Noname manuscript No.
(will be inserted by the editor)

Hedy Attouch · **Jérôme Bolte**

On the convergence of the proximal algorithm for nonsmooth functions involving analytic features

the date of receipt and acceptance should be inserted later

Abstract We study the convergence of the proximal algorithm applied to nonsmooth functions that satisfy the Łojasiewicz inequality around their generalized critical points. Typical examples of functions complying with these conditions are continuous semialgebraic or subanalytic functions. Following Łojasiewicz's original idea, we prove that any bounded sequence generated by the proximal algorithm converges to some generalized critical point. We also obtain convergence rate results which are related to the flatness of the function by means of Łojasiewicz exponents. Apart from the sharp and elliptic cases which yield finite or geometric convergence, the decay estimates that are derived are of the type $O(k^{-s})$, where $s \in (0,+\infty)$ depends on the flatness of the function.

Keywords Proximal algorithm · Łojasiewicz inequality · subanalytic functions

Mathematics Subject Classification (2000) 90C26 · 47N10 · 90C30

1 Introduction

The proximal algorithm has been first introduced by Martinet (1970) [21] and Rockafellar (1976) [28] as an approximation-regularization method in convex optimization and in the study of variational inequalities associated to maximal monotone operators. In the last decades, it has been successfully applied to a wide variety of situations: convex optimization (see for instance [28, 18] and references therein), nonmonotone operators [31, 22, 15, 27, 10, 23] with various applications

Hedy Attouch
I3M UMR CNRS 5149, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier, France. E-mail: attouch@math.univ-montp2.fr

Jérôme Bolte
Equipe Combinatoire et Optimisation (UMR 7090), Case 189, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris Cedex 05, France. E-mail: bolte@math.jussieu.fr

to nonconvex programming. Recent progress in the modelling of decision processes in economics and decision sciences (procedural rationality) provide extra motivation to develop further the study of the proximal algorithm in a nonconvex and possibly nonsmooth setting. Our main concern in the present paper is to present new results in a fairly general setting, namely by considering real-analytic functions, and more generally subanalytic (see [12]) lower semicontinuous functions. First, let us explain briefly the need to go beyond the classical convex setting and then, why analyticity features come naturally in the picture.

In [2] and [3], Attouch and Soubeyran developed a model for "real life" decision making which is an incremental decision process "A worthwhile to move approach of satisficing with not too much sacrificing". In this discrete dynamical model involving both exploration and exploitation aspects, the agent moves from a performance x^k to x^{k+1} when the estimated marginal gain $u(x^{k+1}) - u(x^k)$ is greater than, or equal to, the cost of moving $c(x^k, x^{k+1})$. In this context, optimization features of the decision process are naturally modelled by the proximal algorithm (described below with the maximization version),

$$x^{k+1} \in \operatorname{argmax}\{u(x) - c(x^k, x) : x \in X\}.$$

Classical proximal algorithms correspond to quadratic costs, i.e., $c(x,y) = |x-y|^2$ which expresses that small changes are costless. Because of the cost to change, this process becomes of local nature, which makes it more realistic than the classical global optimization modelling in economics and decision sciences.

The function u measures the quality of the decision or performance $x \in X$, it is the utility function in economics, the valence in cognitive sciences. The opposite function $f = -u$, (which is now to minimize), measures how far is the current performance from a given long term goal. Indeed, the concavity of u (convexity of f) is a too restrictive assumption in order to cover many interesting applications: for example, in economics the utility function u is usually assumed to be quasiconcave. The convergence of the proximal algorithm for quasiconvex functions has been considered only recently, see Goudou-Munier [13] , Attouch-Teboulle [4].

The proximal algorithm can be viewed as an implicit discretization of the continuous steepest descent method (also called continuous gradient method). This important fact has been soon recognized by many authors. It is at the origin of various developments which have been enriching the original algorithm and make it a powerful tool. A striking example is the link between interior point methods, proximal methods associated to Bregman functions, the Riemannian steepest descent and the Lotka-Volterra dynamical systems, see [4] and the references therein. Our special interest for the proximal method for functions involving analytic features comes from the recent developments concerning the convergence of the steepest descent method by Simon [30], Haraux [14] and Bolte-Daniilidis-Lewis [7]. In this last paper, the authors consider the case of subanalytic lower semicontinuous functions. This class of functions is very interesting because it covers many relevant problems in optimization and decision sciences (recall that, by the Stone-Weierstrass theorem, polynomials of several variables and hence analytic functions are dense in the space of continuous functions for the topology of the uniform convergence on bounded sets). A key tool in the mathematical analysis of such continous or discrete dynamical systems is the Łojasiewicz inequality.

It has been first stated by Łojasiewicz in the case of real-analytic functions [20] and, more recently, extended to nonsmooth functions [7].

Our main result (Theorem 1) relies precisely on a judicious use of the Łojasiewicz inequality and proves the convergence of the proximal algorithm to a critical point of the function to which it is applied (f or u). Based on Łojasiewicz's original idea [20] this type of results has already been applied successfully to explicit gradient method for analytic functions [1]. Our main result is completed by studying the rate of convergence of the algorithm (Theorem 2). This rate depends on the value of the so called Łojasiewicz exponents which can be thought as local measures of the flatness of functions around their generalized critical points.

2 The proximal algorithm

2.1 Preliminaries

The Euclidean scalar product of $\mathbb{R}^n$ and its corresponding norm are respectively denoted by $\langle\cdot,\cdot\rangle$ and $|\cdot|$.

Let us recall a few definitions concerning subdifferential calculus.

Definition 1 ([29,9,25]) Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be a proper lower semicontinuous function.

(i) The domain of f, written dom f, is the subset of $\mathbb{R}^n$ on which f is finite-valued.

(ii) For each $x\in$ dom f, the Fréchet subdifferential of f at x, written $\hat\partial f(x)$, is the set of vectors $x^*\in\mathbb{R}^n$ which satisfy

$$\liminf_{\substack{y\neq x\\ y\to x}} \frac{1}{|x-y|}\left[f(y)-f(x)-\langle x^*,y-x\rangle\right]\geq 0.$$

If $x\notin$ dom f, then $\hat\partial f(x)=\emptyset$.

(iii) The limiting-subdifferential ([24]) of f at $x\in\mathbb{R}^n$, written ∂f, is defined as follows

$$\partial f(x):=\{x^*\in\mathbb{R}^n:\exists x_n\to x, f(x_n)\to f(x), x_n^*\in\hat\partial f(x_n)\to x^*\}.$$

Remark 1 The above definition implies that $\hat\partial f(x)\subset\partial f(x)$, where the first set is convex while the second one is closed.

Remark 2 Clearly a necessary condition for $x\in\mathbb{R}^n$ to be a minimizer of f is

$$\partial f(x)\ni 0. \tag{1}$$

Unless f is convex the above is not a sufficient condition. In the remainder, a point $x\in\mathbb{R}^n$ that satisfies (1) is called *limiting-critical* or *critical*. The set of critical points of f is denoted by crit f.

2.2 Proximal algorithm

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Given $x^0 \in \mathbb{R}^n$ we consider the following discrete dynamical system

$$x^{k+1} \in \text{argmin}\,\{f(u) + \frac{1}{2\lambda_k}|u - x^k|^2 : u \in \mathbb{R}^n\}, \tag{2}$$

where $(\lambda_k)_{k\in\mathbb{N}}$ is a positive sequence.

Necessary and sufficient conditions for this algorithm to be well-defined can be found in Rockafellar-Wets [29, Exercise 1.24., p. 20]. We simply assume here that

$$(\mathscr{H}_1) \quad \inf_{\mathbb{R}^n} f > -\infty,$$

which clearly implies that, for all $k \in \mathbb{N}$, the set appearing in (2) is nonempty and compact. Writing down the optimality condition [29, Theorem 10.1] and using the subdifferentiation formula for a sum of functions [29, Exercise 10.10]), it follows that there exists $g^{k+1} \in \partial f(x^{k+1})$ such that

$$x^{k+1} = x^k - \lambda_k g^{k+1}. \tag{3}$$

Let us fix some positive parameters λ_- and λ_+ with $0 < \lambda_- < \lambda_+ < +\infty$.

From now on we assume that $\boxed{\lambda_k \in (\lambda_-, \lambda_+)}$ for all $k \in \mathbb{N}$.

Consider the following assumption:

$(\mathscr{H}_2)$ The restriction of f to its domain is a continuous function (on dom f).

The following result gathers a few elementary facts concerning the dynamical system (2).

Proposition 1 *Let $(x^k)_{k\in\mathbb{N}}$ be a sequence which complies with (2) and denote by $\omega(x^0)$ the set of its limit points. Then*

(i) The sequence $(f(x^k))_{k\in\mathbb{N}}$ is nonincreasing,

(ii) $\sum |x^{k+1} - x^k|^2 < +\infty$,

(iii) If f satifies $(\mathscr{H}_2)$ then $\omega(x^0) \subset \text{crit} f$.

If, in addition, the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded then

(iv) $\omega(x^0)$ is a nonempty compact connected set, and

$$d(x^k, \omega(x^0)) \to 0 \text{ as } k \to +\infty,$$

(v) If f satifies $(\mathscr{H}_2)$ then f is finite and constant on $\omega(x^0)$.

Sketch of the proof Let us prove (i) and (ii). By definition, (2) implies that for all $k \geq 0$ we have

$$f(x^{k+1}) + \frac{1}{2\lambda_k}|x^{k+1} - x^k|^2 \leq f(x^k). \tag{4}$$

This means that $f(x^k)$ is nonincreasing and by summing the inequalities (4) from 0 to $N \geq 0$ we also obtain that

$$\sum_{k=0}^{N} |x^{k+1} - x^k|^2 \leq 2\lambda_+[f(x^0) - f(x^{N+1})] \leq 2\lambda_+[f(x^0) - \inf_{\mathbb{R}^n} f] < \infty.$$

Let us deal with (iii) and (v). For any limit point $\bar{x}$ of f, we can use the lower semicontinuity of f to obtain that $\lim_{k\to\infty} f(x^k) \geq f(\bar{x})$. If, in addition, f satisfies $(\mathscr{H}_2)$ then the above inequality is actually an equality and (v) is proved. By using (ii), (3) and the fact that $\lambda_k \geq \lambda_- > 0$ we can assume that there exists $k_p \to +\infty$ such that $\{(x^{k_p}, g^{k_p})\} \to (\bar{x}, 0)$ with $f(x^{k_p}) \to f(\bar{x})$. Owing to the definition of the limiting subdifferential it follows that $(\bar{x}, 0)$ belongs to the graph of ∂f, so (iii) is proved.

Item (iv) follows by using (ii) together with some classical properties of sequences in $\mathbb{R}^n$.

Remark 3 Even when x^k is bounded, the convergence of the whole sequence x^k may fail even for a finite-valued smooth function f, see Palis-De Melo [26] or Absil-Mahony-Andrews [1].

3 Convergence analysis

3.1 Łojasiewicz inequality and examples

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function that satisfies $(\mathscr{H}_2)$. The function f is said to have the Łojasiewicz property if:

$(\mathscr{H}_3)$ For any limiting-critical point $\hat{x}$, that is $\hat{x} \in \operatorname{crit} f$, there exist $C, \varepsilon > 0$ and $\theta \in [0,1)$ such that

$$|f(x) - f(\hat{x})|^\theta \leq C|x^*|, \forall x \in B(\hat{x}, \varepsilon), \forall x^* \in \partial f(x). \tag{5}$$

Remark 4 When $\theta = 0$ we adopt the convention $0^0 = 0$, and therefore if $|f(x) - f(\hat{x})|^0 = 0$ we have $f(x) = f(\hat{x})$.

Lemma 1 *Assume that f has the Łojasiewicz property.*

(i) If K is a connected subset of the set of critical points of f, then f is constant on K.

(ii) If in addition K is a compact set, then there exist $C, \varepsilon > 0$ and $\theta \in [0,1)$ such that

$$\forall x \in \mathbb{R}^n,\ d(x,K) \leq \varepsilon,\ \forall x^* \in \partial f(x), \quad |f(x) - f(\hat{x})|^\theta \leq C|x^*|. \tag{6}$$

Proof Item (i) is a straightforward consequence of (ii), let us therefore deal with (ii). The compact set K can be covered by a finite number of open balls $B(x_i, \varepsilon_i)$, with $x_i \in K$ ($i = 1, \ldots, p$) on which (5) holds with constants C_i, θ_i. In other words, for each $i \in \{1, \ldots, p\}$ and for each $x \in B(x_i, \varepsilon_i)$ we have

$$|f(x) - f(x_i)|^{\theta_i} \leq C_i |x^*|$$

for all x^* in $\partial f(x)$. As a consequence, f is locally constant (and continuous) on the connected set K, it is therefore constant. By choosing $\varepsilon > 0$ sufficiently small, we obtain that

$$\{x \in \mathbb{R}^n : d(x, K) \leq \varepsilon\} \subset \cup_{i=1}^{p} B(x_i, \varepsilon_i),$$

and the claimed result holds by letting $C = \max C_i$ and $\theta = \max \theta_i$.

Let us give several examples in which the above results can be applied.

Example 1 (a) Real-analytic functions have the Łojasiewicz property, see Łojasiewicz [20].
(b) An interesting class of functions satisfying the Łojasiewicz property is given by *semialgebraic* functions. These are functions whose graphs can be expressed as

$$\bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in \mathbb{R}^n : P_{ij}(x) = 0, Q_{ij}(x) > 0\}, \tag{7}$$

where for all $1 \leq i \leq p, 1 \leq j \leq q$ the $P_{ij}, Q_{ij} : \mathbb{R}^n \mapsto \mathbb{R}$ are polynomial functions. Due to the Tarski-Seidenberg principle [6,5] -which asserts that the linear projection of a set of the type (7) remains of this type- semialgebraic objects enjoy remarkable stability properties.

Let us illustrate these remarks by a nonsmooth semialgebraic minimization problem. Let $g : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ be a polynomial function and let K be a compact semialgebraic subset of $\mathbb{R}^p$, *i.e.* of the form (7). Then

$$f(x) = \max_{y \in K} g(x, y)$$

is a (lower-C^2) locally Lipschitz continuous semialgebraic function which satisfies the Łojasiewicz inequality (see [7] for the details). Let $c_i : \mathbb{R}^n \to \mathbb{R}, i \in \{1, \ldots, m\}$ be a family of polynomial (or semialgebraic) functions, set

$$C := \{x \in \mathbb{R}^n : c_i(x) \leq 0, \forall i \in \{1, \ldots, m\}\},$$

and introduce the indicator function of C, *i.e.* the extended-real-valued function $i_C : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by $i_C(x) = 0$ if $x \in C$, $i_C(x) = +\infty$ otherwise.

The global minimization problem

$$(\mathscr{P}) \quad \min\{f(x) : x \in C\} = \min_{x \in C} \max_{y \in K} g(x, y),$$

is a semialgebraic problem in the sense that the function $f + i_C$ is semialgebraic. This minmax problem occurs in various domains like game theory, shape optimization or mechanics. To simplify this example we can assume that C is a nonempty bounded set. The proximal method (2) can be applied to $(\mathscr{P})$ as follows

$$x^{k+1} \in \operatorname{argmin} \{f(u) + i_C(u) + \frac{1}{2}|u - x^k|^2 : u \in \mathbb{R}^n\}.$$

In view of Theorem 1, each sequence x^k generated by the above recursion is bounded and converges to a critical point of $f + i_C$.
(c) Convex functions satisfying the following growth conditions

$$\forall \hat{x} \in \operatorname{argmin} f, \exists C > 0, r \geq 1, \varepsilon > 0, \forall x \in B(\hat{x}, \varepsilon),\ f(x) \geq f(\hat{x}) + Cd(x, \operatorname{argmin} f)^r$$

comply with (5) (with $\theta = 1 - \frac{1}{r}$), see [7]. This is in particular the case of strongly convex functions, that is proper lower semicontinuous functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ which satisfy

$$\langle x^* - y^*, x - y \rangle \geq k|x - y|^2 \text{ for all } (x, x^*), (y, y^*) \in \operatorname{graph} \partial f,$$

where $k > 0$ and $\operatorname{graph} \partial f := \{(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n : x^* \in \partial f(x)\}$.
(d) Infinite-dimensional versions of (5) have been developed in view of the asymptotic analysis of dissipative evolution equations. These can be found in Simon [30], and Haraux [14].
(e) Kurdyka has recently established a Łojasiewicz-like inequality for functions definable in an arbitrary o-minimal structure [16]. O-minimal structures have been introduced in view of working with sets and functions which enjoys the qualitative properties of semialgebraic sets. A good introduction to o-minimal structures is Coste [11]. In a forthcoming paper we shall tackle the convergence issues of the proximal algorithm in that framework.

3.2 Convergence results

The proofs we develop here are adapted from Łojasiewicz's original idea [20].

Theorem 1 (Convergence result) *Assume that f satisfies $(\mathcal{H}_1)$, $(\mathcal{H}_2)$, $(\mathcal{H}_3)$ and let $(x^k)_{k\in\mathbb{N}}$ be a sequence generated by the proximal algorithm.*
If the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded, then

$$\sum_{k=0}^{+\infty} |x^{k+1} - x^k| < +\infty,$$

in particular the whole sequence $(x^k)_{k\in\mathbb{N}}$ converges to some critical point of f.

Proof Changing f into $f - \inf_{k\geq 0} f(x^k)$ we can assume with no loss of generality that $f(x^k)$ converges to 0. The case when $x^{k+1} = x^k$ for some $k \geq 1$ has no consequence on the asymptotic analysis, so that we may suppose that $|x^{k+1} - x^k| > 0$ for all $k \geq 0$. In view of (4), we obtain also that $f(x^k)$ is positive and decreases (strictly) to 0.

By using the convex inequality for the function $s > 0 \to -s^{1-\theta}$ and (4) for all $k \geq 0$ we obtain that

$$\begin{aligned} f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta} &\geq (1-\theta) f(x^k)^{-\theta} (f(x^k) - f(x^{k+1})) \\ &\geq (1-\theta) f(x^k)^{-\theta} \frac{1}{2\lambda_k} |x^{k+1} - x^k|^2. \end{aligned} \tag{8}$$

By Proposition 1 (iii) and (iv), and Lemma 1 (take $K = \omega(x_0)$) there exist an integer N_0, real numbers C and $\theta \in (0,1)$ such that

$$0 < |f(x^k)|^\theta \leq C|g^k| = \frac{C}{\lambda_k}|x^k - x^{k-1}|$$

for all $k \geq N_0$.
Combining the above result with (8) (recall $\lambda_k > \lambda_- > 0$) yields the existence of a positive constant M such that

$$\frac{|x^{k+1} - x^k|^2}{|x^k - x^{k-1}|} \leq M(f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta}) \tag{9}$$

for all $k \geq N_0$.

Fix $r \in (0,1)$ and take $k \geq N_0$. If $|x^{k+1} - x^k| \geq r|x^k - x^{k-1}|$, (9) implies that

$$|x^{k+1} - x^k| \leq \frac{M}{r}[f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta}],$$

and thus we have for all $k \geq N_0$

$$|x^{k+1} - x^k| \leq r|x^k - x^{k-1}| + \frac{M}{r}[f(x^k)^{1-\theta} - f(x^{k+1})^{1-\theta}].$$

If $N \geq N_0$ an easy induction yields

$$\sum_{k=N_0}^{N} |x^{k+1} - x^k| \leq \frac{r}{1-r}|x^{N_0} - x^{N_0 - 1}| + \frac{M}{r(1-r)}[f(x^{N_0})^{1-\theta} - f(x^{N+1})^{1-\theta}], \tag{10}$$

and the conclusion follows from the fact that f is bounded from below. □

Remark 5 Similar convergence results could be obtained for functions belonging to some o-minimal structure, see Kurdyka [16] and references therein.

Remark 6 It would be interesting to compare the results of Theorem 1 with those obtained in Combettes-Pennanen [10] under a cohypomonotonicity assumption. Indeed if ∂f happens to be cohypomonotone [10, Definition 2.2] a variant of the proximal algorithm (see [10, Algorithm 1.1]) can be applied giving rise to local convergence results. In our specific framework we obtain however stronger results since the orbit generated by algorithm (2) has a finite length.

If $(x^k)_{k\in\mathbb{N}}$ is a bounded sequence generated by (2), let us denote by x^∞ its (unique) limit point. Applying Proposition (1) (iii) and (5) we obtain the existence of a neighborhood around x^∞ such that (5) holds. The number θ appearing in (5) is then called *a Łojasiewicz exponent of* x^∞.

Theorem 2 (Rate of convergence) *Assume that f satisfies $(\mathscr{H}_1)$, $(\mathscr{H}_2)$, $(\mathscr{H}_3)$ and let $(x^k)_{k\in\mathbb{N}}$ be a sequence generated by the proximal algorithm. Assume that $(x^k)_{k\in\mathbb{N}}$ is* bounded *and denote by θ a Łojasiewicz exponent of x^∞. The following estimations hold*

(i) If $\theta=0$, the sequence $(x^k)_{k\in\mathbb{N}}$ converges in a finite number of steps,

(ii) If $\theta\in(0,\frac{1}{2}]$ then there exist $c>0$ and $Q\in[0,1)$ such that

$$|x^k-x^\infty|\le c\,Q^k,$$

(iii) If $\theta\in(\frac{1}{2},1)$ then there exists $c>0$ such that

$$|x^k-x^\infty|\le c\,k^{-\frac{1-\theta}{2\theta-1}}.$$

Proof The notations are those of the previous proof. For any $k\ge 0$, set $\Delta_k=\sum_{p=k}^{\infty}|x^{p+1}-x^p|$ which is finite by Theorem 1. The triangle inequality yields $\Delta_k\ge|x^k-x^\infty|$, it is therefore sufficient to establish the estimations appearing in (ii) and (iii) for Δ_k. With no loss of generality we may assume that $\Delta_k>0$ for all $k\ge 0$.

Using (10), and the fact that $f(x^k)$ decreases to zero we obtain for k sufficiently large (recall that $r\in(0,1)$ can be taken arbitrarily)

$$\begin{aligned}\Delta_k &\le \frac{1}{1-r}(\Delta_{k-1}-\Delta_k)+\frac{M}{r(1-r)}f(x^k)^{1-\theta}\\ &\le \frac{1}{1-r}(\Delta_{k-1}-\Delta_k)+\frac{M}{r(1-r)}(C|g^k|)^{\frac{1-\theta}{\theta}} \qquad (11)\\ &\le \frac{1}{1-r}(\Delta_{k-1}-\Delta_k)+(\lambda_-)^{1-1/\theta}\frac{MC^{\frac{1-\theta}{\theta}}}{r(1-r)}(\Delta_{k-1}-\Delta_k)^{\frac{1-\theta}{\theta}} \qquad (12)\end{aligned}$$

where (11) and (12) follow respectively from (5) and (3).

Assume that θ belongs to $(\frac{1}{2},1)$, so that $\frac{1-\theta}{\theta}<1$. Since $\Delta_k\to 0$ as $k\to\infty$, we deduce from (12) that there exist an integer $N_1\ge N_0$ and a positive constant C_1 such that

$$\Delta_k^{\frac{\theta}{1-\theta}}\le C_1(\Delta_{k-1}-\Delta_k), \qquad (13)$$

for all $k\ge N_1$. Define $h:(0,+\infty)\to\mathbb{R}$ by $h(s)=s^{-\frac{\theta}{1-\theta}}$ and let $R\in(1,+\infty)$. Take $k\ge N_1$ and assume first that $h(\Delta_k)\le R\,h(\Delta_{k-1})$. By rewriting (13) as

$$1\le\frac{C_1(\Delta_{k-1}-\Delta_k)}{\Delta_k^{\frac{\theta}{1-\theta}}},$$

we obtain that

$$\begin{aligned}1&\le C_1(\Delta_{k-1}-\Delta_k)h(\Delta_k)\\ &\le RC_1(\Delta_{k-1}-\Delta_k)h(\Delta_{k-1})\\ &\le RC_1\int_{\Delta_k}^{\Delta_{k-1}}h(s)ds\\ &\le RC_1\frac{1-\theta}{1-2\theta}[\Delta_{k-1}^{\frac{1-2\theta}{1-\theta}}-\Delta_k^{\frac{1-2\theta}{1-\theta}}].\end{aligned}$$

Thus if we set $\mu = \frac{2\theta-1}{(1-\theta)RC_1} > 0$ and $\nu = \frac{1-2\theta}{1-\theta} < 0$ one obtains that

$$0 < \mu \leq \Delta_k^\nu - \Delta_{k-1}^\nu. \tag{14}$$

Assume now that $h(\Delta_k) > Rh(\Delta_{k-1})$ and set $q = (\frac{1}{R})^{\frac{1-\theta}{\theta}} \in (0,1)$. It follows immediately that $\Delta_k \leq q\Delta_{k-1}$ and furthermore - recalling that ν is negative - we have

$$\begin{aligned} \Delta_k^\nu &\geq q^\nu \Delta_{k-1}^\nu \\ \Delta_k^\nu - \Delta_{k-1}^\nu &\geq (q^\nu - 1)\Delta_{k-1}^\nu. \end{aligned}$$

Since $q^\nu - 1 > 0$ and $\Delta_p \to 0^+$ as $p \to +\infty$, there exists $\bar{\mu} > 0$ such that $(q^\nu - 1)\Delta_{p-1}^\nu > \bar{\mu}$ for all $p \geq N_1$. Therefore we obtain that

$$\Delta_k^\nu - \Delta_{k-1}^\nu \geq \bar{\mu}. \tag{15}$$

If we set $\hat{\mu} = \min\{\mu, \bar{\mu}\} > 0$, one can combine (15) and (14) to obtain that

$$\Delta_k^\nu - \Delta_{k-1}^\nu \geq \hat{\mu} > 0$$

for all $k \geq N_1$. By summing those inequalities from N_1 to some N greater than N_1 we obtain that $\Delta_N^\nu - \Delta_{N_1}^\nu \geq \hat{\mu}(N - N_1)$ and consequently (iii) follows from

$$\Delta_N \leq [\Delta_{N_1}^\nu + \hat{\mu}(N - N_1)]^{1/\nu} \leq cN^{-\frac{1-\theta}{2\theta-1}} \text{ (c being a positive constant).}$$

When $\theta \in (0, \frac{1}{2}]$, (12) shows that Δ_k complies with the following inequality (for k sufficiently large)

$$\Delta_k \leq C_2(\Delta_{k-1} - \Delta_k),$$

where C_2 is a positive constant. This implies that $\Delta_k \leq \frac{C_2}{1+C_2}\Delta_{k-1}$ and item (ii) follows easily with $Q = \frac{C_2}{1+C_2} \in (0,1)$.

Assume that $\theta = 0$, set $I := \{k \in \mathbb{N} : x_{k+1} \neq x_k\}$ and take k in I. If k is sufficiently large we have

$$\frac{1}{\lambda_k^2}|x^{k+1} - x^k|^2 = |g^{k+1}|^2 \geq C_3 > 0,$$

so that (4) implies that

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2\lambda_k}|x^{k+1} - x^k|^2 \leq f(x^k) - C_3\frac{\lambda_-}{2}.$$

Since $f(x_k)$ is known to converge to zero the above inequality clearly implies that I is finite and (i) follows immediately. □

Remark 7 When f is convex and satisfies the assumptions of Theorems 1 and 2, which means in particular that f has at least one minimizer, any sequence x^k has a *finite length* in the sense that $\sum |x^{k+1} - x^k| < \infty$. This makes contrast with the classical results for general convex functions [18,28], for which the convergence of the trajectories of the proximal algorithm hold, but with only a square summability property, *i.e.* $\sum |x^{k+1} - x^k|^2 < \infty$.

A criterion to determine if a convex function satisfies the Łojasiewicz property has been given in Example 1 (c), but in practice an easier assumption to check is the semialgebraicity (or the subanalyticity) of the utility function (see [7]). In that case, if f has a minimizer it satisfies automatically the assumptions of Theorem 1 so that the proximal method enjoy the properties described above. To our knowledge the finite length property and the rate estimates given in Theorem 2 are new even in this convex setting. Of course, similar results hold for quasiconvex functions.

References

1. Absil, P.-A., Mahony, R., Andrews, B. : Convergence of the Iterates of Descent Methods for Analytic Cost Functions. SIAM J. Optim. **16**, 531-547 (2005)
2. Attouch, H., Soubeyran, A.: Inertia and reactivity in decision making as cognitive variational inequalities. J. of Convex Analysis **13**, (2006)
3. Attouch, H., Soubeyran, A.: From procedural rationality to routines: a worthwhile to move approach of satisficing with not too much sacrificing. Submitted paper, (2005)
4. Attouch, H., Teboulle, M.: Regularized Lotka-Volterra Dynamical system as continuous proximal-like method in optimization. J. Optim. theory Appl. **121**, 541-580 (2004)
5. Benedetti, R. & Risler, J.-J.: Real Algebraic and Semialgebraic Sets. Hermann, Éditeurs des sciences et des Arts, Paris (1990)
6. Bochnak, J., Coste, M., Roy, M.-F.: Real Algebraic Geometry, Springer (1998)
7. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. Submitted for publication (2004)
   Available electronically at: `http://pareto.uab.es/~adaniilidis/`
8. Bolte, J., Daniilidis, A., Lewis, A.: The Morse-Sard theorem for non-differentiable subanalytic functions. To appear in J. Math. Analysis & Applications (2005)
9. Clarke, F.H., Ledyaev, Yu., Stern, R.I., Wolenski, P.R.: Nonsmooth analysis and control theory. Graduate texts in Mathematics **178**, Springer-Verlag, New-York (1998)
10. Combettes, P., Pennanen, T.: Proximal methods for cohypomonotone operators. SIAM J. Control Optim. **43**, 731-742 (2004)
11. Coste, M.: An introduction to o-minimal geometry. RAAG Notes, 81 p., Institut de Recherche Mathématiques de Rennes, (1999)
12. van den Dries, L. Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**, 497-540 (1996)
13. Goudou, X., Munier, J.: The heavy ball with friction method: the quasiconvex case. Submitted (2005)
14. Haraux, A.: A hyperbolic variant of Simon's convergence theorem. Evolution equations and their applications in physical and life sciences (Bad Herrenalb, 1998), Lecture Notes in Pure and Appl. Math. **215**, 255-264, Dekker-New York (2001)
15. Kaplan, A., Tichatschke, R.: Proximal point methods and nonconvex optimization. J. Glob optimization **13**, 389-406 (1998)
16. Kurdyka, K.: On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier **48**, 769-783 (1998)
17. Kurdyka, K., Parusinski, A.: w_f-stratification of subanalytic functions and the Łojasiewicz inequality. C. R. Acad. Paris **318**, 129-133 (1994)

18. Lemaire, B.: The proximal algorithm. New methods in optimization and their industrial uses, Pau/Paris, (1987) 73-87, Internat. Schriftenreihe Numer. Math. **87**, Birkhäuser, Basel, (1989)
19. LEWIS A.S., Active sets, nonsmoothness and sensitivity. SIAM Journal on Optimization, 13:702725, 2003.
20. Łojasiewicz, S.: "Une propriété topologique des sous-ensembles analytiques réels.". In Ĺes Équations aux Dérivées Partielles, Editions du Centre National de la Recherche Scientifique, Paris, 87-89 (1963)
21. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. (French) Rev. Française Informat. Recherche Opérationnelle **4**, Ser. R-3 154-158 (1970)
22. Miettinen, M., Mkel, M.M., Haslinger, J.: On numerical solution of hemivariational inequalities by nonsmooth optimization methods. J. Glob. Optimization **6**, 401-425 (1995)
23. Mifflin, R., Sagastizabal, C.: $\mathscr{V}\mathscr{U}$-smoothness and proximal point results for some nonconvex functions. Optim. Methods Softw. **19**, 463-478 (2004)
24. Mordukhovich, B.: Maximum principle in the problem of time optimal response with nonsmooth constraints. J. Appl. Math. Mech. **40**, 960-969 (1976); [translated from Prikl. Mat. Meh. **40**, 1014-1023 (1976)]
25. Mordukhovich, B.: Variational analysis and generalized differentiation. Grundlehren der Mathematischen, Wissenschaften, **330**, Springer (1998)
26. Palis, J., De Melo, W.: Geometric theory of dynamical systems. An introduction (Translated from the Portuguese by A. K. Manning). Springer-Verlag, New York-Berlin (1982)
27. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**, 170-191 (2002)
28. Rockafellar, R. T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97-116 (1976)
29. Rockafellar, R.T., Wets, R.: Variational Analysis. Grundlehren der Mathematischen, Wissenschaften, **317**, Springer (1998)
30. Simon, L.: Asymptotics for a class of non-linear evolution equations, with applications to geometric problems. Ann. Math. **118**, 525-571 (1983)
31. Spingarn, J.E.: Submonotone mappings and the proximal point algorithm. Numer. Funct. Anal. Optim. **4**, 123-150 (1981)
32. WRIGHT S.J. Identifiable surfaces in constrained optimization. SIAM Journal on Control and Optimization, 31:10631079, 1993.