paper is organized as follows. Before reaching the main objective stated above, we show in section 2, that the trajectory $u$ is always (even $S$ is empty) minimizing and, if $S$ is non empty, we give a localization result of $u_\infty$ in $S$ with respect to $u_0$. Section 3 is devoted to the variational principle for an infinite horizon. In section 4 we state the asymptotic variational principle characterizing $u_\infty$ in $S$. The analogue for the discrete case is discussed in section 5.

In sections 2, 3, 4, $u$ always denotes the solution mentioned in Theorem 1.1.

## 2. Localization of $u_\infty$

First let us recall an estimate.

**Proposition 2.1.** [9]

$$\forall x \in X, \forall t > 0, \quad \frac{\|u(t) - x\|^2}{2t} + f(u(t)) \leq f(x) + \frac{\|u_0 - x\|^2}{2t}$$

**Proof.** By definition of the subdifferential we have

$$\forall x \in X, \ \textit{for a.e. } s > 0, \ f(x) \geq f(u(s)) - \langle u'(s), x - u(s) \rangle$$

Then, integrating between $\delta$ and $t$ with $0 < \delta < t$, and thanks to the non increasingness of $f(u(.))$ and the absolute continuity of $u$ on $[\delta, +\infty[$, we get

$$(t-\delta)f(x) \geq (t-\delta)f(u(t)) + (\|u(t)-x\|^2 - \|u(\delta)-x\|^2)/2$$

Finally, since $u$ is continuous at 0 we get the result. □

**Corollary 2.2.**
(i) *In all cases, that is, whether* $\inf_X f$ *be finite or not, achieved or not,* $\lim_{t\to+\infty} f(u(t)) = \inf_X f$.
(ii) *If* $S \neq \emptyset$, *then* $\|u_\infty - proj_S u_0\| \leq d(u_0, S)$ *and therefore,* $\|u_\infty - u_0\| \leq 2d(u_0, S)$, *where* $proj_S$ *denotes the projection operator onto the closed convex set* $S$.
(iii) *If* $S$ *is an affine subspace then* $u_\infty = proj_S u_0$.

**Proof.** (i) Trivial.
(ii) Taking $x \in S$ in the estimate we get $\|u(t)-x\| \leq \|u_0 - x\|$. Passing to the lim inf as $t \to +\infty$, thanks to the weak lower semi-continuity of the norm, and taking as $x$ the projection of $u_0$ onto $S$, we get the result.
(iii) Let $e := u_\infty - proj_S u_0$ and assume $e \neq 0$. As $S$ is an affine subspace, $x := proj_S u_0 - td(u_0, S)e/\|e\|$ is in $S$ for all $t \geq 0$. We have $\|u_\infty - x\| = td(u_0, S) + \|e\|$ and, thanks to Pythagore, $\|u_0 - x\| = \sqrt{t^2+1}\, d(u_0, S)$. Therefore, $\|u_0 - x\| - \|u_\infty - x\| = (\sqrt{t^2+1} - t)d(u_0, S) - \|e\|$. Referring to the proof of (ii) above we get $\|e\| \leq (\sqrt{t^2+1} - t)d(u_0, S)$ for all $t \geq 0$, which is a contradiction. □

**Remark 2.3.** (iii) in corollary 2.2 gives a complete and simple characterization of $u_\infty$ in $S$, actually the same than the Tikhonov regularization method. This characterization has yet been obtained in [4] for the discrete method (cf. section 5 below) when $f$ is a quadratic form (a case where $S$ is a subspace). Unfortunately this simple characterization fails to be true in general.

## 3. Variational principle, infinite horizon

First, let $T > 0$ be fixed. We denote by $\mathcal{L}^1_{sc}(0, T)$ the set of functions $\psi$ from $[0, T]$ into $\mathbb{R}$ that belong to $\mathcal{L}^1(\delta, T)$ for all $\delta$, $0 < \delta < T$, and such that $\int_0^T \psi(t)dt := \lim_{\delta\to 0} \int_\delta^T \psi(t)dt$ exists in $\mathbb{R}$.
We define the set of feasible trajectories which reach some $v_\infty \in X$ from $u_0$, possibly using infinite time, arriving asymptotically with nul velocity:

$$K(u_0) := \{v \in \mathcal{A};\ f \circ v + f^* \circ (-v') \in \mathcal{L}^1_{sc}(0, T)\ \ \forall T > 0,$$

$$v(0) = u_0,\ v_\infty := w - \lim_{t\to+\infty} v(t) \text{ exists},\ s - ess - \lim_{t\to+\infty} v'(t) = 0\},$$

where $w$ is short notation for *weak*, $s$ is for *strong*, $ess$ is for *essential*, and the $*$ denotes the Fenchel conjugacy.
We note that, if $v$ is in $K(u_0)$ then $v(t)$ is in $\operatorname{dom} f$ for all $t > 0$ and $-v'(t)$ is in $\operatorname{dom} f^*$ for a.e. $t > 0$.

**Proposition 3.1.** *$K(u_0)$ is a convex subset of $\mathcal{A}$ containing $u$.*

**Proof.** Convexity is easy. For the second assertion, integrating the equality

$$f(u(t)) + f^*(-u'(t)) = -\langle u'(t), u(t)\rangle \text{ for a.e. } t > 0$$

between $\delta$ and $T$ for $\delta > 0$ and passing to the limit as $\delta$ tends to 0, we get

$$\int_0^T [f(u(t)) + f^*(-u'(t))]dt = \frac{1}{2}(\|u_0\|^2 - \|u(T)\|^2)$$

The last condition is fulfilled since (cf. Theorem 1.1) $\lim_{t\to+\infty} \|\frac{du}{dt}^+(t)\| = 0$. □

Motivated by Proposition 3.1 we define the cost function for the finite horizon $T$, $J_T : K(u_0) \to \mathbb{R}$ by

$$J_T(v) := \int_0^T [f(v(t)) + f^*(-v'(t)) + \langle v'(t), v(t)\rangle]dt$$

$$= \int_0^T [f(v(t)) + f^*(-v'(t))]dt + \frac{1}{2}(\|v(T)\|^2 - \|u_0\|^2)$$

As $u_0$ is fixed, we note that $J_T$ is convex. Moreover, via Fenchel's inequality we have $J_T \geq 0$.

**Remark 3.2.** According to Moreau's theory for elastoplastic systems [11,12,13], $J_T(v)$ can be interpreted as the total energy of such a system in the time interval $[0, T]$ if it moves with the velocity $v$. Indeed, $f$ is then the support function of some closed convex subset $C$ containing the origin (therefore Argmin $f$ is the normal cone to $C$ at 0), $f(v(t))$ is the dissipated power and $\langle v'(t), v(t)\rangle$ is the kinetic power.

Now let us define the (convex) cost function for an infinite horizon $J : K(u_0) \to \mathbb{R}_+ \cup \{+\infty\}$ by

$$J(v) := \lim_{T\to+\infty} J_T(v) = \sup_{T>0} J_T(v)$$

**Theorem 3.3.** (Variational principle, variant of [3]). *$u$ is the unique minimizer of $J$ on $K(u_0)$ and the minimum value is nul.*

**Proof.** Clearly $J(u) = 0$ and $J(v) = 0$ if and only if $J_T(v) = 0$ for all $T > 0$ i.e. $v$ satisfies the differential inclusion on $[0, T]$ for all $T > 0$ i.e. $v = u$. □

The following result will be of importance in section 5.

**Proposition 3.4.**

$$\forall v \in K(u_0), \quad J(v) \geq \frac{1}{2}\|v_\infty - u_\infty\|^2.$$

**Proof.** From the definition of a subgradient and as $-u' \in \partial f(u) \Leftrightarrow u \in \partial f^*(-u')$, we have

$$f(v) - f(u) + f^*(-v') - f^*(-u') \geq -\langle u', v - u\rangle + \langle u, u' - v'\rangle$$

Therefore, adding $\langle v', v\rangle - \langle u', u\rangle$ to both sides, we obtain

$$\forall T > 0, \quad J_T(v) = J_T(v) - J_T(u) \geq \lim_{\delta\to 0}\int_\delta^T \langle v' - u', v - u\rangle dt = \frac{1}{2}\|v(T) - u(T)\|^2$$

Then use the weak lower semi-continuity of the norm as $T \to +\infty$. □

We end this section by studying some properties of the cost function $J$ that make the optimization problem involved in theorem 3.3 non trivial in the sense that $J$ is finite not only at $u$.

**Proposition 3.5.**
(i) *For all $v \in K(u_0)$ such that $v_\infty \notin \operatorname{Argmin} f$ then $J(v) = +\infty$.*
(ii) *If $f$ is coercive then for all $x \in \operatorname{Argmin} f$ there exists $v$ in $K(u_0)$ such that $v_\infty = x$ and $J(v) < +\infty$.*

**Proof.** (i) By (weak and strong) lower semi-continuity of $f$ and $f^*$, we get

$$\liminf_{t\to+\infty}[f(v(t)) + f^*(-v'(t)) + \langle v'(t), v(t)\rangle] \;\geq\; f(v_\infty) + f^*(0) \;>\; 0$$

where the last inequality follows from $0 \notin \partial f(v_\infty)$. Evidently, this inequality implies $J(v) = +\infty$.
(ii) Let $0 < \delta < \tau$. For $x \in X$, let us define the function $v$ from $[0, +\infty[$ into $X$ by

$$v(t) := \begin{cases} u(t) & 0 \leq t \leq \delta \\ u(\delta) + (t-\delta)\dfrac{x - u(\delta)}{\tau - \delta} & \delta \leq t \leq \tau \\ x & \tau \leq t \end{cases}$$

Clearly $v$ is in $\mathcal{A}$, $v_\infty = v(\tau) = x$ and $v'(t) = 0$ for all $t > \tau$.
Since $x$ is in $\operatorname{dom} f$ then $v(t)$ is in $\operatorname{dom} f$ for all $t > 0$. Since $f$ is weakly inf-compact *i.e.* $f^*$ is strongly continuous at 0, if $\tau$ is large enough then $-v'(t)$ is in $\operatorname{dom} f^*$ *for a.e.* $t > 0$. Finally, because $f(x) + f^*(0) = 0$ and $f \circ v$ is integrable on $[\delta, \tau]$, we have

$$\forall\ T \geq \tau, \qquad J(v) = J_T(v) = J_\tau(v) = \int_\delta^\tau f(v(t))dt + (\tau - \delta)f^*\left(\frac{u(\delta) - x}{\tau - \delta}\right)$$

$$+\frac{1}{2}(\|x\|^2 - \|u(\delta)\|^2) < +\infty,$$

taking into account that $J_\delta(v) = J_\delta(u) = 0$. □

## 4. Asymptotical variational principle

Let us define the asymptotic cost function $\varphi_{u_0} : X \to \mathbb{R}_+ \cup \{+\infty\}$ by

$$\varphi_{u_0}(x) := \inf\{J(v);\ \ v \in K(u_0),\ \ v_\infty = x\}$$

**Proposition 4.1.** *$\varphi_{u_0}$ is convex;*
$\varphi_{u_0}(u_\infty) = J(u) = 0$;
$\forall x \in X,\ \varphi_{u_0}(x) \geq \frac{1}{2}\|x - u_\infty\|^2$.

**Proof.** Let $x^1, x^2 \in X$, $\theta \in [0,1]$ and $v^i \in K(u_0)$ such that $v^i_\infty = x^i$, $i = 1,2$. As $K(u_0)$ is convex, $v^\theta := \theta v^1 + (1-\theta)v^2 \in K(u_0)$. Moreover $v^\theta_\infty = x^\theta := \theta x^1 + (1-\theta)x^2$. Then

$$\varphi_{u_0}(x^\theta) \leq J(v^\theta) \leq \theta J(v^1) + (1-\theta)J(v^2)$$

As this holds for all $v^i \in K(u_0)$ such that $v^i_\infty = x^i$, we are done for the first statement. The second statement is immediate, the third one comes directly from proposition 3.4. □

Now, as a direct consequence of proposition 4.1, we can state the announced asymptotic variational principle.

**Theorem 4.2.** *$u_\infty$ is the unique minimizer of $\varphi_{u_0}$ on* Argmin $f$ *(or on $X$) and the minimum value is zero.* □

Finally, in order that the optimization problem involved in theorem 4.2 be non trivial, refereing to proposition 3.5, we get

**Proposition 4.3.** *If $f$ is coercive then* $\operatorname{dom} \varphi_{u_0} = \operatorname{Argmin} f$. □

## 5. Discrete case

For the same convex function $f$ there is an intimate relationship between the continuous steepest descent method and its Euler's *implicit* discrete version, *i.e.*, the proximal method of Martinet-Rockafellar [14], [8]. Thanks to this connexion, the proximal method inherits many of the nice asymptotical properties of the continuous steepest descent method [9]. For the problem under consideration here, concerning characterization of the asymptotical limit, a similar discussion for the proximal method is entirely parallel to the one above. The statements are the same, albeit now "time" is discrete.
More precisely, let $X$ and $f$ be as in section 1 and $\{\lambda_k\}$ be a sequence of positive reals. Let $u_0 \in X$ be given. The proximal method [14] generates a sequence $u = \{u_k\} \in X, u_k$ being the unique solution of the iterative scheme

$$\frac{u_{k-1} - u_k}{\lambda_k} \in \partial f(u_k) \quad \forall\ k \geq 1.$$

Thus this method coincides with the Euler's *implicit* discretization of the differential inclusion arising in the continuous steepest descent (section 1). So, for all $n \in \mathbb{N}$, $u_n$ approximates the continuous steepest descent trajectory at the point $t_n := \sum_1^n \lambda_k$.
Let us recall some basic known results about the asymptotic behaviour.

**Proposition 5.1.** [14] *If $\lambda_k$ is bounded away from 0 and* Argmin $f \neq \emptyset$ *then $u_k$ weakly converges to $u_\infty$ some minimizer of $f$.*

Note that here the asymptotic limit $u_\infty$ may be different from that one of the continuous case (for the same $u_0$).

**Remark 5.2.** In proposition 5.1, the convergence holds true for the norm topology if $f$ is even [4] or if $\operatorname{Int}\operatorname{Argmin} f \neq \emptyset$ [9] or if $f$ is well-posed [10].

Proposition 2.1 holds true replacing $u(t)$ by $u_n$ and $t$ by $t_n := \sum_1^n \lambda_k$ ([6]), implying $\lim_{n\to+\infty} f(u_n) = \inf_X f$ and the same localization results as in corollary 2.2.

Now, all statements in sections 3 and 4 still hold replacing $\mathcal{A}$ by $X^{\mathbb{N}}$ and with the following new definitions.

**Set of discrete feasible trajectories from $u_0$:**

$$K(u_0) := \{v = (v_0, v_1, \cdots);\ \ \forall\ \ k \geq 1$$

$$v_k \in \operatorname{dom} f,$$

$$-d_k := \frac{v_{k-1} - v_k}{\lambda_k} \in \operatorname{dom} f^*,$$

$$v_0 := u_0,\ \ v_k \overset{w}{\rightharpoonup} v_\infty,\ \ \|d_k\| \to 0\ \}$$

**Cost function for a finite horizon:**

$$J_n(v) := \sum_{k=1}^{n} \lambda_k [f(v_k) + f^*(-d_k) + \langle v_k, d_k\rangle]$$

$$= \sum_{k=1}^{n} \lambda_k [f(v_k) + f^*(-d_k) + \frac{1}{2}\lambda_k \|d_k\|^2]$$

$$+\frac{1}{2}(\|v_n\|^2 - \|u_0\|^2)$$

**Cost function for an infinite horizon:**

$$J(v) := \lim_{n\to+\infty} J_n(v) = \sup_n J_n(v)$$

**Asymptotic cost function:** exactly the same as in section 4.

The adaptation of proofs is left out as a simple exercise. The crucial trick is

$$2\langle u_{k-1} - u_k, x - u_k\rangle = \|u_{k-1} - u_k\|^2 + \|x - u_k\|^2 - \|x - u_{k-1}\|^2$$

For the proof of the analogue of proposition 3.5 (ii), define $v \in K(u_0)$ by:

$$v_k := \begin{cases} u_k & 0 \leq k \leq 1 \\ u_1 + (t_k - t_1)\dfrac{x - u_1}{t_N - t_1} & 1 < k \leq N \\ x & k > N \end{cases}$$

where $N > 1$ large enough in order that, for $1 < k \leq N$, $-d_k = \frac{u_1 - x}{t_N - t_1}$ be in $\operatorname{dom} f^*$ since $\lim_{n\to+\infty} t_n = +\infty$.

**Acknowledgment.** The author would like to thank the referees for their valuable remarks and suggestions for improving the presentation.

## References

[1] H. Attouch: Viscosity methods. Application to Mechanics and Mathematical Programming, Working paper Université Montpellier II Mathématiques, 1993.

[2] H. Brezis: Opérateurs maximaux monotones, Lecture Note no. 5, North-Holland, 1973.

[3] H. Brezis, I. Ekeland: Un principe variationnel associé à certaines équations paraboliques. Le cas indépendant du temps, C.R. Acad. Sc. Paris, t. 282 (3 mai 1976) Série A, 971–974.

[4] H. Brezis, P.L. Lions: Produits infinis de résolvantes, Israël Journal of Mathematics, Vol 29, no. 4, 1978.

[5] R.E Bruck: Asymptotic Convergence of Nonlinear Contration Semigroups in Hilbert Space, Journal of Functional Analysis 18 (1975) 15–26.

[6] O. Güler: On the convergence of the proximal point algorithm for convex minimization, SIAM J. Control and Optimization, Vol 29, no. 2 (1991) 403–419.

[7] B. Lemaire: Régularisation et pénalisation en optimisation convexe, Séminaire d'Analyse Convexe, Institut de Mathématiques USTL, exposé no. 21, 1971.

[8] B. Lemaire: The proximal algorithm, in New Methods in Optimization and their Industrial Uses, Proc. Symp. Pau and Paris 1987, Int. Ser. Numer. Math., Vol 87, Birkhäuser Verlag (1989) 73–89.

[9] B. Lemaire: About the convergence of the proximal method, in Advances in Optimization, Proceedings Lambrecht 1991, Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag (1992) 39–51.

[10] B. Lemaire: Bounded Diagonally Stationary Sequences in Convex Optimization, Journal of Convex Analysis, Heldermann Verlag Berlin, 1 (1994) 75–86.

[11] J.J Moreau: Sur les lois de frottement, de plasticité et de viscosité, C.R. Acad. Sci. Paris, Sér. A, 271 (1970) 608–611.

[12] J.J Moreau: Sur l'évolution d'un système élasto-visco-plastique, C.R. Acad. Sci. Paris, Sér. A, 273 (1971) 118–121.

[13] J.J Moreau: Application of Convex Analysis to the treatment of elastoplastic systems, in Proceedings Symposium IUTAM/IMU, edited by P.Germain & B.Nayroles, Lecture Notes in Mathematics, vol. 503, Springer-Verlag (1976) 56–89.

[14] R.T. Rockafellar: Monotone Operators and the proximal point algorithm, SIAM J. Control and Optimisation (1976) 877–898.

[15] A. Tikhonov, V. Arsenine: Méthodes de résolution de problèmes mal posés, M.I.R, 1974.