A Generalized Inexact Proximal Point Method for Nonsmooth Functions that Satisfies Kurdyka Lojasiewicz Inequality

G. C. Bento* A. Soubeyran †

Abstract

In this paper, following the ideas presented in Attouch et al. (Math. Program. Ser. A, 137: 91-129, 2013), we present an inexact version of the proximal point method for nonsmoth functions, whose regularization is given by a generalized perturbation term. More precisely, the new perturbation term is defined as a "curved enough" function of the quasi distance between two successive iterates, that appears to be a nice tool for Behavioral Sciences (Psychology, Economics, Management, Game theory, . . .). Our convergence analysis is a extension, of the analysis due to Attouch and Bolte (Math. Program. Ser. B, 116: 5-16, 2009) or, more generally, to Moreno et al. (Optimization, 61:1383-1403, 2011), to an inexact setting of the proximal method which is more suitable from the point of view of applications. We give, in a dynamic setting, a striking application to the famous Nobel Prize Kahneman and Tversky [11], Tversky and Kahneman [10] "loss aversion effect" in Psychology and Management. This application shows how the strength of resistance to change can impact the speed of formation of an habituation/routinization process.

AMS Classification. 49J52 · 49M37 · 65K10 · 90C30 · 91E10.

*IME-Universidade Federal de Goiás, Goiânia-GO 74001-970, BR (glaydston@ufg.br)

†Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, FR (antoine.soubeyran@gmail.com).

keywordsNonconvex optimizationKurdyka-Lojasiewicz inequalityInexact proximal algorithmsSpeed convergence

1 Introduction

The origins of the proximal method can be traced back to the 1960s as an approximation-regularization algorithms in convex optimization as well as in the study of variational inequalities associated to maximal monotone operators; see Moreau [12], Martinet [13] and Rockafellar [14]. In this paper, we show when, in a quasi metric space, and a Kurdyka-Lojasiewicz inequality for the objective function, an inexact proximal algorithm with a generalized perturbation term converges to a critical point, depending of the curvature of the pertubation term. Our convergence analysis is an extension of the analysis due to Attouch and Bolte [7] or, more generally, due to Moreno et al. [9], to an inexact setting of the proximal method, which is more suitable from the point of view of applications because, in the exact case, the proximal algorithm would transform an optimization problem in an infinite sequence of optimization problems. Actually, this is one usual criticism for exact proximal algorithms which is even stronger for behavioral applications. Our main convergence result is completed by the analysis of the rate converge, which is a natural extension of [7, Theorem 2]. We examine, as a striking application, how the strength of resistance to change, which is modelized by the generalized perturbation term, can influence the speed of convergence of an habituation/routinization process. This offers a striking example of the famous "loss aversion effect" in a dynamic context, when resistance to change is low enough.

The motivation to introduce a generalized regularization term comes from a recent approach in Behavioral Sciences, the Variational rationality (VR) approach (see Soubeyran [1, 2]), which helps to unify a lot of stay/stability and change dynamics (as course of human activities) in different disciplines. Stays and stability refer to habits, routines, equilibrium, traps,.... Changes represent, among so many things, creations, destructions, innovations, the formation and revision of attitudes and beliefs, a lot of self regulation problems including goal setting, goal striving and goal revision, habit formation, breaking and forming habits, learning and knowledge management,....

Central to this VR approach are the original behavioral concepts of "worthwhile changes", "non marginal worthwhile changes", and "variational traps", see [1, 2]). Worthwhile changes balance, each period, motivation and resistance to change (the utility of advantages to change and the disutility of inconvenients to change). In the context of inexact proximal algorithms, these worthwhile and non marginal worthwhile changes represent, each round, sufficient descent conditions as well as an internal stopping rule. This very simple idea that worthwhile changes and non worthwhile marginal changes are core concepts to modelize the human course of actions was the essential step to see the perturbation term of a proximal algorithm as a crude formulation of the complex concept of resistance to change, while advantages to change refer to a change in the objective function.

Then, the main message of this paper is that, using the behavioral context of the "Variational rationality approach" [1, 2], a generalized proximal algorithm can modelize fairly well, in Psychology, for an agent, an habituation process as the end of a learning process, and, In Management Sciences, for an organization, a routinization process, as the end of a knowledge management process. This is the case even when resistance to change (inertia) is weak (our paper). The easier case where resistance to change is strong requires less mathematical structure and has been examined in [1, 2]). This opens the door to a new vision of proximal algorithms. They are not only very nice mathematical tools in Optimization theory, with striking computational aspects. They can also be nice tools to modelize the dynamics of human behaviors. This variational approach have provided an extra motivation to develop further the study of proximal algorithms in a nonconvex and possibly nonsmooth setting (for some references, that include the proximal point method for certain nonconvex minimization problems see, for example, [16, 18, 17, 19, 20, 21, 22]). Among recent applications of this simple idea, see Attouch and Soubeyran [3] for local search proximal algorithms, Flores-Bazan et al. [4] for worthwhile to change games, alternating inertial games with costs to move, Attouch et al. [5] and Cruz Neto et al. [6], for the "how to play Nash" problem,.... In all these papers the perturbation term of the usual proximal point algorithm is a linear or a quadratic function of the distance or quasi distance between two successive iterates. They modelize the case of "strong enough resistance" to change. Our paper examines the opposite case of "weak enough" resistance to change, where the pertubation term modelizes the difficulty (relative resistance) to be able to change as a "curved enough" function of the quasi distance between two successive iterates. A quasi distance modelizes costs to be able to change as an index of dissimilarity between actions, where the cost to be able to change from an action to an other one is not the same as the cost to be able to change in the other way.

In the present paper, the end of an inexact proximal algorithm process is a critical point. For applications it would be useful to characterize situations where critical points are variational traps (see [1, 2]), "easy enough to reach" and "difficult enough to leave". These traps can represent the end of a succession of worthwhile temporary stays and changes, where, each period, the agent can prefer to stay for a while and then prefers to change rather than to stay again, and, at the end, prefers to stay than to change. This have been done in a second paper, Bento and Soubeyran [41]. The authors show how such a generalized inexact proximal algorithm can help to modelize, in Psychology, habit formation for an agent and, in Management Sciences, routine formation and knowledge management for an organization. In this second paper, using the context of the variational rationality approach, the authors give sufficient conditions for critical points to be variational traps.

Our paper is organized as follows. Section 2 presents the formulation of our generalized proximal point method, where the perturbation term is given by a "curved enough" function of the quasi distance between two successive iterates. Section 3 examines a generalized inexact proximal algorithm which converges to a critical point, when the objective function satisfies a Kurdyka-Lojasiewicz inequality. This section considers also the speed convergence of this process. Section 4 examines how the strength of resistance to change (which generalizes the loss aversion concept) can influence the speed of convergence of an habituation/routinization process. The conclusion follows.

2 A Generalized Proximal Formulation

In this section we have made explicit a formulation for generalized proximal algorithms where the perturbation term of the usual proximal point algorithm is replaced by a "curved enough" function of the quasi distance between two successive iterates.

Consider the following minimization problem

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function bounded from below.

Definition 2.1. *A mapping $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is said to be a quasi distance iff, for all $x, y, z \in \mathbb{R}^n$,*

i) $q(x, y) = q(y, x) = 0$ if, only if, $x = y$;

ii) $q(x, y) \leq q(x, z) + q(z, y)$,

Given $x \in \mathbb{R}^n$ and $\epsilon > 0$ fixed, we denote by $B_q(x, \epsilon)$ the open ball, with respect to the quasi distance q, of center x and radius $\epsilon > 0$, defined as follows: $B_q(x, \epsilon) = \{y \in M : q(x, y) < \epsilon\}$. In particular, if q is the Euclidean distance, $B_q(x, \epsilon)$ will be denoted $B(x, \epsilon)$.

Throughout the paper q represents a quasi distance that satisfy the following assumption:

Assumption 2.1. *There exist $\beta_1, \beta_2 \in \mathbb{R}_{++}$ such that: $\beta_1 \|x - y\| \leq q(x, y) \leq \beta_2 \|x - y\|$, $x, y \in \mathbb{R}^n$.*

In [9] the authors present several examples of quasi distances highlighting two that satisfy Assumption 2.1.

For each $x \in \mathbb{R}^n$ given, $y \in \mathbb{R}^n$ is said be worthwhile change iff there exists $\lambda(x) > 0$ such that the motivation to change from the status x to new position or action y instead of staying at x, given by $(f(x) - f(y))/\lambda(x)$, is greater that, or equal, to the resistance for change given by $\Gamma(q(x, y))$, namely,

$$f(x) - f(y) \geq \lambda(x)\Gamma(q(x, y)),$$

where Γ be a twice differentiable function such that:

$$\Gamma[0] = \Gamma'[0] = 0, \quad \text{and} \quad \Gamma'[q] > 0, \quad \Gamma''[q] > 0, \quad q > 0, \tag{2}$$

and there exist constants $r, \bar{q}, \bar{\rho}_\Gamma(r) > 0$, satisfying the following condition:

$$\Gamma'[q/r] \leq \bar{\rho}_\Gamma(r)\Gamma[q]/q, \qquad 0 < q \leq \bar{q}. \tag{3}$$

Note that for different choices of Γ and q, the worthwhile to change condition which defines the worthwhile to change set is a descent condition; see, for instance, [7, 9, 8]. This variational approach provides us an extra motivation to develop further the study of proximal algorithms where the perturbation term of the usual proximal point algorithm becomes a "curved enough" function of the quasi distance between two successive iterates. More precisely, given $x^0 \in \mathbb{R}^n$ and a bounded sequence of positive real numbers $\{\lambda_k = \lambda(x^k)\}$ (called regularization parameters), the next step is such that

$$x^{k+1} \in \operatorname{argmin}_{y \in \mathbb{R}^n} \left\{ f(y) + \lambda_k \Gamma[q(x^k, y)] \right\}. \tag{4}$$

Let us consider a generalized rate of curvature of Γ given by:

$$\rho_\Gamma(q, r) := \frac{\Gamma'[q/r]}{(\Gamma[q]/q)}, \qquad 0 < q \leq \bar{q}. \tag{5}$$

In the particular case $r = 1$, (5) represents, in Economics, the elasticity of the desutility curve Γ; see, for instance, [1, 2]. From (5), condition (3) is equivalent to the condition:

$$\bar{\rho}_\Gamma(r) = \sup\{\rho_\Gamma(q, r) : 0 < q < \bar{q}\} < +\infty, \qquad r \in]0, 1[\quad \text{fixed}.$$

For each $\alpha > 1$ fixed, consider the function $\Gamma[q] := q^{\alpha}$. It is easy to see that, in this case, $\bar{\rho}_D(r) \in [\alpha r^{1-\alpha}, +\infty)$. In particular, we can take

$$\bar{\rho}_\Gamma(q, r) = \alpha r^{1-\alpha} = \bar{\rho}_\Gamma(r) < +\infty. \tag{6}$$

More accurately, for each $\alpha > 1$, $\Gamma[q] = q^{\alpha}$ represents a desutility of costs to change. It is strictly increasing and satisfies (2) and (3). Note that, when q is the Euclidean distance and $\Gamma[q] = q^2/2$, then (4) retrieves the classical proximal method. The Method (4) is new and more adapted for applications in Behavioral Sciences. Moreover, it retrieves formulations that have recently been considered for nonconvex functions, see [7, 9].

3 An Inexact Descent Method for KL Functions: Convergence to a Critical Point

In this section, following the ideas presented in [8], we propose and study an inexact version of the proximal method (4) whose full convergence is assured for objective functions that satisfy the Kurdyka-Lojasiewicz inequality. This approach includes an inexact version of the proximal point method studied in [9].

3.1 Some Definitions from Subdifferential Calculus

In this section some elements concerning the subdifferential calculus are recalled; see, for instance, [15, 23].

From the assumption H1., $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function. The domain of f, which we denote by $\mathrm{dom} f$, is the subset of $\mathbb{R}^n$ on which f is finite-valued. Since f is proper, then $\mathrm{dom} f \neq \emptyset$.

Definition 3.1.

i) The Fréchet subdifferential of f at $x \in \mathbb{R}^n$, denoted by $\hat{\partial} f(x)$, is the set given by:

$$\hat{\partial} f(x) := \begin{cases} \{x^* \in \mathbb{R}^n : \liminf_{y \to x; y \neq x} \frac{1}{\|x-y\|}(f(y) - f(x) - \langle x^*, y-x \rangle) \geq 0\}, & if\ x \in domf, \\ \emptyset, & if\ x \notin domf. \end{cases}$$

ii) The limiting Fréchet subdifferential (or simply subdifferential) of f at $x \in \mathbb{R}^n$, denoted by $\partial f(x)$, is the set given by:

$$\partial f(x) := \begin{cases} \{x^* \in \mathbb{R}^n | \exists x_n \to x, f(x_n) \to f(x), x_n^* \in \hat{\partial} f(x_n); x_n^* \to x^*\}, & if\ x \in domf. \\ \emptyset, & if\ x \notin domf. \end{cases}$$

In our convergence analysis, as well as in any limiting processes used in an algorithmic context, we need consider a subdifferential T of f that satisfies the following closedness property:

Property 3.1. *Let $\{(x^k, v^k)\}$ be a sequence in $Gr(T)$. If (x^k, v^k) converges to $(x, v) \in \mathbb{R}^n \times \mathbb{R}^n$ and $f(x^k)$ converges to $f(x)$, then $(x, v) \in Gr(T)$.*

In this sense, throughout the paper we consider the subdifferential ∂f. A necessary condition for a given point $x \in \mathbb{R}^n$ to be a minimizer of f is

$$0 \in \partial f(x). \tag{7}$$

It is known that, unless f is convex, (7) is not a sufficient condition. In the remainder, a point that satisfies (7) is called limiting-critical or simply critical.

Proposition 3.1. *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ be functions such that f_1 is locally Lipschitz continuous at $\bar{x} \in \mathbb{R}^n$ while f_2 is proper lower semicontinuous, with $f_2(\bar{x})$ finite. Then, $\partial(f_1 + f_2)(\bar{x}) \subset \partial f_1(\bar{x}) + \partial f_2(\bar{x})$.*

Proof. See [15, pages 350,431]. □

Proposition 3.2. *Let $f_1 : \mathbb{R}_+ \to \mathbb{R}_+$ be continuously differentiable, $f_2 : \mathbb{R}^n \to \mathbb{R}_+$ a locally Lipschitz function at $\bar{x} \in \mathbb{R}^n$. Then, $\partial(f_1 \circ f_2)(\bar{x}) = f_1'(f_2(\bar{x}))\partial f_2(\bar{x})$, for each $x \in [0 < f_2 < 1]$.*

Proof. See [24, Lemma 43]. □

3.2 The algorithm

Let us consider the proximal point method (4) and assume that f is bounded below. Since f is proper lower semicontinuous, from the first inequality in (2) and (3) combined with Assumption 2.1, it is easy to see that this method is well-defined. In particular, $x^{k+1} \in \mathrm{dom} f$, for all $k \in \mathbb{N}$. From the definition of x^{k+1}, we have

$$f(x^{k+1}) + \lambda_k \Gamma[q(x^k, x^{k+1})] \leq f(x^k), \qquad k = 0, 1, \ldots. \tag{8}$$

Let $\{x^k\}$ be a sequence generated by (4) and assume that:

$$\{x^k\} \subset \mathrm{dom} f, \qquad x^k \neq x^{k+1}, \qquad 0 < \bar{\lambda} \leq \lambda_k \leq \tilde{\lambda}, \qquad k \in \mathbb{N}.$$

As $x^k \neq x^{k+1}$, $k \in \mathbb{N}$, combining definition of q with definition of Γ, it follows that

$$\Gamma[q(x^k, x^{k+1})] > 0, \qquad k = 0, 1 \ldots.$$

Hence, the inequality (8) implies that $\{f(x^k)\}$ is strictly decreasing. Moreover, since $\lambda_k \geq \bar{\lambda} > 0$ and f is bounded below, using again inequality (8), it is easy to see that $\{\Gamma[q(x^k, x^{k+1})]\}$ is a summable sequence. In particular, taking into account that Γ is a continuous function, we conclude that there exists $k_0 \in \mathbb{N}$ sufficiently large such that

$$q(x^k, x^{k+1}) < 1, \qquad k \geq k_0. \tag{9}$$

On the other hand, from the definition of x^{k+1} together with optimality condition (7), it follows that

$$0 \in \partial(f + \Gamma[q(x^k, \cdot)])(x^{k+1}), \qquad k = 0, 1, \ldots. \tag{10}$$

Note that Γ is a continuously differentiable function and $q(x^k, \cdot)$ is a Lipschitz function, for each $k \in \mathbb{N}$ (see [9, Proposition 3.6]). In particular, $\Gamma[q(x^k, \cdot)]$ is a locally Lipschitz function. So, applying Proposition 3.1 with $f_1 = \Gamma[q(x^k, \cdot)]$ and $f_2 = f$, from inclusion (10), we obtain:

$$0 \in \partial f(x^{k+1}) + \lambda_k \partial \Gamma[q(x^k, \cdot)](x^{k+1}). \tag{11}$$

Now, since there exists $k_0 \in \mathbb{N}$ such that (9) holds, using Proposition 3.2 with $f_1 = \Gamma$ and $f_2 = q(x^k, \cdot)$, we conclude that there exist $w^{k+1} \in \partial f(x^{k+1})$ and $v^{k+1} \in \partial q(x^k, \cdot)(x^{k+1})$ such that:

$$0 = w^{k+1} + \lambda_k \Gamma'[q(x^k, x^{k+1})]v^{k+1}, \qquad k \geq k_0. \tag{12}$$

As mentioned previously, this version of the proximal point method is new and generalizes the methods considered in [7, 9]. However, here we are interested in an inexact version of (4) characterized from the two conditions (8) and (12). For inexact versions in the convex case see, for instance, Rockafellar [14], Solodov and Svaiter [25, 26] and references therein. In the particular case where $\Gamma[q] = q^2/2$, $q(x, y) = \|x - y\|$ and f is convex, in [26] the authors consider that $(x^{k+1}, w^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n$ is an inexact solution, with tolerance $\sigma \in [0, 1[$, for the subproblem (11) if:

$$w^{k+1} \in \partial f(x^{k+1}), \qquad \tilde{\varepsilon}^k = w^{k+1} + \lambda_k(x^{k+1} - x^k), \tag{13}$$

and

$$\varepsilon_k := \|\tilde{\varepsilon}^k\| \leq \sigma \max\{\|w^{k+1}\|, \lambda_k\|x^{k+1} - x^k\|\}. \tag{14}$$

Note that (14) implies the following weaker condition:

$$\exists\, b > 0 : \ \ \|w^{k+1}\| \leq b\|x^{k+1} - x^k\|, \tag{15}$$

which coincides with the inexact optimality condition proposes in [8].

Next, we introduce an inexact version of the proximal point method (4):

Algorithm 3.1. *Take $x^0 \in domf$, $0 < \bar{\lambda} \leq \tilde{\lambda} < +\infty$, $\sigma \in [0, 1[$ and $b > 0$. For each $k = 0, 1, \ldots$, choose $\lambda_k \in [\bar{\lambda}, \tilde{\lambda}]$ and find $(x^{k+1}, w^{k+1}, v^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ such that:*

$$f(x^k) - f(x^{k+1}) \geq \lambda_k(1 - \sigma)\Gamma[q(x^k, x^{k+1})], \tag{16}$$

$$w^{k+1} \in \partial f(x^{k+1}), \quad v^{k+1} \in \partial q(x^k, \cdot)(x^{k+1}), \tag{17}$$

$$\|w^{k+1}\| \leq b\Gamma'[q(x^k, x^{k+1})]\|v^{k+1}\|, \tag{18}$$

Remark 3.1. *Note that Algorithm 3.1 is in fact an inexact version of the proximal method* (4) *that retrieves the inexact algorithm proposed in [8, Algorithm 2] in the particular case* $\Gamma[q] = q^2/2$, $q(x, y) = \|x - y\|$ *and* $1 - \sigma = \theta$.

If $\{x^k\}$ is a sequence finitely generated by Algorithm 1, from the definition of q together with the first inequality in (2) and (18), it is easy to see that it terminates at a critical point. Unless stated to the contrary, in the remainder of this paper we assume that $\{x^k\}$ is an infinite sequence generated by Algorithm 1 and f is bounded from below and continuous on $\mathrm{dom} f$.

3.3 Convergence and speed of convergence

Next we present a partial convergence result for Algorithm 1. We observe that in [9, Lemma 5.1] the proof presented by the authors holds for any bounded sequence $\{y^k\}$ and any function Ψ_k which is locally Lipschitz for each $k \in \mathbb{N}$. Hence, taking into account that $q(x^k, \cdot)$ is a locally Lipschitz function, for each $k \in \mathbb{N}$ (see [9, Proposition 3.6]), in the particular case where $\{x^k\}$ is a bounded sequence and $\Psi_k = q(x^k, \cdot)$, it follows that the sequence $\{v^k\}$ defined in Algorithm 1 is bounded. In the remainder of this paper we assume that there exists $L > 0$ such that

$$\|v^k\| \leq L, \qquad k \in \mathbb{N}. \tag{19}$$

Proposition 3.3. *The following statements hold:*

i) *The sequence* $\{f(x^k)\}$ *is strictly increasing;*

ii) $\sum_{k=0}^{+\infty} \Gamma[q(x^k, x^{k+1})] < +\infty$;

iii) $\lim_{k\to+\infty} q(x^k, x^{k+1}) = 0$;

iv) Each accumulation point of the sequence $\{x^k\}$, if any, is a critical point of f.

Proof. The proof of the items *i*), *ii*) and *iii*) are of simple verification (see what was presented at the beginning of Section 3.2). Let us deal with item *iv*). Suppose that $\bar{x} \in \mathbb{R}^n$ be an accumulation point of $\{x^k\}$ and let $\{x^{k_j}\}$ be a subsequence converging to $\bar{x}$. The item *i*) implies that $\bar{x} \in \text{dom} f$ and $f(x^{k_j})$ converges to $f(\bar{x})$. Since $\{x^k\}$ is generated by Algorithm 1, there exist sequences $\{w^k\}$ and $\{v^k\}$ such that $w^{k+1} \in \partial f(x^{k+1})$ and $v^{k+1} \in \partial(q(x^k, \cdot)(x^{k+1})$ satisfying (18). Now, $\{v^k\}$ being bounded, take $\bar{v} \in \mathbb{R}^n$ and assume that $\{v^{k_j}\}$ is a subsequence of $\{v^k\}$ converging to $\bar{v}$. Hence, taking into account that $\{\lambda_k\}$ is bounded, inequality (18) allows us to conclude that $\{w^k\}$ has a subsequence converging to zero. Without loss of generality we can assume that $\{w^{k_j}\}$ converges to zero. Therefore, as ∂f satisfies Property 3.1, the proof is complete. □

As in [7, 9, 8], our main convergence result is restricted to functions that satisfy the so-called Kurdyka-Lojasiewicz inequality. This was first introduced by Lojasiewicz [27], to real analytic functions, and extended by Kurdyka [29] to differentiable definable functions in an o-minimal structure (for a detailed discussion on o-minimal structures see, for example, Dries and Miller [28]). For extensions of the Kurdyka-Lojasiewicz inequality, in the Euclidean context, to the classe of nonsmooth functions see Bolte et al. [30], Bolte et al. [31] and Attouch et al. [32]. For extensions of the Kurdyka-Lojasiewicz property to functions defined on non-linear spaces, see Kurdyka et al. [33], Lageman [34], Bolte et al. [24] and Bento et al. [35]. Next formal definition of the Kurdyka-Lojasiewicz inequality can be finding in [32], where it is also possible to find several examples and a good discussion over important classes of functions which satisfy the mentioned inequality.

Definition 3.2. *A proper lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to have the Kurdyka-Lojasiewicz property at $\bar{x} \in$ dom ∂f if there exists $\eta \in]0, +\infty]$, a neighborhood U of $\bar{x}$ and a continuous* *concave function* $\varphi : [0, \eta[\to \mathbb{R}_+$ *such that:*

$$\varphi(0) = 0, \quad \varphi \in C^1(0, \eta), \quad \varphi'(s) > 0, \quad s \in]0, \eta[; \tag{20}$$

$$\varphi'(f(x) - f(\bar{x})) dist(0, \partial f(x)) \geq 1, \quad x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta], \tag{21}$$

- $dist(0, \partial f(x)) := inf\{\|v\| : v \in \partial f(x)\}$,
- $[\eta_1 < f < \eta_2] := \{x \in M : \eta_1 < f(x) < \eta_2\}, \quad \eta_1 < \eta_2$.

In the remainder of this paper we assume that f is a KL function, i.e., a function which satisfies the Kurdyka-Lojasiewicz inequality at each point of dom∂f.

Theorem 3.1. *Assume that* $\{x^k\}$ *is bounded,* $\tilde{x} \in \mathbb{R}^n$ *is an accumulation point of* $\{x^k\}$ *and Assumption 2.1 holds. Let* $U \subset \mathbb{R}^n$ *be a neighborhood of* $\tilde{x}$*,* $\eta \in]0, +\infty]$ *and* $\varphi : [0, \eta[\to \mathbb{R}_+$ *a continuous concave function such that* (20) *and* (21) *hold. If* $\delta \in (0, \bar{q})$ *(see condition* (3)*) and* $r \in]0, 1[$ *are fixed constants,* $B(\tilde{x}, \delta/\beta_1) \subset U$*,* $a := \bar{\lambda}(1 - \sigma)$ *and* $M := \frac{Lb}{a}$*, then there exists* $k_0 \in \mathbb{N}$ *such that:*

$$f(\tilde{x}) < f(x^k) < f(\tilde{x}) + \eta, \qquad k \geq k_0, \tag{22}$$

$$q(\tilde{x}, x^{k_0}) + \frac{1}{1-r}\Gamma^{-1}\left[\frac{f(x^{k_0}) - f(\tilde{x})}{a}\right] + \frac{r}{1-r}\Gamma^{-1}\left[\frac{f(x^{k_0-1}) - f(\tilde{x})}{a}\right] + \frac{\bar{\rho}_\Gamma(r)M}{1-r}\left[\varphi(f(x^{k_0}) - \varphi(f(\tilde{x}))\right] < \delta, \tag{23}$$

$$x^{k_0+j} \in B_q(\tilde{x}, \delta), \qquad j = 1, 2, \ldots. \tag{24}$$

Moreover,

$$\sum_{k=0}^{+\infty} q(x^k, x^{k+1}) < +\infty. \tag{25}$$

In particular, the whole sequence $\{x^k\}$ *converges to* $\tilde{x}$ *and it is a critical point of* f*.*

Proof. From item $i)$ of Proposition 3.3, it follows that $x^k, \tilde{x} \in$ dom$f$, for all $k \in \mathbb{N}$. In particular, the sequence $\{f(x^k)\}$ is well defined and converges to $\inf_{k\geq 0} f(x^k) = f(\tilde{x})$. Changing f into $f - \inf_{k\geq 0} f(x^k)$ we can assume, without loss of generality, that $f(\tilde{x}) = 0$. Since $\{f(x^k)\}$ is a decreasing sequence converging to 0, we have

$$0 < f(x^k), \qquad k \in \mathbb{N}. \tag{26}$$

In particular, there exists $N \in \mathbb{N}$ such that

$$0 < f(x^k) < \eta, \qquad k \geq N. \tag{27}$$

Since (26) holds, let us define the sequence $\{b_k\}$ given by

$$b_k = q(x^k, \tilde{x}) + \frac{1}{1-r}\Gamma^{-1}\left[\frac{f(x^k)}{a}\right] + \frac{r}{1-r}\Gamma^{-1}\left[\frac{f(x^{k-1})}{a}\right] + \frac{\bar{\rho}_\Gamma(r)M}{1-r}\varphi(f(x^k)).$$

As $q(\cdot, \tilde{x})$, Γ^{-1} and φ are continuous functions, $\varphi(0) = 0$ and $\{f(x^k)\}$ converges to 0, it follows that 0 is an accumulation point of the sequence $\{b_k\}$. Hence, there exists $k_0 := k_{j_0} > N$ such that (23) holds. In particular, as $k_0 > N$, from (27) it also holds (22).

Note that (23) implies $x^{k_0} \in B_q(\tilde{x}, \delta)$. On the other hand, from Assumption 2.1 there exists $\beta_1 > 0$ such that $q(\tilde{x}, x^{k_0}) \geq \beta_1 \|x^{k_0} - \tilde{x}\|$. This tell us that $x^{k_0} \in B(\tilde{x}, \delta/\beta_1) \subset U$ (latter inclusion follows by hypothesis). Hence, from (22), we have

$$x^{k_0} \in U \cap [0 < f < \eta].$$

Then, since $\tilde{x}$ is a point where f satisfies the Kurdyka-Lojasiewicz inequality it follows that $0 \notin \partial f(x^{k_0})$. Moreover, conditions (19) and (18) combined with the definition of $dist(0, \partial f(x^k))$, yield

$$b\Gamma'[q(x^{k-1}, x^k)]L \geq b\Gamma'[q(x^{k-1}, x^k)]\|v^k\| \geq \|w^k\| \geq dist(0, \partial f(x^k)), \qquad k = 1, 2, \ldots.$$

Hence, again from the Kurdyka Lojasiewicz inequality of f at $\tilde{x}$, it follows that

$$\varphi'(f(x^{k_0})) \geq \frac{1}{bL\Gamma'[q(x^{k-1}, x^k)]}. \tag{28}$$

On the other hand, the concavity of the function φ implies that

$$\varphi(f(x^{k_0})) - \varphi(f(x^{k_0+1})) \geq \varphi'(f(x^{k_0}))(f(x^{k_0}) - f(x^{k_0+1})),$$

which, combined with $\varphi' > 0$ and condition (16) (taking into account that $\lambda_k \geq \bar{\lambda} = \frac{a}{1-\sigma}$, $k \in \mathbb{N}$) yields

$$\varphi(f(x^{k_0})) - \varphi(f(x^{k_0+1}) \geq \varphi'(f(x^{k_0}))a\Gamma[q(x^{k_0}, x^{k_0+1})], \tag{29}$$

So, using the inequalities (29) and (28), we get

$$M\left[\varphi(f(x^{k_0})) - \varphi(f(x^{k_0+1}))\right] \geq \frac{\Gamma[q(x^{k_0}, x^{k_0+1})]}{\Gamma'[q(x^{k_0-1}, x^{k_0})]}. \tag{30}$$

We state that

$$q(x^k, x^{k+1}) \leq rq(x^{k-1}, x^k) + \bar{\rho}_\Gamma(r)M[\varphi(f(x^k)) - \varphi(f(x^{k+1}))], \tag{31}$$

holds for $k = k_0$. Indeed, we have two possibilities to consider:

a) $q(x^{k_0}, x^{k_0+1}) \geq rq(x^{k_0-1}, x^{k_0})$;

b) $q(x^{k_0-1}, x^{k_0}) > rq(x^{k_0}, x^{k_0+1})$.

Let us suppose that holds *a*). Then, $q(x^{k_0-1}, x^{k_0}) \leq q(x^{k_0}, x^{k_0+1})/r$. Hence, from the last inequality in (2), we obtain

$$\Gamma'[q(x^{k_0-1}, x^{k_0})] \leq \Gamma'[q(x^{k_0}, x^{k_0+1})/r] \leq \bar{\rho}_\Gamma(r)\left[\Gamma[q(x^{k_0}, x^{k_0+1})]/q(x^{k_0}, x^{k_0+1})\right],$$

where the last inequality follows from condition (3). Therefore, (31) follows by combining last inequality with (30), taking into consideration that $rq(x^{k_0}, x^{k_0+1}) \geq 0$. If happens item *b*) the statement is immediately verified.

Let us prove (24) by induction on j. Suppose that $j = 1$. Since Γ is invertible, $f(x^k) > 0$, $\lambda_k \geq \bar{\lambda} = \frac{a}{1-\sigma}$ ($k \in \mathbb{N}$) and $r \in]0, 1[$, condition (16) implies that

$$q(x^{k_0}, x^{k_0+1}) \leq \Gamma^{-1}\left[\frac{f(x^{k_0})}{a}\right] \leq \frac{1}{1-r}\Gamma^{-1}\left[\frac{f(x^{k_0})}{a}\right]. \tag{32}$$

On the other hand, combining the triangle inequality with last inequality and condition (20) we obtain, from (23):

$$q(\tilde{x}, x^{k_0+1}) \leq q(\tilde{x}, x^{k_0}) + q(x^{k_0}, x^{k_0+1}) \leq q(\tilde{x}, x^{k_0}) + \frac{1}{1-r}\Gamma^{-1}\left[\frac{f(x^{k_0})}{a}\right] < \rho.$$

But this tell us that (24) holds with $j = 1$. Take $j \geq 1$ and let us suppose that (24) holds for all $k = k_0 + 1, \ldots, k_0 + j - 1$. In this case, (31) holds for $k = k_0 + 1, \ldots, k_0 + j - 1$ and, hence, we get

$$\sum_{i=0}^{j-1} q(x^{k_0+i}, x^{k_0+i+1}) \leq \frac{r}{1-r} q(x^{k_0-1}, x^{k_0}) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}[\varphi(f(x^{k_0})) - \varphi(f(x^{k_0+j}))]. \tag{33}$$

From the triangle inequality, we have:

$$q(\tilde{x}, x^{k_0+j}) \leq q(\tilde{x}, x^{k_0}) + q(x^{k_0}, x^{k_0+1}) + \sum_{i=0}^{j-1} q(x^{k_0+i}, x^{k_0+i+1}).$$

Thus, combining these two last inequalities with (32) and taking into account the inequality $-\varphi(f(x^{k_0+j})) < 0$, we have

$$q(\tilde{x}, x^{k_0+j}) \leq q(\tilde{x}, x^{k_0}) + \frac{1}{1-r}\Gamma^{-1}\left[\frac{f(x^{k_0})}{a}\right] + \frac{r}{1-r}\Gamma^{-1}\left[\frac{f(x^{k_0-1})}{a}\right] + \frac{\bar{\rho}_\Gamma(r)M}{1-r}[\varphi(f(x^{k_0}))],$$

which, from (23), allows us to conclude the induction proof.

Note that (25) follows immediately from (33). Now, combining Assumption 2.1 with (25), we conclude that $\{x^k\}$ is a Cauchy sequence and hence converges. The conclusion of the proof follows from item *iii*) of Proposition 3.3. □

Under the conditions of the last theorem, for k sufficiently large ($k \geq k_0$), we have

$$\sum_{p=k}^{N} q(x^p, x^{p+1}) \leq \frac{r}{1-r} q(x^{k-1}, x^k) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}[\varphi(f(x^k)) - \varphi(f(x^{N+1}))], \qquad N \geq k. \tag{34}$$

Letting N goes to infinity and taking into account that $f(x^k)$ decreases to zero and $\varphi(0) = 0$, last inequality yields

$$\sum_{p=k}^{+\infty} q(x^p, x^{p+1}) \leq \frac{r}{1-r} q(x^{k-1}, x^k) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}\varphi(f(x^k)), \qquad k \geq k_0. \tag{35}$$

Let us suppose that $\varphi(s) = s^{1-\theta}$, $\theta \in [0,1[$ and, for simplicity of notation, define $\Delta_k := \sum_{k=k_0}^{+\infty} q(x^k, x^{k+1})$. So, considering that $q(x^{k-1}, x^k) = \Delta_{k-1} - \Delta_k$, we can rewrite inequality (35) as it follows:

$$\Delta_k \leq \frac{r}{1-r}(\Delta_{k-1} - \Delta_k) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}(f(x^k))^{1-\theta}, \qquad k \geq k_0. \tag{36}$$

Since $f(\tilde{x}) = 0$ and $x^k \in U \cap [0 < f < \eta]$, for $k \geq k_0$, combining (21) with (36), we obtain

$$\Delta_k \leq \frac{r}{1-r}(\Delta_{k-1} - \Delta_k) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}\left[(1-\theta)\mathrm{dist}(0, \partial f(x^k))\right]^{\frac{1-\theta}{\theta}}, \qquad k \geq k_0. \tag{37}$$

Now, from the definition of the sequence $\{x^k\}$, there exist $b > 0$, $w^k \in \partial f(x^k)$, $v^k \in \partial(q(x^{k-1}, \cdot)(x^k)$ such that

$$\|w^k\| \leq b\Gamma'[q(x^{k-1}, x^k)]\|v^k\| \leq bL\Gamma'[q(x^{k-1}, x^k)], \qquad k \in \mathbb{N},$$

where the last inequality follows from (19). As $\mathrm{dist}(0, \partial f(x^k)) \leq \|w^k\|$, combining last inequality with (37) and taking into account that $q(x^{k-1}, x^k) = \Delta_{k-1} - \Delta_k$, we get

$$\Delta_k \leq \frac{r}{1-r}(\Delta_{k-1} - \Delta_k) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}\left[(1-\theta)bL\Gamma'[\Delta_{k-1} - \Delta_k]\right]^{\frac{1-\theta}{\theta}}, \qquad k \geq k_0. \tag{38}$$

Assumption 3.1. *For each $\beta > 0$ fixed, there exist positive constants $\gamma = \gamma(\beta)$ and $h = h(\beta)$ such that*

$$\Gamma'[q] \leq \gamma q^{h\beta}, \qquad 0 < q < \tilde{q} < 1, \qquad \tilde{q} \quad \textit{fixed.}$$

If Γ is a desutility of costs to change satisfying last assumption with $\beta = \frac{\theta}{1-\theta}$, then there exist $\gamma(\theta), h(\theta) > 0$ such that (38) becomes

$$\Delta_k \leq \frac{r}{1-r}(\Delta_{k-1} - \Delta_k) + \frac{\bar{\rho}_\Gamma(r)M}{1-r}\tilde{C}^{h(\theta)}\left[\Delta_{k-1} - \Delta_k\right]^{h(\theta)}, \qquad k \geq k_0, \tag{39}$$

where $\tilde{C} = (1-\theta)bL\gamma(\theta)$.

Remark 3.2.

a) Given $\beta > 0$, the desutility of costs to change, defined by $\Gamma[q] = q^{\alpha}$, satisfies Assumption 3.1 for $\gamma = \alpha$ and $h = \frac{\alpha-1}{\beta}$;

b) Note that (34) (resp. (39)) is a generalization of (10) (resp. (12)) in [7]. Indeed, this can be easily observed when $\Gamma[d] = d^2$, $q(x,y) = \|y - x\|$ and $\varphi(s) = s^{1-\theta}$ together with (6) ($\alpha = 2$)

Next we present an analysis of the rate convergence in the particular case where the desutility of costs to change D satisfies Assumption 3.1 and $\varphi(s) = s^{1-\theta}$, $\theta \in [0,1[$. This resulted is a generalization of [7, Theorem 2] and its proof follows the same idea presented in the refereed paper.

Theorem 3.2. *Let Γ be a relative resistance to change function satisfying Assumption 3.1, $\varphi(s) = s^{1-\theta}$, $\theta \in [0,1[$. Assume also that all assumptions of the previous theorem hold. Let $\tilde{x}$ be the limit point of the sequence $\{x^k\}$. Then,*

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

ii) If $h(\theta) \geq 1$, then, there exist $c_1 > 0$, $Q \in]0,1[$ and $k_1 \in \mathbb{N}$ such that

$$q(\tilde{x}, x^k) \leq c_1 Q^k, \qquad k \geq k_1;$$

iii) If $h(\theta) \in]0,1[$, then, there exits $c_2 > 0$ such that

$$q(\tilde{x}, x^k) \leq c_2 k^{\frac{h(\theta)}{h(\theta)-1}}, \qquad k \geq k_1.$$

Proof. First, let us suppose that $h(\theta) \in]0,1[$. Since Δ_k converges to zero as k goes to infinity, there exists $k_1 \geq k_0$ such that

$$\Delta_{k-1} - \Delta_k \leq [(\Delta_{k-1} - \Delta_k)]^{h(\theta)}, \qquad k \geq k_1.$$

Hence, (39) shows that there exists a positive constant $\mathcal{C} = \left[\frac{r+\bar{\rho}_{\Gamma}(r)M\tilde{C}^{h(\theta)}}{1-r}\right]^{\frac{1}{h(\theta)}}$, such that

$$\Delta_k^{\frac{1}{h(\theta)}} \leq \mathcal{C}(\Delta_{k-1} - \Delta_k), \qquad k \geq k_1. \tag{40}$$

Define $\psi :]0, +\infty[\to \mathbb{R}$ given by $\psi(t) = t^{-\frac{1}{h(\theta)}}$ and take $R \in]1, +\infty[$. For each $k \geq k_1$ fixed, we state that there exist $\mu > 0$ and $\nu < 0$ such that

$$\Delta_k^{\nu} - \Delta_{k-1}^{\nu} \geq \mu. \tag{41}$$

Indeed, we have two cases to consider:

a) $\psi(\Delta_k) \leq R\psi(\Delta_{k-1})$;

b) $\psi(\Delta_k) > R\psi(\Delta_{k-1})$.

Let us suppose firstly that $a)$ holds. Then, from (40) together with definition of the function h, it follows that

$$1 \leq \mathcal{C}R(\Delta_{k-1} - \Delta_k)\psi(\Delta_{k-1}).$$

Now, since $\psi(\Delta_{k-1}) = \min_{\Delta_k \leq t \leq \Delta_{k-1}} \psi(t)$, last inequality implies

$$1 \leq \mathcal{C}R \int_{\Delta_k}^{\Delta_{k-1}} \psi(t)dt,$$

and, consequently, we have

$$1 \leq \mathcal{C}R \frac{h(\theta)}{h(\theta) - 1} \left[\Delta_{k-1}^{\frac{h(\theta)-1}{h(\theta)}} - \Delta_k^{\frac{h(\theta)-1}{h(\theta)}} \right].$$

Note that $\frac{h(\theta)-1}{h(\theta)} < 0$ and, hence, $\frac{1-h(\theta)}{R\mathcal{C}h(\theta)} > 0$. Therefore, the desired resulted follows from the last inequality with

$$\mu = \frac{1 - h(\theta)}{R\mathcal{C}h(\theta)} \quad \text{and} \quad \nu := \frac{h(\theta) - 1}{h(\theta)}. \tag{42}$$

Let us suppose now that holds $b)$ and define $q = \left(\frac{1}{R}\right)^{h(\theta)}$. Then, from the definition of h, it follows that $\Delta_k \leq q\Delta_{k-1}$ and, taking into account that $\nu < 0$, we get

$$\Delta_k^{\nu} \geq q^{\nu} \Delta_{k-1}^{\nu}.$$

Hence, subtracting Δ_{k-1}^{ν} from both sides of the last inequality, we have

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

Recall that Δ_p converges to zero as p goes to infinity. Without loss of generality, we can suppose that $\Delta_k \in]0,1[$ for $k \geq k_1$ and $\Delta_{k_1}^{\nu} \geq \mu k_1$ (this is possible because $\nu < 0$). Since $q \in]0,1[$ and $\nu < 0$, then, there exists $\bar{\mu} > 0$ such that $(q^{\nu} - 1)\Delta_{k-1} > \bar{\mu}$. This tell us that the desired resulted holds with $\mu = \bar{\mu}$ and ν as defined in (42). Therefore, the statement is proved.

Take $k \in \mathbb{N}$ greater that k_1. By summing the inequality (41) from k_1 to k, we obtain

$$\Delta_k^{\nu} - \Delta_{N_1}^{\nu} \geq \mu(k - k_1),$$

and, hence (since $\nu < 0$),

$$\Delta_k \leq \left[\Delta_{N_1}^{\nu} + \mu(k - k_1)\right]^{\frac{1}{\nu}}.$$

Using again that $\nu < 0$, it follows from the last inequality that there exists a positive constant ω_1, for example, $\omega_1 = \mu^{1/\nu}$, such that

$$\Delta_k \leq \omega_1 k^{\frac{h(\theta)}{h(\theta)-1}}. \tag{43}$$

On the other hand, combining the first inequality in Assumption 2.1 with definition of Δ_k, we have

$$\Delta_k \geq \beta_1 \lim_{N\to+\infty} \sum_{p=k}^{N} \|x^{p+1} - x^p\| \geq \beta_1 \lim_{N\to+\infty} \sum_{p=k}^{N} \left[\|x^p - \tilde{x}\| - \|x^{p+1} - \tilde{x}\|\right] = \beta_1 \|x^k - \tilde{x}\|, \tag{44}$$

where the second inequality and equality follow, respectively, from the triangle inequality and definition of $\tilde{x}$. Thus, item *iii)* follows by combining inequalities (43), (44) and the second inequality in Assumption 2.1 with $c_2 = (\beta_2\omega_1)/\beta_1$.

Suppose now that $h(\theta) \geq 1$. From (39), we get

$$\Delta_k \leq \mathcal{C}_1(\Delta_{k-1} - \Delta_k), \qquad k \geq k_1 \geq k_0,$$

where $\mathcal{C}_1 = \mathcal{C}^{h(\theta)}$. So, last inequality implies $\Delta_k \leq \frac{\mathcal{C}_1}{1+\mathcal{C}_1}\Delta_{k-1}$, for $k \geq k_0$ and, hence,

$$\Delta_k \leq \omega_2 Q^k, \tag{45}$$

where $Q = \frac{\mathcal{C}_1}{1+\mathcal{C}_1} \in]0,1[$ and $\omega_2 = Q^{k_0}\Delta_{k_0}$. Using again (44), item $ii)$ follows with $c_1 = (\beta_2\omega_2)/\beta_1$.

Now, assume that $\theta = 0$ and suppose, for contradiction, that the sequence $\{x^k\}$ is infinitely generated. Take $k_0 \in \mathbb{N}$ sufficiently large such that $x^k \in B_q(\tilde{x}, \delta) \cup [0 < f < \eta]$, $k \geq k_0$, (we are assuming that $f(\tilde{x}) = 0$). Since $\varphi(s) = s$, from (21), we obtain

$$1 \leq \operatorname{dist}(0, \partial f(x^k)), \qquad k \geq k_0. \tag{46}$$

On the other hand, from the definition of the sequence $\{x^k\}$, for each $k \in \mathbb{N}$ there exist $w^k \in \partial f(x^k)$ and $v^k \in \partial q(x^{k-1}, \cdot)(x^k)$ such that

$$\|w^k\| \leq b\Gamma'[q(x^{k-1}, x^k)]\|v^k\| \leq bL\Gamma'[q(x^{k-1}, x^k)], \tag{47}$$

where the last inequality follows from (19). Taking into account that D satisfies (2), it is easy to see that $\Gamma'[q(x^{k-1}, x^k)]$ converges to zero as k goes to $+\infty$. Hence, combining (46) with (47) and letting k goes to $+\infty$, we obtain a contradiction. This proves item $i)$ and conclude the proof. □

Remark 3.3. *The constants Q, c_1 and c_2, present in Theorem 3.2, may be made explicit as follows:*

$$c_1 = \frac{\beta_2}{\beta_1}Q^{k_0}\Delta_{k_0}, \quad \textit{where} \quad Q = \frac{r + \bar{\rho}_\Gamma(r)M[(1-\theta)bL\gamma(\theta)]^{h(\theta)}}{1 + \bar{\rho}_\Gamma(r)M[(1-\theta)bL\gamma(\theta)]^{h(\theta)}},$$

$$c_2 = \frac{\beta_2}{\beta_1}\frac{(1-h(\theta))^{\frac{h(\theta)}{h(\theta)-1}}(1-r)^{\frac{1}{h(\theta)-1}}}{[Rh(\theta)]^{\frac{h(\theta)}{h(\theta)-1}}\left[r + \bar{\rho}_\Gamma(r)M[(1-\theta)bL\gamma(\theta)]^{h(\theta)}\right]^{\frac{1}{h(\theta)-1}}}.$$

Note that if $\Gamma[q] = q^\alpha$, $\alpha > 1$, from item a) of Remark 3.2 it is easy to see that the last theorem holds with $h(\theta) = \frac{(1-\theta)(\alpha-1)}{\theta}$. In particular, [7, Theorem 2] is obtained for $\alpha = 2$.

4 Application. The strength of habituation and routinization processes and the "loss aversion effect"

4.1 The Variational approach

Successions of worthwhile temporary stays and changes

The recent VR variationality approach (see [1, 2]) helps to unify a lot of stay/stability and change behavioral dynamics in Behavioral Sciences (Psychology, Economics, Management Sciences, Decision theory, Philosophy, Game theory, Political Sciences, Artifical Intelligence...), as a succession of worthwhile temporary stays and changes $x^{k+1} \in W_{e_k,\xi_{k+1}}(x^k)$, $k \in \mathbb{N}$, ending in variational traps $x^* \in X$. Stays $x = x^k \curvearrowright y = x^{k+1} = x^k$ refer to habits, routines, equilibria, rules, norms, and traps. Changes $x = x^k \curvearrowright y = x^{k+1} \neq x^k$ represent creations, destructions, innovations, learning processes,.... A point $x \in X$ can be an action (doing), or a state (being, having). In the current period $k+1$, an agent considers that a change $x^k \curvearrowright x^{k+1} \in W_{e_k,\xi_{k+1}}(x^k)$ is worthwhile if his motivation to change $M_{e_k}(x^k, x^{k+1})$ is higher than his resistance to change $R_{e_k}(x^k, x^{k+1})$ up to a satisficing worthwhile to change ratio $\xi_{k+1} > 0$. At the end of the last period k, the experience of the agent is $e^k \in E$.

Within the current period $k+1$, a worthwhile change is both feasible and desirable enough, i.e, acceptable, improving with no too high costs to be able to improve, satisficing with not too much sacrificing. Each period, the agent chooses the ratio $\xi_{k+1} > 0$, which represents how worthwhile a change must be to accept to move rather than to stay. The famous (see [43]) satisficing principle is a specific case (see [1, 2]). Marginally not worthwhile changes play also a major role. Within the same period, the agent must also have to know when he must stop changing. This is the case when one step more is not worthwhile. More formally, this change is not "marginaly worthwhile", when the ex ante marginal motivation to change is sufficiently lower than the ex ante marginal resistance to change. In this case the agent does not regret ex ante to do not go one step further. The motivation to change again next period comes from residual unsatisfied needs or variable preferences.

Then, $x^{k+1} \in W_{e_k,\xi_{k+1}}(x^k) \Longleftrightarrow M_{e_k}(x^k, x^{k+1}) \geq \xi_{k+1} R_{e_k}(x^k, x^{k+1})$. Starting from an initial position x^0, a variational trap $(x^*, e_*, \xi_*) \in X \times E \times \mathbb{R}_+$ is the end of a worthwhile to change process $x^{k+1} \in W_{e_k,\xi_{k+1}}(x^k)$, $k \in \mathbb{N}$, that is to say a position where, at the end, it is worthwhile to stay (i.e., not worthwhile to move again): $W_{e_*,\xi_*}(x^*) = \{x^*\}$. Then, the pilar of the VR approach are worthwhile changes and variational traps. The main question is to know when successions of worthwhile temporary stays and changes end in variational traps.

Given the current experience $e = e^k$, motivation to change $M_e(x, y) = U_e\left[A_e(x, y)\right]$ from $x = x^k$ to $y = x^{k+1}$ is defined as the utility $U_e\left[A_e\right]$ of advantages to change $A_e = A_e(x, y)$. Resistance to change $R_e(x, y) = D_e\left[I_e(x, y)\right]$ refers to the disutility $D_e\left[I_e\right]$ of inconvenients to change $I_e = I_e(x, y)$. Suppose that U^{-1} exists and that $U_e\left[0\right] = 0 = D_e\left[0\right]$. Let $\Gamma_e\left[I_e\right] = U_e^{-1}\left[D_e\left[I_e\right]\right]$ be the relative resistance to change function. It defines how strong is the utility of advantages to change with respect to the disutility of inconvenients to change.

Inertia or the degree of resistance to change

The variational rationality approach defines the degree of resistance to change $\delta_e\left[Z\right] = D_e\left[Z\right]/U_e\left[Z\right]$, for $Z > 0$. It says that the agent resists to small change when $\delta_e\left[Z\right] > 1 \Longleftrightarrow D_e\left[Z\right] > U_e\left[Z\right]$, for $Z > 0$ small enough. In this case inconvenients to change $I_e = Z > 0$ seem larger (enough) than the same size advantages to change $Z = A_e > 0$, this common size Z being small enough.

Advantages and inconvenients to change

They admit a lot of variants. Let us choose, among all of them, one which retrieves the formulation of proximal algorithms and gives a direct generalization. We will suppose that, when they exist, advantages and inconvenients to change are experience free and separable, i.e.,

$$A_e(x,y) = \chi\left[g(y) - g(x)\right] = \chi\left[f(x) - f(y)\right] \geq 0 \quad \text{and} \quad I_e(x,y) = C(x,y) - C(x,x) \geq 0, \quad \chi > 0,$$

where:

- $g : x \in X \longmapsto g(x) \in R$ is a "to be increased" payoff of the agent, say his satisfaction for a consumer, or his profit for an entrepreneur;
- $g : x \in X \longmapsto g(x) \in R$ is a "to be decreased" payoff of the agent, say his unsatisfied needs for a consumer, or his residual profit "to be exhausted" for an entrepreneur;
- $C(x,y) \geq 0$ is his costs to be able to move from x to y;
- $C(x,x) \geq 0$ is his costs to be able to stay at x;
- $Q_\lambda(x,y) = g(y) - \lambda\, \Gamma\left[I(x,y)\right]$ is the current "to be increased" proximal payoff of the agent;
- $P_\lambda(x,y) = f(y) + \lambda\, \Gamma\left[I(x,y)\right]$ is the current "to be decreased" proximal payoff of the agent.

In this case, a current change from x to y is worthwhile iff $M(x,y) \geq \xi R(x,y)$, i.e., $U\left[A(x,y)\right] \geq \xi D\left[I(x,y)\right]$ i.e.,

$$g(y) - g(x) \geq \lambda U^{-1}\left[\xi D\left[I(x,y)\right]\right] = \lambda \Gamma_\xi\left[I(x,y)\right] \quad \text{or} \quad f(x) - f(y) \geq \lambda \Gamma_\xi\left[I(x,y)\right], \quad \lambda = 1/\chi.$$

For simplification we will take $\xi = 1$ and $\Gamma_1\left[I(x,y)\right] = \Gamma\left[I(x,y)\right]$. Then a change is worthwhile if the "to be increased" proximal payoff of the agent increases, or if the "to be decreased" proximal payoff of the agent decreases. $M(x,y) \geq \xi R(x,y) \Longleftrightarrow Q_\lambda(x,y) \geq Q_\lambda(x,x) \Longleftrightarrow P_\lambda(x,y) \leq P_\lambda(x,x)$. In this situation a variational trap $(x^*, \lambda_*) \in X \times \mathbb{R}_+$ satisfies the stability condition

$$g(y) - g(x^*) < \lambda_* \Gamma\left[I(x^*, y)\right] \quad \text{or} \quad f(x^*) - f(y) < \lambda_* \Gamma\left[I(x^*, y)\right], \quad y \neq x^*.$$

4.2 An example

A simple model of knowledge management

Because the second paper, Bento and Soubeyran [41], is a continuation of the present paper, for easier understanding, this example is the same in both papers. It modelizes a very simple case of knowledge management within an organization, to determine the optimal size and shape of an innovative firm driven by a leader. In Management Sciences, the literature on this topic is enormous and represents one of its main area of research. Consider an entrepreneur (leader) who, each period, can hire and fire different kinds and numbers of skilled and specialized workers $\{1, 2, \ldots, j, \ldots l\} = J$ to produce a chosen quantity of a final good of a chosen quality. These workers are knowledge workers (see Long et al. [42]). The endogenous quality $q(x)$ of this final good changes with the chosen profile of skilled workers $x = (x^1, x^2, \ldots, x^j, \ldots, x^l) \geq 0$, where $x^j \geq 0$ is a number of workers of type j. To save space and for simplification, each period, each employed skilled worker utilizes one unit of a specific non durable mean, to produce, using his specific knows how, one unit of a specific component of type j. Then, the entrepreneur combines these different components to produce $\mathfrak{q}(x)$ units of a final good of endogenous quality $s(x)$. This production function is original because it combines both variable quantity and quality. The revenue of the entrepreneur is $\varphi\left[\mathfrak{q}(x), s(x)\right]$. His operational costs $\rho(x)$ are the sum of his costs to buy the non durable means used by each worker, and the wages paid to each employed worker. Then, in a given period, the profit of the entrepreneur who employs the profile $x \in X = \mathbb{R}^l$ of skilled workers is $g(x) = \varphi\left[\mathfrak{q}(x), s(x)\right] - \rho(x) \in \mathbb{R}$. For a famous example of an endogenous production function of quality, see Kremer [40].

To fit with the formulation of inexact proximal algorithms, where mathematicians consider to be decreased cost functions, let us consider the residual profit that the entrepreneur expects to exhaust in the future, $f(x) = \overline{g} - g(x) \geq 0$, where $\overline{g} = \sup\{g(y), y \in X\} < +\infty$ is the highest finite profit that the entrepreneur can hope to get.

Costs to be able to change (to stay)

To be able to hire one skilled worker of type j, ready to work, costs $h_+^j > 0$. These costs include search and training costs. To be able to fire one worker of type j, costs $h_-^j > 0$. These costs represent separation and compensation costs. To keep a worker, ready to work, one period more, costs $h_=^j \geq 0$. These conservation costs include knowledge regeneration and motivation costs. Then, in the current period, costs to be able to conserve the same profile of workers as before are $C(x,x) = \Sigma_{j=1}^n h_=^j x^j$ while costs to utilize the profile of skilled workers y are

$$C(x,y) = \sum_{j \in J_+(x,y)} \left[h_=^j x^j + h_+^j (y^j - x^j)\right] + \sum_{j \in J_-(x,y)} \left[h_=^j y^j + h_+^j (x^j - y^j)\right],$$

where $J_+(x,y) = \left\{j \in J : y^j \geq x^j\right\}$ and $J_-(x,y) = \left\{j \in J : y^j < x^j\right\}$. For simplification, suppose that conservation costs are zero, i.e., $h_=^j = 0$. Then, $C(x,x) = 0$ and $I(x,y) = \Sigma_{j \in J_+(x,y)} h_+^j (y^j - x^j) + \Sigma_{j \in J_-(x,y)} h_-^j (x^j - y^j)$. Then, $I(x,y)$ is a quasi distance $q(x,y) = I(x,y) \geq 0$ such that

$$q(x,y) = 0, \ iff \ y = x \quad \text{and} \quad q(x,z) \leq q(x,y) + q(y,z), \quad x,y,z \in X.$$

The more general case where $h_=^j > 0$ works as well.

Motivation and resistance to change functions

Take $U[A] = A^{\mu}$ and $D[I] = I^{\nu}$ with $\mu, \nu > 0$. Then, the relative resistance to change function is

$$\Gamma[I(x,y)] = I(x,y)^{\alpha}, \qquad \alpha = \nu/\mu > 0.$$

An habituation/routinization process

It is such that, step by step, gradually, the agent carries out a more and more similar action. This is equivalent to say than the quasi distance $C(x^k, x^{k+1})$ converges to zero as k goes to infinity.

When a worthwhile to change process converges to a variational trap, this variational formulation offers a model of trap as the end point of a path of worthwhile temporary stays and changes.

4.3 Inexact proximal algorithms as succession of worthwhile temporary stays and changes

Succession of worthwhile temporary stays and changes

In this example they are $x^{k+1} \in W_{\lambda_{k+1}}(x^k)$, $k \in \mathbb{N}$, i.e.,

$$g(x^{k+1}) - g(x^k) \geq \lambda_{k+1}\Gamma\left[q(x^k, x^{k+1})\right] \quad \text{or} \quad f(x^k) - f(x^{k+1}) \geq \lambda_{k+1}\Gamma\left[q(x^k, x^{k+1})\right], \quad k \in \mathbb{N}.$$

Sufficient descent conditions

They are

$$g(x^k) - \lambda_k\Gamma\left[q(x^k, x^{k+1})\right] \geq g(x^k) \quad \text{or} \quad f(x^{k+1}) + \lambda_k\Gamma\left[q(x^k, x^{k+1})\right] \leq f(x^k), \qquad k \in \mathbb{N}.$$

Then, they represent worthwhile changes $x^{k+1} \in W_{\lambda_k}(x^k)$, $k \in \mathbb{N}$, where, in Mathematics, λ_k replaces λ_{k+1}.

Inexact proximal algorithms

They include, each step $k+1$, a sufficient descent condition and a stopping rule. Then, they modelize particular instances of worthwhile temporary stays (where it is marginally not worthwhile to change) and changes processes.

4.4 Speed of convergence, loss averstion and the degree of resistance to change

4.4.1 Inertia (to resist to change) and loss aversion

Degree of resistance to change

As seen before the variational rationality approach defines relative resistance to change as $\Gamma[I] = U^{-1}[D(I)]$ and inertia, or degree of resistance to change, as $\delta_e(Z) = D_e[Z]/U_e[Z]$, for $Z > 0$. It says that the agent resists to small change when $\delta_e(Z) > 1 \Longleftrightarrow D_e[Z] \geq U_e[Z]$ for small enough changes $Z > 0$. It also tells us that the agent favors small change when $\delta_e(Z) < 1 \Longleftrightarrow D_e[Z] < U_e[Z]$, for small enough changes $Z > 0$.

Loss aversion

In Behavioral Economics, Kahneman and Tversky and their followers have defined several famous "loss aversion" indexes, both in case of riskless choices (see Tversky and Kahneman [10]) and for the case of risky choices (see Kahneman and Tversky [11]). In their case, there are no experience effects, and the loss aversion index $l(Z) = D[Z]/U[Z]$ is a particular case of the degree of resistance to change $\delta_e(Z)$. Other indexes of "loss aversion" have been given (for example $\widetilde{l}(Z) = D'[Z]/U'[Z]$, see (Abdellaoui et al. [44] and Köbberling and Wakker [45]). To simplify we will consider the leading example where $U[A] = A^\mu, \mu > 0$, and $D[I] = I^\nu, \nu > 0$. Then, $\Gamma[Z] = Z^\alpha$ with $\alpha = \nu/\mu > 0$, $\delta(Z) = l(Z) = Z^{\nu-\mu} = Z^{\mu(\alpha-1)}$ and $\widetilde{l}[Z] = (\nu/\mu)Z^{\nu-\mu}$.

Loss aversion and the strength of resistance to change

Then, when $\alpha > 1$, and for a given $0 < Z < 1$, the formula $\delta(Z) = e^{\mu(-\log Z)(\alpha-1)}$ shows that when resistance to change in the small increases (that is to say α decreases), $\delta(Z)$ decreases.

Weak resistance for small changes

Our present paper considers the leading example of $\Gamma[q(x,y)] = q(x,y)^{\alpha}$, where $\alpha > 1$. If $\alpha = \nu/\mu$, then $\nu > \mu$. This represents the case of weak resistance to change "in the small" $(0 < q < 1)$ but strong resistance to change "in the large" $(q \geq 1)$, because $q^{\alpha} < q$ for $0 < q < 1$ and $q^{\alpha} \geq q$ for $q \geq 1$.

Strong resistance for small changes

In the opposite case $0 < \alpha \leq 1$, where there is strong resistance to change "in the small" (but weak resistance to change "in the large"), Soubeyran [1, 2] examined the case of "loss aversion for small changes" which represents also "strong resistance to small changes", where $\Gamma[q] \geq \sigma q$, $\sigma > 0$, for $q > 0$ small enough.

4.4.2 Convergence. How the strength of resistance to change impacts hibt/routine formation

- **Problem**. In the context of habit/routine formation, the main point is to know how more or less bounded needs and the strength of resistance to change (weak, strong) favors or not the speed of convergence. This point is very important because an habituation/routinization process is gradual. It represents a progressive increase in automaticity. The speed of convergence of the inexact proximal algorithm modelizes in a nice way a slow learning process of automatization (how the same way of doing an action emerges gradually from repetitions in the same recurrent context). Actions become more and more similar and converge to a limit action which is a variational trap, a permanent habit, where the agent prefers to stay than to move. In this behavioral context, Theorem 3.2 of our paper is a powerful result. It shows the importance of the strength of resistance to change for the formation of habits and routines.

- **Findings**. Let us consider the case of weak resistance to change in the small where $\Gamma\left[q(x,y)\right] = q(x,y)^{\alpha}$, $\alpha > 0$. Then $h(\theta) = (\alpha - 1)(1-\theta)/\theta > 0$ for $0 < \theta < 1$. Theorem 3.2 of our paper exhibits two polar cases:

  ii) very weak resistance to change in the small (when α is > 1 and very high):

  $$h(\theta) \geq 1 \Longleftrightarrow \alpha \geq 1/(1-\theta) > 1.$$

  In this case, $q(\widetilde{x}, x^k) \leq c_1 Q^k = c_1 e^{(\log Q)k}$ with $\log Q < 0$ from $0 < Q < 1$;

  iii) weak, but stronger resistance to change in the small (when α is > 1, but lower than before):

  $$0 < h(\theta) < 1 \Longleftrightarrow 1 < \alpha < 1/(1-\theta).$$

  In this case where α is lower,

  $$q(\widetilde{x}, x^k) \leq c_2 k^{\omega(\theta)} = c_2 e^{\omega(\theta)(\log k)},$$

  with $\omega(\theta) = h(\theta)/\left[h(\theta) - 1\right] < 0$ from $0 < h(\theta) < 1$.

  Then, when k goes to infinity, the speed of convergence is higher in case of very weak resistance to change (case *ii*)) than in case of stronger resistance to change (case *iii*)), because $k > \log k$ for k high enough. This shows that when resistance to change is small reamins weak, i.e., $\alpha > 1$, when α decreases, resistance to change in the small increases. Then, moving from ii) to iii) show that,

- **Finding** as long as resistance to change is weak ($\alpha > 1$), if resistance to change in the small increases, then, the speed of convergence of the habituation/routinization process decreases. This is a striking result which is also intuitive. It means that, each period, the agent will make smaller steps when resistance to change increases.

Finally, when $\theta = 1$, habit/routine formation is in finite time, a nice result.

5 Conclusion

The consideration of relative resistance to change functions $\Gamma[\cdot]$ helps to classify exact and inexact proximal algorithms in two separate groups. The first case considers strong resistance to change, where $\Gamma[q] = q$ is linear for all $q \geq 0$. The second case examines weak resistance to change, where, $\Gamma[\cdot]$ is "convex enough", for example, $\Gamma[q] = q^{\alpha}$, $\alpha > 1$ for all $q \geq 0$. In both cases $q = q(x, y)$ is a distance or a quasi distance. Applications in Behavioral Sciences needs convergence to a variational trap, which is worthwhile to approach, but not worthwhile to leave. The question becomes: when a critical point is a variational trap. This point has been examined in Bento and Soubeyran [41]. In the case of weak resistance to change (the present paper) a striking result for applications in Behavioral Sciences has been to show how the strength of resistance to change can impact the speed of convergence of an habituation/routinization process. In the opposite case of strong resistance to change, this impact is much easy to show (see [1, 2]). Then, the variational approach which considers relative resistance to change as a core concept, which balances the strength of motivation and resistance to change, provides us an extra motivation to develop further the study of proximal algorithms in a nonconvex and possibly nonsmooth setting.

Acknowledgment

First author was partially supported by CNPq Grant 471815/2012-8, Project CAPES-MES-CUBA 226/2012, PROCAD-nf-UFG/UnB/IMPA and FAPEG/CNPq.

References

1. Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. Pre-print. GREQAM, Aix Marseillle University, (2009).

2. Soubeyran, A.: Variational rationality and the "unsatisfied man": routines and the course pursuit between aspirations, capabilities and beliefs. Preprint. GREQAM, Aix Marseillle University (2010).

3. Attouch.H, Soubeyran, A.: Local Search Proximal Algorithms as Decision Dynamics with Costs to Move. Set-Valued Var. Anal. **19**(1), 157–177 (2011)

4. Flores-Bazán.F., Luc. D, Soubeyran, A.: Maximal Elements Under Reference-Dependent Preferences with Applications to Behavioral Traps and Games. J Optim Theory Appl. **155**, 883–901 (2012)

5. Attouch, H., Redont, P., Soubeyran, A.: A new class of alternating proximal minimization algorithms with costs-to-move, SIAM J. Optimiz. **18**, 1061-1081 (2007)

6. Cruz Neto, J. X., Oliveira, P. R., Soares Jr, P. A., Soubeyran, A.: Learning how to play Nash, potential games and alternating minimization method for structured nonconvex problems on Riemannian manifolds. To appear in J. Convex Anal. (2013)

7. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Programming, Ser. B. **116**(1-2), 5-16 (2009)

8. Attouch, H., Bolte, J., Svaiter, B. F.: Convergence of descent methods for semi-algebric and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Programming, Ser. A. **137**, 91-129 (2012)

9. Moreno, F. G., Oliveira, P. R., Soubeyran, A.: A Proximal Algorithm with Quasi Distance. Application to Habit's Formation. Optimization **61**(12), 1383-1403 (2011)

10. Tversky, A., Kahneman, D.: Loss aversion in riskless choice: a reference dependent model. Q. J. Econ. **106**(4), 1039-1061 (1991)

11. Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica **47**(2), 263-291 (1979)

12. Moreau, J.: Proximité et dualité dans un espace hilbertien. (French) Bull. Soc. Math. **93**, 273-299 (1965)

13. Martinet, B.: Régularisation, d'inéquations variationelles par approximations successives. (French) Rev. Française Informat. Recherche Opérationnelle **4**, Ser. R-3, 154-158 (1970)

14. Rockafellar, R. T.: Monotone operators and the proximal point algorithm. SIAM J. Control. Optim. **14**, 877-898 (1976)

15. Rockafellar, R. T., Wets, R.: Variational Analysis, **317** of Grundlehren der Mathematischen Wissenschafte. Springer, 1998 (3rd printing 2009).

16. Fukushima, M., Mine, H.: A generalized proximal point algorithm for certain nonconvex minimization problems. Int. J. Systems Sci. **12**, 989-1000 (1981)

17. Kaplan, A., Tichatschke, R.: Proximal point methods and nonconvex optimization. J. Global Optim. **13**(4), 389-406 (1998)

18. Spingarn, J. E.: Submonotone mappings and the proximal point algorithm. Numer. Funct. Anal. Optim. **4**(2), 123-150 (1981)

19. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**, 170-191 (2002)

20. Iusem, A. N., Penannen, T., Svaiter, B. F.: Inexact variants of the proximal point algorithm without monotonicity. SIAM J. Optim. **13**(4), 1080-1097 (2003)

21. Combettes, P. L., Pennanen, T.: Proximal methods for cohypomonotone operators. SIAM J. Control Optim. **43**, 731–742 (2004)

22. Gárciga-Otero, R., Iusem, A. N.: Proximal methods in reflexive Banach spaces without monotonicity. J. Math. Anal. Appl. **330**(1), 433-450 (2007)

23. Mordukhovich, B.: Variational Analysis and Generalized Differentiation I: Basic Theory (Grundlehren der mathematischen Wissenschaften), (2010)

24. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Lojasiewicz inequalities: Subgradient flows, talweg, convexity. Trans. Amer. Math. Soc. **362**, 3319-3363 (2010)

25. Solodov, M. V., Svaiter, B. F.: A hybrid approximate extragradient - proximal algorithm using the enlargement of a maximal monotone operator. Set-Valued Var. Anal. **7**, 323-345 (1999)

26. Solodov, M. V., Svaiter, B. F.: Forcing strong convergence of proximal point iterations in a Hilbert space. Math. Program. Ser. A **87**, 189-202 (2000)

27. Lojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. Les Équations aux Dérivées Partielles, Éditions du centre National de la Recherche Scientifique, 87-89 (1963)

28. van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**, 497-540 (1996)

29. Kurdyka, K.: On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier, **48**, 769—783 (1998)

30. Bolte, J., Daniilidis, J. A., Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205-1223 (2006)

31. Bolte, J., Daniilidis, J. A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556-572 (2007)

32. Attouch, H., Redont, P., Bolte, J., Soubeyran, A.: Proximal Alternating Minimization and Projection Methods for Nonconvex Problems. An Approach Based on the Kurdyka-Lojasiewicz Inequality. Math. Oper. Res. **35**(2), 438-457 (2010)

33. Kurdyka, K., Mostowski, T., Parusinski, A.: Proof of the gradient conjecture of R. Thom. Ann. Math, **152**, 763-792 (2000)

34. Lageman, C.: Convergence of gradient-like dynamical systems and optimization algorithms, Ph.D. Thesis, Universität Würzburg (2007)

35. Bento, C. G., Cruz Neto, J. X. and Oliveira, P. R.: Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds. arXiv:1103.4828 (2011)

36. Zaslavski, A. J.: Inexact Proximal Point Methods in Metric Spaces. Set-Valued Anal. **19**, 589-608 (2011)

37. Yu, P. L., Chen,Y. C.: Dynamic MCDM, habitual domains and competence set analysis for effective decision making in changeable spaces. Chaper 1. In Trends in Multiple Criteria Decision Analysis. Matthias Ehrgott Jose Rui Figueira Salvatore Greco. Springer (2010)

38. Larbani, M., Yu, P. L.: Decision Making and Optimization in Changeable Spaces, a New Paradigm. J. Optim. Theory Appl., **155**(3), 727-761 (2012)

39. Cohen, M., Grossberg, S.: Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE T. Syst. Man. CYB, **13**, 815-826 (1983)

40. Kremer, M.: The O-ring theory of economic development. The Quarterly Journal of Economics, **108**(3), 551-575 (1993)

41. Bento, G., Soubeyran, A.: Generalized Inexact Proximal Algorithms: Habit's/ Routine's Formation with Resistance to Change, following Worthwhile Changes. Submitted (2014).

42. Van Long, N., Soubeyran, A., Soubeyran, R.: Knowledge Accumulation within an Organization. To appear in Int. Econ. Rev. (2014)

43. Simon. H.: A behavioral model of rational choice. Q. J. Econ. **69**, 99-118 (1955)

44. Abdellaoui, M., Bleichrodt, H., Paraschiv, C.: Loss aversion under prospect theory: A parameter-free measurement. Manage. Sci. **53**(10), 1659-1674 (2007)

45. Köbberling, V., Wakker, P.: An index of loss aversion. J. Econ. Theory. **122**, 119–131 (2005)