

Generalized Inexact Proximal Algorithms: Routine's Formation with Resistance to Change, Following Worthwhile Changes

G. C. Bento · A. Soubeyran

Received: 13 June 2013 / Accepted: 30 January 2015 / Published online: 12 February 2015
© Springer Science+Business Media New York 2015

Abstract This paper shows how, in a quasi-metric space, an inexact proximal algorithm with a generalized perturbation term appears to be a nice tool for Behavioral Sciences (Psychology, Economics, Management, Game theory,...). More precisely, the new perturbation term represents an index of resistance to change, defined as a “curved enough” function of the quasi-distance between two successive iterates. Using this behavioral point of view, the present paper shows how such a generalized inexact proximal algorithm can modelize the formation of habits and routines in a striking way. This idea comes from a recent “variational rationality approach” of human behavior which links a lot of different theories of stability (habits, routines, equilibrium, traps,...) and changes (creations, innovations, learning and destructions,...) in Behavioral Sciences and a lot of concepts and algorithms in variational analysis.

Keywords Nonconvex optimization · Kurdyka–Lojasiewicz inequality · Inexact proximal algorithms · Habits · Routines · Worthwhile changes

Mathematics Subject Classification 49J52 · 49M37 · 65K10 · 90C30 · 91E10

Communicated by Po-Lung Yu.

G. C. Bento (✉)
Universidade Federal de Goiás, Goiânia, Brazil
e-mail: glaydston@ufg.br

A. Soubeyran
CNRS & EHESS, Aix-Marseille University (Aix-Marseille School of Economics),
Marseille, France
e-mail: antoine.soubeyran@gmail.com

1 Introduction

The main message of this paper is that, using the behavioral context of a recent “variational rationality” approach of worthwhile stay and change dynamics proposed by Soubeyran [1,2], a generalized proximal algorithm can modelize fairly well an habituation process as described in Psychology for an agent, or a routinization process, in Management Sciences, for an organization. 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, but also nice tools to modelize the dynamics of human behaviors.

Theories of stability and change consider successions of stays and changes. Stays refer to habits, routines, equilibria, traps, rules and conventions, etc. Changes represent creations, destructions, learning processes, innovations, attitudes as well as the formation and revision of beliefs, self-regulation problems, including goal setting, goal striving and goal revision, the formation and break of habits and routines. In the interdisciplinary context which characterizes all these theories in Behavioral Sciences, the “variational rationality approach” (see [1,2]) shows how to modelize the course of human activities as a succession of worthwhile temporary stays and changes which balance, at each step, the motivation to change (the utility of advantages to change) and the resistance to change (the disutility of inconveniences to change). This very simple idea has allowed to see proximal algorithms as an important tool to modelize the human course of actions, where the perturbation term of a proximal algorithm can be seen as a crude formulation of the complex concept of resistance to change, while the utility generated by a change in the objective function can represent a crude formulation of the motivation to change concept. The variational rationality approach considers three original concepts: (i) “worthwhile changes,” when, at each step, the motivation to change is higher enough with respect to resistance to change, (ii) “non-marginal worthwhile changes” and, (iii) “variational traps,” “easy enough to reach,” that the agent can reach by using a succession of worthwhile changes, and “difficult enough to leave,” such that, being there, it is not worthwhile to move from there.

These three concepts represent the pillar of the variational rationality approach; see [1,2], which has provided an extra motivation to further develop the study of proximal algorithms in a nonconvex and possibly nonsmooth setting. Among other recent applications of this simple idea see, for instance, Attouch and Soubeyran [3] for local search proximal algorithms, Flores-Bazán et al. [4] for worthwhile to change games, Attouch et al. [5] for alternating inertial games with costs to move and Cruz Neto et al. [6] for the “how to play Nash” problems. In all these papers, the perturbation term of the usual proximal point algorithm is a linear or quadratic function of the distance or quasi- distance between two successive iterates. It modelizes the case of “strong enough resistance” to change. Our paper examines the opposite case of “weak enough” resistance to change where the perturbation 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 another one is not the same as the cost to be able to change in the other way. In a first paper, Bento and Soubeyran [7] showed when, in a

quasi-metric space, a generalized inexact proximal algorithm (equipped with such a generalized perturbation term) defined, at each step, by a sufficient descent condition and a stopping rule, converges to a critical point, and that the speed shown that the speed of convergence and the convergence in finite time depend on the curvature of the perturbation term and on the Kurdyka–Lojasiewicz property associated to the objective function. A striking and new application is given. It concerns the impact of the famous “loss aversion effect” (Nobel Prize Kahneman and Tversky [8], Tversky and Kahneman [9]) on the speed of convergence of the generalized inexact proximal algorithm.

In the present paper, inspired by the VR “variational rationality” approach, we consider a new inexact proximal algorithm whose sufficient descent condition is, at each step, a little more demanding, using the same stopping rule. Applying the convergence result of the first paper [7], this simple modification is a way to force convergence even more. It gives an intuitive sufficient condition for the critical point to be a variational trap (weak or strong). In this case, changes are required to be “worthwhile enough,” the stopping rule is the same, and the end of the convergence worthwhile stay and change process is both a critical point and a variational trap. Doing so, this paper extends the convergence result to a critical point of Attouch and Bolte [10], Attouch et al. [11] and Moreno et al. [12], using a fairly general “convex enough” perturbation term. It is important to note that, as an application, it is possible to consider the formation of habits and routines as an inexact proximal algorithm in the context of weak resistance to change. However, because of its strongly interdisciplinary aspect (Mathematics, Psychology, Economics, Management), to be carefully justified, this application needs several steps. Due to space constraints here, these considerations are given in Bento and Soubeyran [13]. Then, at the behavioral level, the main message of this paper is to advocate that our generalized proximal algorithm is well suited to modelize the formation of habitual/ routinized human behaviors. The list of the main (VR) concepts is presented through an example in Sect. 2. To get a better perspective, in [13], the VR approach of stability and change dynamics is compared, from our point of view, with that of the HD habitual domain theory (see Yu [14]) and its application to DMCS (decision making with changeable spaces; see Larbani and Yu [15]).

At a higher dimensional level, VR and HD can be complementary to each other. Both consider stability and change dynamics, but in different formulations. Deterministic worthwhile temporary stays and changes transitions are the bases for the VR approach, and dynamic charge structures and attention allocation are that for HD theory. Both put focus on optimization and satisfying processes. VR is variable and possibly intransitive preference (utility) based, while HD is based on charge structures resulting from goal setting and state evaluation. VR’s main topic is the self-regulation problem (goal setting, goal striving, goal revision and goal disengagement) at the individual level or for interactive agents. In addition to this, the HD also contains some main topics of “how agents expand and enrich their habitual domain” as to solve challenging decision problems.

Our paper is organized as follows. Section 2 gives an example which helps to list the main variational tools necessary to define the central concepts of “worthwhile change” and “variational trap” for behavioral applications. Section 3 shows how inexact proximal algorithms can represent adaptive satisfying processes. Section 4 examines a

generalized inexact proximal algorithm which converges to a critical point which is also a variational trap (weak or strong), when the objective function satisfies a Kurdyka–Lojasiewicz inequality. In [13], the authors compare, on their point of view, the VR and HD approaches relative to habituation/routinization processes.

2 Variational Rationality: How Successions of Worthwhile Stays and Changes End in Variational Traps

2.1 Worthwhile Stay and Change Dynamics

A recent variational rationality approach, see [1,2], gives a common background to a lot of theories of stability/stay and change in Behavioral Sciences (Psychology, Economics, Management Sciences, Decision theory, Philosophy, Game theory, Political Sciences, Artificial Intelligence. . .), using as a central building bloc the three concepts of “worthwhile change,” “marginal worthwhile change” and “variational trap.” All these behavioral dynamics can be seen 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 $x^* \in X$, where X is the universal space of actions (doing), having or being, depending on the applications. X includes all past elements and all the new elements that can be discovered as time evolves.

The main idea is quite evident. If a behavioral theory wants to explain “why, where, how and when” agents perform actions and change, this theory must define, at each period, along a path of changes $\{x^0, x^1, \dots, x^k, x^{k+1}, \dots\}$ why the agent has, first, an incentive to take some steps away from his current position and, then, an incentive to stop changing one step more within this period. In the current period $k + 1$, a change is such that $x^{k+1} \neq x^k$, while a stay is $x^{k+1} = x^k$. Let $e_k \in E$ be the experience of the agent at the end of the last period k . A change $x^k \rightsquigarrow x^{k+1} \in W_{e_k, \xi_{k+1}}(x^k)$ is worthwhile, when its ex ante motivation to change $M_{e_k}(x^k, x^{k+1})$ is sufficiently higher (more than $\xi_{k+1} > 0$) than his ex ante resistance to change, $R_{e_k}(x^k, x^{k+1})$. Then, $x^{k+1} \in W_{e_k, \xi_{k+1}}(x^k) \iff M_{e_k}(x^k, x^{k+1}) \geq \xi_{k+1} R_{e_k}(x^k, x^{k+1})$. Motivation and resistance to change are two complex variational concepts which admit a lot of variants (see [1,2]). Motivation to change $M_{e_k}(x^k, x^{k+1}) = U_{e_k}[A_{e_k}(x^k, x^{k+1})]$ is the utility $U_{e_k}[\cdot]$ of the advantages to change, $A_{e_k}(x^k, x^{k+1})$, while resistance to change $R_{e_k}(x^k, x^{k+1}) = D_{e_k}[I_{e_k}(x^k, x^{k+1})]$ is the disutility $D_{e_k}[\cdot]$ of the inconveniences to change $I_{e_k}(x^k, x^{k+1})$.

Worthwhile changes are generalized satisfying changes Within a period, a worthwhile change $x^k \rightsquigarrow x^{k+1} \in W_{e_k, \xi_{k+1}}(x^k)$ is desirable and feasible enough, i.e, acceptable, improving with not too high costs to be able to improve. Then, a worthwhile change is a generalized satisfying change where, at 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 Simon [16] satisfying principle is a specific case (see [1,2]). Second, 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 “marginally 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 not to go one step further. The motivation to change again next period comes from residual unsatisfied needs or variable preferences.

A variational trap x^* is such that, starting from an initial point $x^0 \in X$, it exists a path of worthwhile changes $x^{k+1} \in W_{e_k, \xi_{k+1}}(x^k)$ which ends in x^* , i.e., such that, being there, it is not worthwhile to move again, i.e., $W_{e_*, \xi_*}(x^*) = \{x^*\}$.

2.2 Variational Concepts: An Example

To save space and to fix ideas, let us define these variational rationality concepts through a simple example. This being done, we can easily show how an inexact proximal algorithm represents a nice benchmark process of worthwhile temporary stays and changes in term of, at each period, a sufficient descent condition and a stopping rule. For more comments, and a more complete formulation of each of these variational concepts, with references to a lot of different disciplines in Behavioral Sciences which will help to justify their unifying power; see [1,2].

A simple model of knowledge management: This example modelizes a very simple case of knowledge management within an organization, to determine a satisfying or, as an extreme case, 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, at each period, can hire and fire different kinds and numbers of skilled and specialized workers $\{1, 2, \dots, j, \dots, l\} = J$ (say knowledge workers; see Long et al. [17]) to produce a chosen quantity of a final good of a chosen quality. The endogenous quality $q(x)$ of this final good changes with the chosen profile of skilled workers $x = (x^1, x^2, \dots, x^j, \dots, x^l) \geq 0$, where $x^j \geq 0$ is a number of workers in the coordinate j . To save space and for simplification, at each period, each employed skilled worker of type j (located in the coordinate j) utilizes one unit of a specific non durable mean to produce, using his specific know-how, a unit of a specific component j . Then, the entrepreneur combines the quantities $x = (x^1, x^2, \dots, x^j, \dots, x^l)$ of these different components, to produce $q(x)$ units of a final good of endogenous quality $s(x)$. This general production function mixes both quantitative and qualitative variables. Our formulation generalizes the O-Ring production function which appeared in the O-Ring theory of the firm; see Kremer [18]. The revenue of the entrepreneur is $\varphi[q(x), s(x)]$. 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 employee 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 $g(x) = \varphi[q(x), s(x)] - \rho(x) \in \mathbb{R}$.

Advantages to change: Let $x = x^k$ and $y = x^{k+1}$, be respectively, the last period, and current period profiles of skilled workers chosen by the entrepreneur. Then, if this is the case, his advantages to change his profile of skilled workers from one period to the next is $A(x, y) = g(y) - g(x) \geq 0$;

Inconveniences to change: They represent the difference $I(x, y) = C(x, y) - C(x, x) \geq 0$ between costs $C(x, y)$ to be able to change from profile x to profile y and costs $C(x, x)$ to be able to stay with the same profile x used in the last period; *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 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 $h^j_{\underline{\underline}}} \geq 0$. These conservation costs include knowledge regeneration and motivation costs. Then, in the current period, (i) costs to conserve the same profile of workers as in the last period are $C(x, x) = \sum_{j=1}^n h^j_{\underline{\underline}}} x^j$ while, (ii) costs to utilize the profile of skilled workers y are:

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

where $J_+(x, y) = \{j \in J, y^j \geq x^j\}$ and $J_-(x, y) = \{j \in J, y^j < x^j\}$. For simplification, suppose that conservation costs are zero, i.e., $h^j_{\underline{\underline}}} = 0$. Then,

$$C(x, x) = 0 \quad \text{and} \quad I(x, y) = \sum_{j \in J_+(x, y)} h^j_+(y^j - x^j) + \sum_{j \in J_-(x, y)} h^j_+(x^j - y^j).$$

So, $I(x, y)$ is a quasi-distance $q(x, y) := I(x, y) \geq 0$ such that

- (i) $q(x, y) = 0$ iff $y = x$,
- (ii) $q(x, z) \leq q(x, y) + q(y, z)$, $x, y, z \in X$.

The more general case where $h^j_{\underline{\underline}}} > 0$ works as well.

Motivation and resistance to change functions: They are, moving from the past profile of knowledge workers x to the current profile y :

$$M(x, y) = U[A(x, y)] = [g(y) - g(x)]^\mu \quad \text{and} \\ R(x, y) = D[I(x, y)] = q(x, y)^\nu, \quad \mu, \nu > 0,$$

where the utility and disutility functions are $U[A] = A^\mu$ and $D[I] = I^\nu$.

Relative resistance to change function: It is $\Gamma[q(x, y)] = U^{-1}[D[I(x, y)]] = q(x, y)^{\nu/\mu}$, $\nu/\mu > 0$.

Worthwhile changes: In this setting, a change from profile x to y is worthwhile if $M(x, y) \geq \xi R(x, y)$, i.e., $[g(y) - g(x)]^\mu \geq \xi q(x, y)^\nu$, where $\xi > 0$ is the current and chosen “worthwhile enough” satisfying ratio. Then, a worthwhile change is such that $y \in W_\xi(x)$ iff

$$g(y) - g(x) \geq \lambda \Gamma[q(x, y)], \quad \lambda = (\xi)^{1/\mu} > 0.$$

Succession of worthwhile temporary stays and changes: In this example they are:

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

Variational traps: In the example, given the initial profile of skilled workers $x^0 \in X$, and a final worthwhile enough to change ratio $\lambda_* > 0, x^* \in X$ is a variational trap if it exists a path of worthwhile temporary stays and changes $\{x^0, x^1, \dots, x^k, x^{k+1}, \dots\}$ such that,

- (i) $g(x^{k+1}) - g(x^k) \geq \lambda_{k+1} \Gamma[q(x^k, x^{k+1})]$,
- (ii) $g(y) - g(x^*) < \lambda_* \Gamma[q(x^k, y)]$, $y \neq x^*, y \in X, k \in \mathbb{N}$;

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 that the quasi-distance $C(x^k, x^{k+1})$ converges to zero as k goes to infinite.

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

3 Inexact Proximal Algorithms as Worthwhile Stay and Change Processes

3.1 Inexact Proximal Formulation of Worthwhile Changes

Proximal intransitive preferences. Let us define, in the current period $k + 1$, the “to be increased” entrepreneur proximal payoff to change from $x = x^k$ to $y = x^{k+1}$ as $Q_\lambda(x, y) = g(y) - \lambda \Gamma[q(x, y)]$ with $\lambda > 0$. Then, the proximal payoff to stay at $x = x^k = y = x^{k+1}$ is $Q_\lambda(x, x) = g(x) - \lambda \Gamma[q(x, x)] = g(x)$. It follows that it is worthwhile to change from profile x to y iff $Q_\lambda(x, y) \geq Q_\lambda(x, x)$, i.e., $y \in W_\lambda(x)$. This defines a variable and possibly non transitive preference $z \geq_{x,\lambda} y \iff Q_\lambda(x, z) \geq Q_\lambda(x, y)$.

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, $f(x) = \bar{g} - g(x) \geq 0$, where $\bar{g} = \sup \{g(y), y \in X\} < +\infty$ is the highest finite profit that the entrepreneur can hope to get. Then, the “to be decreased” proximal payoff of the entrepreneur is

$$P_\lambda(x, y) = f(y) + \lambda \Gamma[q(x, y)]. \tag{1}$$

In this case, to move from profile x to profile y is a worthwhile change $y \in W_\lambda(x)$ iff $P_\lambda(x, y) \leq P_\lambda(x, x)$.

Sufficient descent methods. The entrepreneur performs, each period $k + 1$, a sufficient descent, if he can choose a new profile x^{k+1} such that $f(x^k) - f(x^{k+1}) \geq \lambda_{k+1} \Gamma[q(x^k, x^{k+1})]$. This means that the entrepreneur follows a path of worthwhile changes $x^{k+1} \in W_{\lambda_{k+1}}(x^k), k \in \mathbb{N}$. Since $q(x^k, x^k) = 0$, this comes from definition of $W_{\lambda_{k+1}}(x^k)$ combined with (1) for $x = x^k, y = x^{k+1}$ and $\lambda = \lambda_{k+1}$. In this case, each new worthwhile change is not necessarily optimal, contrary to each step of an exact proximal algorithm.

Exact proximal algorithms. The entrepreneur follows an exact proximal algorithm if, at each current period $k + 1$, he can choose a new profile x^{k+1} which minimizes his “to be decreased” proximal payoff $P_{\lambda_{k+1}}(x^k, y) = f(y) + \lambda_{k+1}\Gamma[q(x^k, y)]$ on the whole space X ,

$$x^{k+1} \in \operatorname{argmin}_{y \in X} \left\{ f(y) + \lambda_{k+1}\Gamma[q(x^k, y)] \right\}, \quad k \in \mathbb{N}, \tag{2}$$

which allows us to obtain $x^{k+1} \in W_{\lambda_{k+1}}(x^k), k \in \mathbb{N}$. In Mathematics the formulation is

$$x^{k+1} \in \operatorname{argmin}_{y \in X} \left\{ f(y) + \lambda_k\Gamma[q(x^k, y)] \right\}, \quad k \in \mathbb{N}. \tag{3}$$

It takes λ_k instead of λ_{k+1} . In this case, the entrepreneur follows a path of optimal worthwhile changes, $x^{k+1} \in W_{\lambda_k}(x^k), k \in \mathbb{N}$. In this paper, we will adopt the mathematical formulation.

Epsilon inexact proximal algorithms. Several variants about this subject can be found in the literature. Let us consider the version given in Attouch and Soubeyran [3] following a long tradition, starting with Rockafellar [19]. In our context, the entrepreneur follows an inexact proximal algorithm if, at each period $k + 1$, he can choose a new profile x^{k+1} such that

$$f(x^{k+1}) + \lambda_k\Gamma[q(x^k, x^{k+1})] \leq f(y) + \lambda_k\Gamma[q(x^k, y)] + \varepsilon_k, \quad y \in X,$$

given a sequence of nonnegative error terms $\{\varepsilon_k\}$, i.e., $P_{\lambda_k}(x^k, x^{k+1}) \leq P_{\lambda_k}(x^k, y) + \varepsilon_k, y \in X$. The term λ_k can be replaced by λ_{k+1} .

Epsilon inexact proximal algorithms represent a succession of adaptive satisfying processes. Let $\overline{Q}_{\lambda_k}(x^k) = \sup \{Q_{\lambda_k}(x^k, y) : y \in X\} < +\infty$ and $\underline{P}_{\lambda_k}(x^k) = \inf \{P_{\lambda_k}(x^k, y) : y \in X\} > -\infty$ be, for each current period $k + 1$, the optimal past values of the “to be increased” and “to be decreased” proximal payoffs of this entrepreneur. Let $\overline{Q}_{\lambda_k}(x^k) - s_{k+1}$ and $\underline{P}_{\lambda_k}(x^k) + s_{k+1}$ be, in this current period $k + 1$, the current satisfying levels of the “to be increased” and “to be decreased” proximal payoffs of the entrepreneur. In this current period, $s_{k+1} > 0$ represents, for the VR approach, a given satisfying rate; see [1, 2]. For an inexact proximal algorithm, $s_{k+1} = \varepsilon_k > 0$ is a given error term. Then, in the context of the VR theory, an inexact proximal algorithm has a new interpretation. It means that, for each period $k + 1$, the new profile x^{k+1} must be satisfying. That is to say, “to be increased” and “to be decreased” proximal payoffs of the entrepreneur must be higher or lower than the current satisfying level, i.e., $Q_{\lambda_k}(x^k, x^{k+1}) \geq \overline{Q}_{\lambda_k}(x^k) - \varepsilon_k$ for a “to be increased” proximal payoff and $P_{\lambda_k}(x^k, x^{k+1}) \leq \underline{P}_{\lambda_k}(x^k) + \varepsilon_k$ for a “to be decreased” proximal payoff. For each period $k + 1$, let us consider the variable satisfying set $S_{\lambda_k, \varepsilon_k}(x^k) = \{y \in X : P_{\lambda_k}(x^k, y) \leq \underline{P}_{\lambda_k}(x^k) + \varepsilon_k\}$. Then, an epsilon inexact proximal algorithm is defined by a succession of repeated decision making problems with changeable spaces and goals (satisfying levels): find $y \in S_{\lambda_k, \varepsilon_k}(x^k), k \in \mathbb{N}$. They are decision making problems with changeable spaces. See Larbani and Yu [15] for different aspects of what can be changed and how (their DMCS approach).

3.2 Marginally Worthwhile Changes

Consider the current period $k+1$. Let $x^k \rightsquigarrow y = x^{k+1}$ be a worthwhile change from x^k to $x^{k+1} \in W_{\lambda_k}(x^k)$ and let $x^{k+1} \rightsquigarrow z \in \mathfrak{M}(x^{k+1}) \subset X$ be a marginal change, where $\mathfrak{M}(x^{k+1})$ is a small neighborhood of x^{k+1} in the quasi-metric space X . Then, at each period $k+1$, the agent who has done the worthwhile change $y = x^{k+1} \in W_{\lambda_k}(x^k)$ will stop to prolong this change if, taking one more step in this period $k+1$, from x^{k+1} to $z \in \mathfrak{M}(x^{k+1})$, this marginal change is not worthwhile, i.e., $z \notin W_{\lambda_k}(x^{k+1})$. This is a generalized stopping rule, a “not worthwhile marginal change” condition, that will be used later in the context of proximal algorithms; see condition (12).

3.3 The Separation Between Weak and Strong Resistance to Change

Two cases. The consideration of relative resistance to change functions $\Gamma[\cdot]$ helps to classify proximal algorithms in two separate groups. The first case is that of strong resistance to change, where $\Gamma[q] = q$ for all $q \geq 0$. This case has been examined in [1–3]. The second case is that of weak resistance to change, where $\Gamma[q] = q^2$ and $q = q(x, y)$ is a distance and not a quasi-distance. This is the traditional case. The literature on this topic is enormous; see, for example, Moreau [20] and Martinet [21], as well as in the study of variational inequalities associated to maximal monotone operators; see Rockafellar [19]. The variational approach, which considers relative resistance to change as a core concept, balances motivation and resistance to change, provides us with an extra motivation to further develop the study of proximal algorithms in a nonconvex and possibly nonsmooth setting, where the perturbation term of the usual proximal point algorithm becomes a “curved enough” function of the quasi-distance between two successive iterates. Soubeyran [1, 2] and, later, Bento and Soubeyran [7], in a first paper which paved the way for the present one, have shown the strong link between a relative resistance to change index with the famous “loss aversion” index; see [8, 9]. The generalized proximal algorithm, examined both in [7] and in the present paper, is new and more adapted for applications in Behavioral Sciences. Moreover, it retrieves recent approaches of the proximal method for nonconvex functions; see [10, 12].

Assumption on the relative resistance to change. In the remainder of this paper we assume that Γ is a twice differentiable function such that:

$$\Gamma[0] = \Gamma'[0] = 0, \quad \text{and} \quad \Gamma'[q] > 0, \quad \Gamma''[q] > 0, \quad q > 0, \quad (4)$$

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

$$\Gamma'[q/r] \leq \bar{\rho}_\Gamma(r) \Gamma[q]/q, \quad 0 < q \leq \bar{q}. \quad (5)$$

Let us consider the following Γ -generalized rate of curvature:

$$\rho_\Gamma(q, r) := \frac{\Gamma'[q/r]}{(\Gamma[q]/q)}, \quad 0 < q \leq \bar{q}. \quad (6)$$

In the particular case $r = 1$, (6) represents, in Economics, the elasticity of the disutility curve Γ ; see, for instance, [1,2]. From (6), condition (5) is equivalent to the condition:

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

Let us consider, for each $\alpha > 1$ fixed, 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

$$\bar{\rho}_\Gamma(q, r) = \alpha r^{1-\alpha} = \bar{\rho}_\Gamma(r) < +\infty. \tag{7}$$

More accurately, for each $\alpha > 1$, $\Gamma[q] = q^\alpha$ represents a disutility of inconveniences to change. It is strictly increasing and satisfies (4) and (5).

4 An Inexact Proximal Point Algorithm: Convergence to a Weak or Strong Variational Trap

4.1 End Points as Critical Points or Variational Traps

In a first paper, Bento and Soubeyran [7] showed when, in a quasi-metric space, a generalized inexact proximal algorithm, equipped with a generalized perturbation term $\Gamma [q(x, y)]$, and defined at each step by: (i) A sufficient descent condition and, (ii) a stopping rule and convergence to a critical point. Then, they have shown that the speed of convergence and the convergence in finite time depend on the curvature of the perturbation term and on the Kurdyka–Lojasiewicz property associated to the objective function. A striking and new application has been given. It concerns the impact of the famous “loss aversion effect,” see [8,9], on the speed of convergence of the generalized inexact proximal algorithm. However, in the context of the “variational rationality approach,” which considers, as central dynamical concepts, worthwhile stay and change processes, these important results in Applied Mathematics are not enough, from the viewpoint of our applications to Behavioral Sciences, unless we can show that this critical point is a stationary and variational trap (strong or weak) where the agent will prefer to stay than to move, because his motivation to change is strictly or weakly lower than his resistance to change. This section presents, under the conditions of [7, Theorem 3.1], a worthwhile stay and change process which converges to a critical point of f which is a weak trap (compare, below, with the definition of a strong stationary and variational trap). Then, let us start with the general definition of a weak stationary and variational trap instead of a strong one ([1,2]).

Definition 4.1 Let $x \in X$ be a given action and $\xi > 0$ be a satisfying rate of change chosen by the agent. Let $W_\xi(x) := \{y \in X, M(x, y) \geq \xi R(x, y)\}$ be his worthwhile to change set, starting from $x \in X$. Then, starting from $x^* \in X$ with a given satisfying worthwhile to change rate $\xi^* > 0$, a strong stationary trap $x^* \in X$ is such that motivation to change is strictly lower than resistance to change, $M(x^*, y) < \xi_* R(x^*, y)$ for all $y \neq x^* \in X$. A weak stationary trap is such that $M(x^*, y) \leq \xi_* R(x^*, y)$, for

all $y \in X$. This defines the stationary side of a trap. The variational aspect comes from being the end of a worthwhile to change process, starting from an initial given point.

- Remark 4.1* (a) Notice that a strong stationary trap is such that $W_{\xi_*}(x^*) = \{x^*\}$ and a weak stationary trap is such that $W_{\xi_*}(x^*) = \{y \in X, M(x^*, y) = \xi_* R(x^*, y)\}$. Being at a strong (weak) stationary trap, the agent strictly (weakly) prefers to stay than to move. Then, when a process of worthwhile stays and changes converges to a strong variational trap, this variational formulation defines, starting from an initial point, a variational trap as the end point of a path of worthwhile changes, worthwhile to approach, but not worthwhile to leave. This happens because, starting from there, there is no way to do any other worthwhile change, except repetitions;
- (b) Assuming that $\{\lambda_k\}$ converges to λ_∞ , our sufficient condition proposes an algorithm which, following a succession of worthwhile changes $x^{k+1} \in W_{\lambda_k}(x^k), k \in \mathbb{N}$, converges to a weak stationary trap x^* such that $W_{\lambda_\infty}(x^*) = \{y \in X : M(x^*, y) = \lambda_\infty R(x^*, y)\}$. Since the agent is free to choose all his satisfying worthwhile to change rates λ_k in an adaptive way, this will show that the agent, choosing at the limit point x^* a satisfying worthwhile to change rate $\lambda_* > \lambda_\infty$, ends in a strong stationary trap x^* , because $M(x^*, y) = \lambda_\infty R(x^*, y) < \lambda_* R(x^*, y)$, for all $y \in X$;
- (c) As observed in Sect. 2, in the specific context of this paper, we have

$$M(x, y) = U [A(x, y)] = f(x) - f(y), \quad R(x, y) = D [C(x, y)],$$

$$\Gamma [q(x, y)] = \lambda U^{-1} [D[q(x, y)]], \xi = 1.$$

Then, in our present paper, a strong (resp;weak) stationary trap is such that $f(x^*) - f(y) < \lambda \Gamma [q(x^*, y)]$, for all $y \neq x^*$ (resp. $f(x^*) - f(y) \leq \lambda \Gamma [q(x^*, y)]$, for all $y \in X$).

4.2 Some Definitions from Subdifferential Calculus

In this section, some elements concerning the subdifferential calculus are recalled; see, for instance, Rockafellar and Wets [22]. Assume that $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function. The domain of f , which we denote by $\text{dom} f$, is the subset of \mathbb{R}^n on which f is finite-valued. Since f is proper, then $\text{dom} f \neq \emptyset$.

Definition 4.2 (i) The Fréchet subdifferential of f at $x \in \mathbb{R}^n$, denoted by $\hat{\partial} f(x)$, is the set:

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

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

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

Throughout the paper, we consider the subdifferential ∂f since it satisfies a closedness property important in our convergence analysis, as well as in any limiting processes used in an algorithmic context.

A necessary condition for a given point $x \in \mathbb{R}^n$ to be a minimizer of f is

$$0 \in \partial f(x). \tag{8}$$

It is known that, unless f is convex, (8) is not a sufficient condition. The domain of ∂f , which we denote by $\text{dom } \partial f$, is the subset of \mathbb{R}^n on which ∂f is a nonempty set. In the remainder, a point that satisfies (8) is called a limiting-critical or simply critical point.

4.3 The Algorithm

In [3] the authors examined the “local epsilon inexact proximal” algorithm,

$$f(x^{k+1}) + \lambda_k d(x^k, x^{k+1}) \leq f(y) + \lambda_k d(x^k, y) + \varepsilon_k, \quad y \in E(x^k, r_k) \subset X,$$

where: (i) d is a distance, (ii) $E(x^k, r_{k+1}) \subset X$ is a variable choice set (a moving ball) for each current period $k + 1$. Following [3] we consider the so-called global epsilon inexact proximal algorithm as follows: starting from the current position x^k , let us define the next iterate x^{k+1} as follows:

$$f(x^{k+1}) + \lambda_k \Gamma[q(x^k, x^{k+1})] \leq f(y) + \lambda_k \Gamma[q(x^k, y)] + \varepsilon_k, \quad y \in X, \tag{9}$$

where $\{\lambda_k\}, \{\varepsilon_k\}$ are given sequences of nonnegative real numbers, and q is a quasi-distance. In the particular case where the generalized perturbation term $\Gamma[q(x, y)] = q(x, y)^2$ and $q(x, y) = d(x, y)$ is a distance, instead of a quasi-distance, our “global epsilon inexact proximal” algorithm coincides with the case considered by Zaslavski [23].

Assumption 4.1 There exist $\beta_1, \beta_2 \in \mathbb{R}_{++}$ such that: $\beta_1 \|x - y\| \leq q(x, y) \leq \beta_2 \| -y \|$, $x, y \in \mathbb{R}^n$.

This is the case in our knowledge management example. For another explicit example where inconveniences to change are a quasi-distance satisfying Assumption 4.1, see [12].

Next, we recall the inexact version of the proximal point method introduced in [7].

Algorithm 4.1 Take $x^0 \in \text{dom } f, 0 < \bar{\lambda} \leq \tilde{\lambda} < +\infty, \sigma \in [0, 1[$ and $b > 0$. For each $k = 0, 1, \dots$, 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:

$$f(x^k) - f(x^{k+1}) \geq \lambda_k(1 - \sigma)\Gamma[q(x^k, x^{k+1})], \tag{10}$$

$$w^{k+1} \in \partial f(x^{k+1}), \quad v^{k+1} \in \partial q(x^k, \cdot)(x^{k+1}), \tag{11}$$

$$\|w^{k+1}\| \leq b\Gamma'[q(x^k, x^{k+1})] \|v^{k+1}\|, \tag{12}$$

The first condition is a sufficient descent condition. It is a (proximal-like) worthwhile to change condition $x^{k+1} \in W_{\xi_{k+1}}(x_k)$, where the proximal perturbation term defines the relative resistance to change function. This condition tells us that it is worthwhile to change from x^k to x^{k+1} , rather than to stay at x^k . In this case, advantages to change from x^k to x^{k+1} , $A(x^k, x^{k+1}) = f(x^k) - f(x^{k+1})$ are, at each period, higher than some adaptive proportion $\xi_{k+1} = \lambda_k(1 - \sigma)$ of the relative disutility of inconveniences to change rather than to stay $\Gamma[q(x^k, x^{k+1})] = U^{-1}[D[I(x^k, x^{k+1})]]$, where, (i) inconveniences to change rather than to stay are $I(x^k, x^{k+1}) = C(x^k, x^{k+1}) - C(x^k, x^k) = q(x^k, x^{k+1})$, (ii) costs to be able to change from x^k to x^{k+1} are $C(x^k, x^{k+1}) = q(x^k, x^{k+1})$, while, (iii) costs to be able to stay $C(x^k, x^k) = q(x^k, x^k) = 0$ are zero as quasi-distances. The second condition defines subgradients of the objective and costs to be able to change functions. The third condition is a stopping rule which tells us, at each period, when the agent prefers not to make a new marginal change, because it is not worthwhile to do it, in this period; see Sect. 3.2 on marginally worthwhile changes.

Remark 4.2 As pointed out by the authors in [7], Algorithm 4.1 retrieves the inexact algorithm proposed in [11, Algorithm 2] in the particular case $\Gamma[q] = q^2/2$, $q(x, y) = \|x - y\|$ and $1 - \sigma = \theta$. Moreover, Algorithm 4.1 is a habituation/routinization process and any sequence generated from it is a path of worthwhile changes with parameter $\lambda_k(1 - \sigma)$ such that, at each step, it is marginally worthwhile to stop. The variational stopping rule condition raises the following question: when, marginally, a change stops to be worthwhile? This strongly depends on the shapes of the utility and disutility functions.

Comparing Algorithm 4.1 with the iterative process (9), we observe the following:

- (i) on one side, the iterative process (9) is much more specific than our Algorithm 4.1. Indeed the weak “worthwhile to change” condition (10) is replaced by the much stronger condition (9).
- (ii) on the other side, the iterative process (9) does not impose the “not worthwhile marginal change condition” (12) as the Algorithm 4.1 does.

Next we propose a new inexact proximal algorithm, combining a particular instance of (9) with the stopping rule (12).

Algorithm 4.2 Take $x^0 \in \text{dom} f$, $0 < \bar{\lambda} \leq \tilde{\lambda} < +\infty$, $\sigma \in [0, 1[$ and $b > 0$. For each $k = 0, 1, \dots$, 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:

$$f(y) - f(x^{k+1}) \geq \lambda_k \left[(1 - \sigma)\Gamma[q(x^k, x^{k+1})] - \Gamma[q(x^k, y)] \right], \quad y \in X, \tag{13}$$

$$w^{k+1} \in \partial f(x^{k+1}), \quad v^{k+1} \in \partial q(x^k, \cdot)(x^{k+1}), \tag{14}$$

$$\|w^{k+1}\| \leq b\Gamma' \left[q(x^k, x^{k+1}) \right] \|v^{k+1}\|. \tag{15}$$

Remark 4.3 This new inexact proximal algorithm imposes a stronger worthwhile to change condition than Algorithm 4.1, because it must be verified, at each period, for each $y \in X$. Setting $y = x^k$ gives the last worthwhile to change condition. The other two conditions remain unchanged. Note that the exact proximal algorithm (3) is a specific case of our new algorithm (it holds by taking $\sigma = 0$). The new inexact worthwhile to change condition is $P_{\lambda_k}(x^k, x^{k+1}) \leq P_{\lambda_k}(x^k, y) + \lambda_k \sigma \Gamma [q(x^k, x^{k+1})]$, for all $y \in X$.

As in [7, 10–12], our main convergence result is restricted to functions that satisfy the so-called Kurdyka–Lojasiewicz inequality; see, for instance, [24–26]. Next formal definition of Kurdyka–Lojasiewicz inequality can be found in [26], where it is also possible to find several examples and a good discussion over important classes of functions which satisfy the mentioned inequality.

Definition 4.3 A proper lower semicontinuous function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to have the Kurdyka–Lojasiewicz property at $\bar{x} \in \text{dom } \partial f$ if there exist $\eta \in]0, +\infty[$, a neighborhood U of \bar{x} and a continuous concave function $\varphi : [0, \eta[\rightarrow \mathbb{R}_+$ such that:

$$\varphi(0) = 0, \quad \varphi \in C^1(]0, \eta[), \quad \varphi'(s) > 0, \quad s \in]0, \eta[; \tag{16}$$

$$\varphi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1, \quad x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta], \tag{17}$$

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

In what follows, we assume that f is bounded from below, continuous on $\text{dom } f$ and KL function, i.e., a function which satisfies the Kurdyka–Lojasiewicz inequality at each point of $\text{dom } \partial f$.

Theorem 4.1 Assume that $\{x^k\}$ is a bounded sequence generated from Algorithm 4.2, \tilde{x} is an accumulation point of $\{x^k\}$ and Assumption 4.1 holds. Let $U \subset \mathbb{R}^n$ be a neighborhood of \tilde{x} , $\eta \in]0, +\infty[$ and $\varphi : [0, \eta[\rightarrow \mathbb{R}_+$ a continuous concave function such that (16) and (17) hold. If $\delta \in (0, \bar{q})$ [see condition (5)] 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 the whole sequence $\{x^k\}$ converges to a critical point x^* of f which is a strong global trap, relative to the worthwhile to change set $W_{\lambda_*}(x^*)$, for any choice of the final satisfying rate $\lambda_* > \lambda_\infty$.

Proof The first part of the theorem follows immediately from [7, Theorem 3.1] because any sequence, generated from Algorithm 4.2, satisfies the conditions (10) and (12) of Algorithm 4.1. Let x^* be the limit point of the sequence $\{x^k\}$. Given that the sequence $\{\lambda_k\} \subset [\bar{\lambda}, \tilde{\lambda}]$ (it is bounded), $0 < \bar{\lambda} \leq \tilde{\lambda} < +\infty$, taking a subsequence, if necessary, we can assume that λ_k converges to a certain $\lambda_\infty \in]0, +\infty[$. For the second part, note that $\{f(x^k)\}$ is a non increasing sequence and $x^* \in \text{dom } f$. Now, given that $q(\cdot, y)$ is continuous for each $y \in X$, see [12], Γ is continuous and f is continuous on $\text{dom } f$,

taking the limit in (13) as k goes to infinity and assuming that λ_k converges to a certain $\lambda_\infty \in]0, +\infty[$, we get:

$$f(x^*) \leq f(y) + \lambda_* \Gamma[q(x^*, y)], \quad y \in X.$$

Therefore, the desired result follows from Remark 4.1. \square

5 Conclusions

In this paper, following [7], and using the recent variational approach presented in [1, 2], we have proposed a generalized “epsilon inexact proximal” algorithm that converges to a critical point which is also a variational trap. In Mathematics, our paper helps to show how the literature on proximal algorithms can be divided into two parts: the case of strong and weak relative resistance to change. In this paper we have considered the most difficult situation, the weak case. For Behavioral Sciences, our paper offers a dynamic model for habituation/routinization processes, and gives a striking and new result on the impact of the famous “loss aversion” index (see [8, 9]) on the speed of convergence of such processes. Given editorial constraints (lack of space in the present paper), this important result appears in the first paper [7]. In [13] the authors compare our VR variational rationality approach of inexact proximal algorithms to the HD habitual domain theory and the DMCS approach; see [14, 15]. Future research will consider the multiobjective case.

Acknowledgments The work was supported by CAPES, CAPES-MES-CUBA 226/2012, FAPEG 201210267000909–05/2012 and CNPq Grants 458479/2014-4, 471815/2012-8, 303732/2011-3, 236938/2012-6.

References

1. Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidexterity behaviors. GREQAM, Aix Marseille University (2009, pre-print)
2. Soubeyran, A.: Variational rationality and the “unsatisfied man”: routines and the course pursuit between aspirations, capabilities and beliefs. GREQAM, Aix Marseille University (2010, preprint)
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. Optim.* **18**, 1061–1081 (2007)
6. Cruz Neto, J.X., Oliveira, P.R., Soares Jr., P.A., Soubeyran, A.: Learning how to play Nash and alternating minimization method for structured nonconvex problems on Riemannian manifolds. *J. Convex Anal.* **20**, 395–438 (2013)
7. Bento, G.C., Soubeyran, A.: A generalized inexact proximal point method for nonsmooth functions that satisfies Kurdyka Lojasiewicz inequality. *Set-Valued Var. Anal.* (2015). doi:10.1007/s11228-015-0319-6
8. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. *Econometrica* **47**(2), 263–291 (1979)
9. Tversky, A., Kahneman, D.: Loss aversion in riskless choice: a reference dependent model. *Q. J. Econ.* **106**(4), 1039–1061 (1991)

10. 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)
11. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program. Ser. A* **137**, 91–129 (2012)
12. Moreno, F.G., Oliveira, P.R., Soubeyran, A.: A proximal algorithm with quasi distance. Application to Habit’s formation. *Optimization* **61**(12), 1383–1403 (2012)
13. Bento, G.C., Soubeyran, A.: Some Comparisons Between the Variational Rationality, Habitual Domain, and DMCS Approaches. [arXiv:submit/1077264](https://arxiv.org/abs/submit/1077264) (2014)
14. Yu, P.L.: *Forming Winning Strategies, An Integrated Theory of Habitual Domains*. Springer, Berlin (1990)
15. Larbani, M., Yu, P.L.: Decision making and optimization in changeable spaces, a new paradigm. *J. Optim. Theory Appl.* **155**(3), 727–761 (2012)
16. Simon, H.: A behavioral model of rational choice. *Q. J. Econ* **69**, 99–118 (1955)
17. Van Long, N., Soubeyran, A., Soubeyran, R.: Knowledge accumulation within an organization. *Int. Econ. Rev.* **55**(4), 1089–1128 (2014)
18. Kremer, M.: The O-ring theory of economic development. *Q. J. Econ.* **108**, 551–575 (1993)
19. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898 (1976)
20. Moreau, J.: Proximité et dualité dans un espace hilbertien. (French) *Bull. Soc. Math.* **93**, 273–299 (1965)
21. Martinet, B.: Régularisation, d’inéquations variationnelles par approximations successives. (French) *Rev. Française Informat. Recherche Opérationnelle* **4**(Ser. R–3), 154–158 (1970)
22. Rockafellar, R.T., Wets, R.: *Variational Analysis*, 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Berlin (1998)
23. Zaslavski, A.J.: Inexact proximal point methods in metric spaces. *Set-Valued Anal.* **19**, 589–608 (2011)
24. Lojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. *Les Équations aux Dérivées Partielles*, Éditions du centre National de la Recherche Scientifique, **117**, 87–89 (1963)
25. Kurdyka, K.: On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier* **48**, 769–783 (1998)
26. 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)