



HAL
open science

General descent method using w-distance. Application to emergence of habits following worthwhile moves

Antoine Soubeyran, João Carlos O. Souza

► To cite this version:

Antoine Soubeyran, João Carlos O. Souza. General descent method using w-distance. Application to emergence of habits following worthwhile moves. *Journal of Nonlinear and Variational Analysis*, 2020, 4 (2), pp.285-300. 10.23952/jnva.4.2020.2.09 . hal-03135309

HAL Id: hal-03135309

<https://amu.hal.science/hal-03135309>

Submitted on 19 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

GENERAL DESCENT METHOD USING W-DISTANCE. APPLICATION TO EMERGENCE OF HABITS FOLLOWING WORTHWHILE MOVES

ANTOINE SOUBEYRAN¹, JOÃO CARLOS DE O. SOUZA^{2,*}

¹*Aix-Marseille University (Aix-Marseille School of Economics), CNRS & EHESS, Marseille, France*

²*Department of Mathematics (CCN), Federal University of Piauí, Teresina, Piauí, Brazil*

Abstract. In this paper, we extend the general descent method proposed by Attouch, Bolte and Svaiter [Math. Program. 137 (2013), 91-129] to deal with possible asymmetric like-distances. Using a w-distance as regularization term, our results guarantee the convergence of bounded sequences under the assumption that the objective function satisfies the Kurdyka-Łojasiewicz inequality. In particular, it improves some existing works on proximal point methods with quasi-distance as regularization term because we prove convergence of bounded sequences without any additional assumption on the w-distance unlike it have been done with quasi-distances. The last section gives an application relative to the emergence of habits after a succession of worthwhile moves which balance motivation and resistance to move.

Keywords. Descent methods; Kurdyka-Łojasiewicz inequality; w-distance; habits; Variational rationality.

1. INTRODUCTION

In this paper, we show how first-order methods can modelize habits using the behavioral context of the recent variational rationality (VR) approach of worthwhile stay and change dynamics proposed by Soubeyran; see [24, 25, 26, 27] and more recently, [28, 29]. 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 characterizing all these theories in Behavioral Sciences, the variational rationality approach shows how to model 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).

In 2013, Attouch, Bolte and Svaiter [2] studied the convergence of a general descent method based on a first-order optimality condition which includes projection methods, proximal point method, Gauss-Seidel method, forward-backward methods, among others. This general descent method is unable to procure such an application to economics and social sciences. In fact, it uses the square of an Euclidean distance as a regularization term, which cannot represent a cost

*Corresponding author.

E-mail addresses: antoine.soubeyran@gmail.com (A. Soubeyran), joacos.mat@ufpi.edu.br (J.C.O. Souza).

Received January 10, 2020, Accepted July 7, 2020.

of moving or some inconveniences to move (as in our case) because in this setting costs of changing from x to y are equal to costs of changing from y to x , a very restrictive symmetric assumption for costs to move. Therefore, a quasi-distance (asymmetric distance) seems to be more appropriate as regularization term rather than the Euclidean norm.

Recently, Bento and Soubeyran [6, 7] showed when, in a quasi-metric space, a generalized inexact proximal algorithm (equipped with a generalized perturbation term) defined, at each step, by a sufficient descent condition and a stopping rule, converges to a critical point, the speed of convergence and the convergence in finite time depend on the curvature of the perturbation term and on the Kurdyka-Łojasiewicz property associated to the objective function. These works are based on the paper of Moreno, Oliveira and Soubeyran [18] who was the first to deal with proximal point method with a quasi-distance as regularization term. Other works have been considered proximal-type methods using a quasi-distance as regularization; see, e.g., [8, 9, 11].

As proposed in [18], all the previous mentioned works impose a restriction on the quasi-metric assuming a kind of “not too asymmetric” condition; see Section 2. Although such a condition makes sense in the VR approach, it restrains the class of quasi-distances, which can be used as regularizations. Actually, we will prove that this assumption implies the quasi-distance to be a particular instance of a w -distance; see Section 2. Thus, we enable not only the proximal point method but a general descent method, which includes projection methods, proximal point method, Gauss-Seidel method, forward-backward methods, among others, to deal with a w -distance as regularization term. An interesting application of the alternating Gauss-Seidel method in how to play Nash potential games can be found in Soubeyran, Souza and Cruz Neto [30]. In this context, our work generalizes and improves the above mentioned papers who perform first-order descent methods using quasi-distances and Euclidean norms as regularizations.

The rest of this paper is organized as follows. In Section 2, some basic definitions and results are stated. The method and its convergence analysis is presented in Section 3. Finally, Section 4 provides an application to emergence of habits following worthwhile moves.

2. PRELIMINARIES

Definition 2.1. A quasi metric space is a pair (X, q) such that X is a nonempty set, and $q : X \times X \rightarrow \mathbb{R}_+$, called a quasi metric or quasi distance, is a mapping satisfying:

- (Q1) For all $x, y \in X$, $q(x, y) = q(y, x) = 0 \Leftrightarrow x = y$;
- (Q2) For all $x, y, z \in X$, $q(x, z) \leq q(x, y) + q(y, z)$.

Clearly, metric spaces are quasi metric spaces satisfying the symmetric property $q(x, y) = q(y, x)$.

The works [6, 7, 11, 18] are devoted to studying algorithms via a quasi distance. Bento and Soubeyran [6] discussed how a proximal point method using a quasi-distance as regularization can be a nice tool to modelize the dynamics of human behavior in the context of the (VR) variational rationality approach; see Soubeyran [24, 25, 26, 27], and more recently in Soubeyran [28, 29]. Applications of quasi metric spaces to Behavioral Sciences (Psychology, Economics, Management, Game theory, etc.) and theoretical computer science can be found, for instance, in [3, 11] and references therein.

The following assumption has been made for algorithms dealing with quasi-distances. There exist real numbers $\alpha > 0$ and $\beta > 0$ such that

$$(H) \text{ (Not too asymmetric) } \alpha \|x - y\| \leq q(x, y) \leq \beta \|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

see, for instance, [11, 18] and the references therein.

Example 2.1. Let $c_1, c_2 \in \mathbb{R}$ be positive constants such that $c_1 \neq c_2$. Then, $q : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ given by

$$q(x, y) = \begin{cases} c_1(x - y), & \text{if } x \geq y, \\ c_2(y - x), & \text{if } x < y, \end{cases}$$

is a quasi-distance satisfying assumption (H) with $\alpha = \min\{c_1, c_2\}$ and $\beta = \max\{c_1, c_2\}$. Clearly, q is not symmetric and hence q is not a distance. See more examples in Moreno, Oliveira and Soubeyran [18].

Definition 2.2. A w-distance on a metric space (X, d) is a mapping $p : X \times X \rightarrow \mathbb{R}_+$ satisfying the following conditions:

- (W1) $p(x, y) \leq p(x, z) + p(z, y)$, for all $x, y, z \in X$;
- (W2) $p(x, \cdot) : X \rightarrow \mathbb{R}_+$ is lower semicontinuous for all $x \in X$;
- (W3) For each $\varepsilon > 0$, there exists $\delta > 0$ such that $p(x, y) < \delta$ and $p(x, z) < \delta$ imply $d(y, z) < \varepsilon$.

This concept was introduced by Kada and Suzuki [16]. Next, we give some classical examples of w-distance, which can be found, for instance, in [31] and Suzuki and Takahashi [32]. To this end, let X be a metric space with metric d . The application of w-distance in organizational change can be found in Bao, Khanh and Soubeyran [4].

Example 2.2. The mapping $p : X \times X \rightarrow \mathbb{R}_+$ given by $p(x, y) = c$, for every $x, y \in X$, where $c > 0$ is real number, is a w-distance.

Example 2.3. Let C be a bounded and closed subset of X containing at least two points and Δ a constant such that $\Delta \geq \text{diam } C$, where $\text{diam } C$ denotes the diameter of C . The function $p : X \times X \rightarrow \mathbb{R}_+$ given by

$$p(x, y) = \begin{cases} d(x, y), & \text{if } x, y \in C \\ \Delta, & \text{if } x \notin C \text{ or } y \notin C \end{cases}$$

is a w-distance.

Let X be a normed linear space with norm $\|\cdot\|$.

Example 2.4. The mapping $p : X \times X \rightarrow \mathbb{R}_+$ given by $p(x, y) = \|x\| + \|y\|$, for every $x, y \in X$ is a w-distance.

Example 2.5. The mapping $p : X \times X \rightarrow \mathbb{R}_+$ given by $p(x, y) = \|y\|$, for every $x, y \in X$ is a w-distance.

Remark 2.1. One can easily check that none of the above w-distances satisfies assumption (H).

For simplicity, from now on, we will consider the w-distances on the Euclidean metric space endowed with Euclidean norm $d(x, y) = \|x - y\|$. Now, let us examine what assumption (H) imposes to the quasi-distance in terms of a w-distance.

Remark 2.2. Let q be a quasi-distance satisfying assumption (H). Clearly, $p(x, y) = \|x - y\|$ is a w-distance on (\mathbb{R}^n, d) with $d(x, y) = \|x - y\|$, i.e., a distance d is a w-distance on (\mathbb{R}^n, d) . Then, from (W3), for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$p(x, y) = \|x - y\| < \delta \quad \text{and} \quad p(x, z) = \|x - z\| < \delta \quad \text{imply} \quad \|y - z\| < \varepsilon. \quad (2.1)$$

Thus, setting $\bar{\delta} = \alpha\delta$, from (H), we have

$$q(x, y) \leq \bar{\delta} \quad \text{and} \quad q(x, z) \leq \bar{\delta},$$

and hence,

$$\|x - y\| \leq \frac{q(x, y)}{\alpha} \leq \frac{\bar{\delta}}{\alpha} = \frac{\alpha\delta}{\alpha} = \delta$$

and

$$\|x - z\| \leq \frac{q(x, z)}{\alpha} \leq \frac{\bar{\delta}}{\alpha} = \frac{\alpha\delta}{\alpha} = \delta,$$

using the left hand-side inequality in (H) in inequalities above. Thus, from (2.2), we obtain that, for every $\varepsilon > 0$, there exists $\bar{\delta} > 0$ such that

$$q(x, y) < \bar{\delta} \quad \text{and} \quad q(x, z) < \bar{\delta} \quad \text{imply} \quad \|y - z\| < \varepsilon.$$

This is condition (W3) in the definition of w-distance. Note that a quasi-distance satisfies (W1) and $q(x, \cdot)$ is Lipschitz continuous (see [18, Proposition 3.6]), in particular, it satisfies (W2). Therefore, a quasi-distance, which satisfies assumption (H), is a w-distance. This means that algorithms dealing with a quasi-distance satisfying (H) actually are handling with a particular class of w-distances.

Let us recall some definitions and properties of the subdifferential theory which can be found, for instance, in [19, 21].

Definition 2.3. Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.

(1) The Fréchet subdifferential of f at x , denoted by $\hat{\partial}f(x)$, is defined as follows

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

(2) The limiting-subdifferential of f at x , denoted by $\partial f(x)$, is defined as follows

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

We denote by $\text{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. Recall that the limiting-subdifferential is closed and $\hat{\partial}f(x) \subset \partial f(x)$. If f is a proper, lower semicontinuous and convex function, and $x \in \text{dom}(f)$, then $\hat{\partial}f(x)$ coincides with the classical subdifferential in the sense of convex analysis and it is a nonempty, closed and convex set.

Dealing with descent methods for convex functions, we can expect that the algorithm provides globally convergent sequences, i.e., the convergence of the whole sequence. When the functions under consideration are neither convex nor quasiconvex, the method may provide sequences that exhibit highly oscillatory behaviors, and partial convergence results are obtained. The Kurdyka-Łojasiewicz property has been successfully applied to analyze various types of asymptotic behavior, in particular, proximal point methods; see, for instance, [2, 7, 13].

Definition 2.4. A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to have the Kurdyka-Łojasiewicz property (here called KL function) at $x^* \in \text{dom } \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of x^* and a continuous concave function $\varphi : [0, \eta] \rightarrow \mathbb{R}_+$ (called desingularizing function) such that:

$$\varphi(0) = 0, \quad \varphi \text{ is } C^1 \text{ on } (0, \eta), \quad \varphi'(s) > 0, \quad \forall s \in (0, \eta); \tag{2.2}$$

$$\varphi'(f(x) - f(x^*)) \text{dist}(0, \partial f(x)) \geq 1, \quad \forall x \in U \cap [f(x^*) < f < f(x^*) + \eta], \tag{2.3}$$

where $[\eta_1 < f < \eta_2] = \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\}$ and C^1 means differentiable with continuous derivative.

One can found the definition of the Kurdyka-Łojasiewicz property and other references on this subject in Attouch, Bolte and Svaiter [2].

Remark 2.3. One can easily check that the Kurdyka-Łojasiewicz property is satisfied at any non-critical point $\hat{x} \in \text{dom } \partial f$. It follows from the Kurdyka-Łojasiewicz property that the critical points of f lying in $U \cap [f(x^*) < f < f(x^*) + \eta]$ have the same critical value $f(x^*)$. If f is differentiable and $f(x^*) = 0$, then (2.3) can be rewritten as

$$\nabla(\varphi \circ f)(x) \geq 1,$$

for each convenient $x \in \mathbb{R}^n$. This property basically expresses the fact that a function can be made sharp by a reparameterization of its values; see [2].

Lemma 2.1 ([10], Lemma 4.1). *Let $\{\alpha_k\}$ be a sequence of positive numbers such that*

$$\sum_{k=1}^{+\infty} \frac{\alpha_k^2}{\alpha_{k-1}} < \infty.$$

Then, $\sum_{k=1}^{+\infty} \alpha_k < \infty$.

3. GENERAL DESCENT METHOD USING A W-DISTANCE

In this section, we study the convergence of a general descent method which uses as regularization a w-distance. The domain of applications of this abstract method includes first-order methods (such as gradient methods), projection methods, proximal point method, Gauss-Seidel method, forward-backward methods, among others; see [2].

(General descent method) We consider a and b two fixed positive constants. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a proper lower semicontinuous function. We suppose that f is bounded from below and its set of critical points S^* is non-empty. Without loss of generality, we can assume that $\inf_{x \in \mathbb{R}^n} f(x) = 0$. In the sequel, we consider a sequence $\{x^k\}$, which satisfies the following conditions:

(C1) (Sufficient descent condition) For each $k \in \mathbb{N}$,

$$f(x^{k+1}) + ap^2(x^k, x^{k+1}) \leq f(x^k);$$

(C2) (Relative error condition) For each $k \in \mathbb{N}$, there exists $w^{k+1} \in \partial f(x^{k+1})$ such that

$$\|w^{k+1}\| \leq bp(x^k, x^{k+1});$$

(C3) (Continuity condition) There exists a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ and $\hat{x} \in \mathbb{R}^n$ such that $x^{k_j} \rightarrow \hat{x}$ and $f(x^{k_j}) \rightarrow f(\hat{x})$ as $j \rightarrow +\infty$.

Remark 3.1. The general descent method was firstly considered in the context of KL functions in [2]. As mentioned in [2], condition (C1) was intended to model a descent property involving a measure of the quality of the descent. The use of a like-distance as regularization comes from the fact that it modelizes costs of moving and, more generally, inconveniences to move in Behavioral Sciences given in the presentation of the Variational Rationality approach, see Soubeyran [24, 25, 26, 27], and more recently in Soubeyran [28, 29]. A quick justification is that costs of moving are non symmetric, because costs of moving from x to y are not the same as costs of moving from y to x . Furthermore, the triangle inequality comes from the fact that detours are costly, because of some kind of fixed starting and stopping costs for each detour. A complete answer for this justification would be too long. It requires to have in mind the framework of the Variational Rationality approach, where the resistance to change plays a major role. The applications of this concept applied to the proximal point method can be found, for instance, in Cruz Neto et al. [11] in optimal size of the firm problem; Bento, Bitar and Cruz Neto [8] in group dynamic problems; Moreno Oliveira and Soubeyran [18] and Bento and Soubeyran [6, 7] in habit and routines formation; Bento et al [9] in compromise problem. It is worth to mention that none of these works deal with a w -distance as we do. All these works consider as regularization term the (square) Euclidean norm or a quasi-distance satisfying assumption (H). Both cases are particular instance of a w -distance. In this context, our results generalize the previously mentioned works.

Remark 3.2. Condition (C2) originates from the well-known fact that most algorithms in optimization are generated by an infinite sequence of subproblems which involve exact or inexact minimization processes. Condition (C2) corresponds to an inexact optimality condition. This assertion will be clear in the following example.

Example 3.1. Under some mild assumptions, the proximal point method applied to find a minimizer of a function f computes at each iteration a point x^{k+1} such that

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \right\}, \quad (3.1)$$

where $\{\lambda_k\}$ is an auxiliary parameter sequence. Using the first-optimality condition in (3.1), we have

$$\lambda_k(x^k - x^{k+1}) \in \partial f(x^{k+1}),$$

i.e., there exists $w^{k+1} \in \partial f(x^{k+1})$ such that $\|w^{k+1}\| = \lambda_k \|x^k - x^{k+1}\|$; see Rockafellar [22]. Comparing (3.1) with (C2), it is clear that x^{k+1} is computed exactly in (3.1) and inexactly in (C2).

Note that if $p(x^k, x^{k+1}) = 0$, then we have from (C2) that x^{k+1} is a critical point of f . So, this is a natural stopping rule for the method. Next, we prove some classical properties of descent methods but in the context of w -distances.

Theorem 3.1. *The following assertions hold:*

- (i) $\{f(x^k)\}$ is convergent;
- (ii) $\lim_{k \rightarrow +\infty} p(x^k, x^{k+1}) = 0$.

Proof. From (C1), we have that $\{f(x^k)\}$ is non-increasing. Since f is bounded from below, then it is convergent and the first assertion is proved. From (C1), we also have

$$p^2(x^k, x^{k+1}) \leq \frac{1}{a}[f(x^k) - f(x^{k+1})], \quad \forall k \in \mathbb{N}.$$

Thus,

$$\sum_{k=0}^n p^2(x^k, x^{k+1}) \leq \frac{1}{a}[f(x^0) - f(x^{n+1})] \leq \frac{f(x^0)}{a},$$

where the second inequality comes from the fact that f is bounded from below. Letting $n \rightarrow +\infty$, we obtain that $\sum_{k=0}^{+\infty} p^2(x^k, x^{k+1}) < \infty$, and hence, we have $\lim_{k \rightarrow +\infty} p(x^k, x^{k+1}) = 0$. \square

Theorem 3.2. *Assume that $\{x^k\}$ is bounded and f has the Kurdyka-Łojasiewicz property at the cluster point \hat{x} specified in (C3). Then, $\{x^k\}$ converges to \hat{x} which is a critical point of f .*

Proof. Let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ such that $x^{k_j} \rightarrow \hat{x}$ as $j \rightarrow +\infty$. Combining Theorem 3.1 (i) with (C3), we have that $f(x^k) \rightarrow f(\hat{x})$ as $k \rightarrow +\infty$ and $f(x^k) \geq f(\hat{x})$, for all $k \in \mathbb{N}$. The function f has the Kurdyka-Łojasiewicz property at \hat{x} . Then there exist φ , U , η as in Definition 2.4. Let $\delta > 0$ be such that $B(\hat{x}, \delta) \subset U$ and $\rho \in (0, \delta)$ (if necessary, shrink η so that $\eta < a(\delta - \rho)^2$; see [2, Corollary 2.8]). Since φ is continuous, we obtain that there exists $\bar{k}_0 \in \mathbb{N}$ such that $f(x^k) \in [f(\hat{x}), f(\hat{x}) + \eta)$, for all $k \geq \bar{k}_0$, and hence, $x^k \in U \cap [f(\hat{x}), f(\hat{x}) + \eta)$, for all $k \geq \bar{k}_0$ and some $k_0 \in \mathbb{N}$. Thus, using the fact that φ is concave, we have

$$\varphi(f(x^k)) - \varphi(f(x^{k+1})) \geq \varphi'(f(x^k))(f(x^k) - f(x^{k+1})) \geq a\varphi'(f(x^k))p^2(x^k, x^{k+1}), \quad (3.2)$$

where the second inequality comes from (C1). On the other hand, from (2.3), we have

$$\varphi'(f(x^k))\|w^k\| \geq 1, \quad \forall w^k \in \partial f(x^k).$$

Using this fact combined with (C2) in (3.2), we obtain

$$\varphi(f(x^k)) - \varphi(f(x^{k+1})) \geq \frac{a}{b} \frac{p^2(x^k, x^{k+1})}{p(x^{k-1}, x^k)}, \quad \forall k \geq k_0,$$

and hence,

$$\sum_{k=k_0}^n \frac{p^2(x^k, x^{k+1})}{p(x^{k-1}, x^k)} \leq \frac{b}{a} \varphi(f(x^{k_0})) - \varphi(f(x^{n+1})) \leq \frac{b}{a} \varphi(f(x^{k_0})).$$

Letting $n \rightarrow +\infty$ and applying Lemma 2.1, we have

$$\sum_{k=0}^{+\infty} p(x^k, x^{k+1}) < \infty. \quad (3.3)$$

Let $\{\alpha_n\}$ be the sequence of real numbers given by $\alpha_n = \sum_{k=n}^{+\infty} p(x^k, x^{k+1})$. From (3.3), we have that $\{\alpha_n\}$ converges to 0 as $n \rightarrow +\infty$. Therefore, for any $\delta > 0$, there exists $n_0 \in \mathbb{N}$ such that $\alpha_n < \delta$, for all $n \geq n_0$. On the other hand, given an arbitrary $\varepsilon > 0$, take $n_0 \in \mathbb{N}$ such that $p(x^{n_0}, x^{n_0}) \leq \alpha_{n_0} < \delta$ and $p(x^{n_0}, x^m) \leq \alpha_{n_0} < \delta$. Thus, from the definition of the w-distance, we have

$$\|x^n - x^m\| < \varepsilon, \quad \forall m, n \geq n_0,$$

i.e., $\{x^k\}$ is a Cauchy sequence. Therefore, $\{x^k\}$ is convergent and it converges to \hat{x} .

Now, from (C2), we have that $0 \leq \|w^{k+1}\| \leq bp(x^k, x^{k+1})$, where $w^{k+1} \in \partial f(x^{k+1})$. Letting $k \rightarrow +\infty$ and using Theorem 3.1 (ii), we have that $w^k \rightarrow 0$ as $k \rightarrow +\infty$. Using the closedness of the subdifferential $\partial f(\cdot)$, we obtain that $0 \in \partial f(\hat{x})$, and hence, \hat{x} is a critical point of f . \square

4. THE EMERGENCE OF HABITS FOLLOWING WORTHWHILE MOVES

In this last section, we give an application of our generalized descent method to the important problem of the emergence of habits, when an individual changes gradually his own behavior to end in a habit. In this way, an agent can promote gradually better habits, like food or health habits. This section represents a simplified but important application, among many possible others, of the variational rationality model of human dynamics (Soubeyran [24, 25, 26, 27, 28, 29]). For an initial application relative to habit formation, see Moreno, Oliveira and Soubeyran [18]. For a general and preliminary presentation of habit forming and breaking in a different context where preferences can change, see Soubeyran [27].

4.1. Modelling an enduring habit. The difficulty to define a habit. To define a habit is not easy, because it has a lot of different aspects. Then, in psychology, there are as much definitions of a habit as aspects an author wants to emphasize. To better see which aspects of a habit the mathematical part of this paper modelizes, we will list several definitions of a habit, and then, choose one on them.

- In psychology, a habit can be defined as:
 - a recurrent/repeated behavior: “a particular act or way of acting that you tend to do regularly” (Cambridge Dictionary), where recurrent and repeated are not synonym terms;
 - a recurrent/repeated and unconscious behavior: “something that you do often and regularly, sometimes without knowing that you are doing it” (Cambridge Dictionary);
 - a recurrent/repeated, acquired and fixed way of behaving: “a more or less fixed way of thinking, willing, or feeling acquired through previous repetition of a mental experience” (Andrews [1]);
 - an acquired and involuntary behavior: “an acquired mode of behaviour that has become nearly or completely involuntary” (Merriam-Webster’s online dictionary);
 - a recurrent/repeated, acquired by repetition and efficient behaviour pattern, “a behaviour pattern acquired by frequent repetition or physiologic exposure that shows itself in regularity or increased facility of performance” (Merriam-Webster’s online dictionary);
 - a context-behavior association in memory that develop as people repeatedly experience rewards for a given action in a given context; see Mazar and Wood [17]. We note that “most modern research begins with a conceptual definition of habits as cue-response associations in memory that are acquired slowly through repetition of an action in a stable circumstance” (Gardner [14], Orbell and Verplanken [20], Wood and Runger [33]);
 - an automatic response: “a settled tendency or usual manner of behaviour” (Merriam-Webster’s online dictionary). That is, “a habit can also be thought of as a link between a stimulus and a response. It serves as a mental connection between a

trigger thought or event (stimulus) and our response to that trigger (the response). Repeating this connection time and again forms a habit and affects all subsequent decisions and actions. If repeated often enough, this connection becomes near permanent unless we take conscious action to change it. For example, a stimulus for overeating might be stress. The stress may be physical, emotional or mental and triggered by such things as a restricted diet, tiredness, an argument, a bad day at work or even negative thinking. A learned response for dealing with this stress may be eating. Over time, the bond may become so strong that our automatic or habitual response to stress is to eat. In psychology, this is known as classical conditioning, as demonstrated by Pavlov's dogs. The dogs learnt to associate a tone with food and would salivate whenever they heard the tone whether there was food present or not."; see Joseph [15].

- The Free Dictionary gives the following definitions:
 - A recurrent, often unconscious pattern of behavior that is acquired through frequent repetition: made a habit of going to bed early.
 - An established disposition of the mind or character: a pessimistic habit.
 - Customary manner or practice: an early riser by habit.
 - An addiction, especially to a narcotic drug.
- The MacMillan Dictionary lists a lot of synonyms:
 - habit: something that you do often or regularly, often without thinking about it,
 - routine: your usual way of doing things, especially when you do them in a fixed order at the same time,
 - daily life: all the things that happen or that you do regularly,
 - custom: something that a particular person regularly does,
 - tendency: an attitude, habit, or situation that is starting to develop in a particular way,
 - ritual: something that you do regularly and always in the same way,
 - way of life: if something is a way of life, it is considered to be what people normally do or have,
 - common practice: something that is done a lot and is considered normal,
 - as is someone's wont: used for saying that someone has a habit of doing a particular thing,
 - proclivity: a tendency to want to do a particular thing, especially something bad.
 - the prevailing disposition or character of a person's thoughts and feelings (Merriam-Webster's online dictionary).

4.1.1. *A simplified definition of a habit.* We will define a habit as a learned repeated activity or bundle of activities done in the same context. As we will see, it fits well with the mathematical part. A more complete definition will include the fact that a habit is usually goal driven at the beginning, and becomes an unconscious and automatic response to the repeated occurrence of a given context (cue).

4.1.2. *A simple representation.* a habit is characterized by the repetition, each period, of the same context $e \in E$ (environment), the same activity or bundle of activities $a \in A$ (or routine) and the same per period reward/utility $g(x) \in \mathbb{R}$, where $x = (a, e) \in A \times E$ is a situated bundle of

activities. This is called the habit loop; see Duhigg [12]. This means that a habit x refers to the choice of a context, and the choice of an activity or a bundle of activities done in this context related to the evaluation of a reward $g(x)$.

4.1.3. *The stability of good habits, being where you want to be at the very beginning.* Let us define, within a period, good (optimal) habits and their utilities. Let $X = \mathbb{R}^n$ and $g : x \in X \mapsto g(x) \in \mathbb{R}$ be a per period utility (reward) function. Let $\bar{g} = \sup \{g(y), y \in X\} < +\infty$ be the highest utility an individual can expect from doing any situated bundle of activities $y \in X$ within a period. It represents the aspiration level of an individual. A situated bundle of activities $x^* \in X$ maximizes his per period utility if $g(x^*) \geq g(y)$ for all $y \in X$. It defines an optimal habit when, by chance, an individual, in the initial period 0, performs the optimal situated bundle of activities $x_0 = x^*$. In this context he will repeat x^* again and again, each successive period, that is, $x^k = x^{k+1} = x^*$, $k = 0, 1, \dots$, because it will not be advantageous to change this situated bundle of activities x^* . In this case this optimal situated bundle of activities (different given activities done in given different environments) becomes an optimal habit.

4.2. Gradual improving processes driven by motivation and resistance to move.

4.2.1. *Wanting to improve his utility, being not where he wants to be.* Suppose now that, contrary to the previous situation, an individual is not initially where he wants to be, wanting to move from bad habits to good habits. This is the case when, initially, in the first period 0, he does not perform the optimal situated bundle of activities x^* . That is, if $x_0 \neq x^*$, this individual, in order to improve his utility, must change this bundle of situated activities step by step, moving from performing x_0 to performing x^* . Let $f : X \rightarrow \mathbb{R}$ given by $f(y) = \bar{g} - g(y) \geq 0$ be the unsatisfied need of this individual at $y = x^0$, that is, a need for a higher utility which is equal to the discrepancy $f(y) = \bar{g} - g(y) \geq 0$. This gap represents the dissatisfaction feeling of this agent when he performs y instead of x^* . The problem of the individual being to reduce this discrepancy until it disappears at x^* where $f(x^*) = \bar{g} - g(x^*) = 0$. In this setting the problem of the individual is to minimize gradually his dissatisfied needs until he finds $x^* \in X$ such that $f(x^*) \leq f(y)$ for all $u \in X$. There are two benchmark cases:

- i) global and simultaneous improving processes;
- ii) local and sequential improving processes.

4.2.2. *Global and simultaneous improving processes.* From a practical point of view, consider a consumer and food habits. In economics, the neoclassical theory supposes that, in a first stage, a consumer can explore the whole space X of bundles of goods that we identify in this paper to bundles of situated consumption activities. Then, he will discover the whole graph of utilities, $g(y)$ for all bundle of goods $y \in X$. As a consequence he will find an optimal bundle of consumption goods $x^* \in X$ by direct comparison (brute force) of pairs $(g(x), g(y))$ and elimination of x or y , the one with the lowest utility. This is the essence of the dichotomy principle of substantive rationality (see Simon [23]) where optimization is ideally done in three steps:

- i) first exploration of the whole space to discover $\{g(y) : y \in X\}$;
- ii) comparison by pairs x, y and successive elimination of x or y to select x^* ;
- iii) finally, the bundle of goods x^* is consumed.

4.2.3. *Local and sequential improving processes.* Instead of considering a global optimization approach done in only one step, we will consider, using the variational rationality approach of human dynamics (Soubeyran [24, 25, 26, 27, 28, 29]), a local and sequential improving process driven by motivation and resistance to move. The individual will make a sequential exploration instead of a simultaneous ex ante exploration of the whole space. In this setting, each step, exploration must be local. These “local action” aspects are left for future research. This individual will try to improve step by step his utility, in a bounded rationality framework (limited resources, no way to know the utility function in one step, ...). The agent will not move from x^0 to a known and given optimal x^* , but will make a succession of bounded rational moves (no optimization, only local improvements) : he will improve enough in a given sense. The question is to know where this process ends (in a non optimal habit, which can be a trap or not, an inefficient or an optimal end). As Mazar and Wood [17, page 8] said “many habits begin with goal pursuit”. In this paper we will consider, as a rather vague goal, the desire to improve utility, starting from an undesirable initial situation.

4.2.4. *A variational rationality dynamic driven by motivation and resistance to move.* Let consider an individual who chooses to perform, in the current period, a bundle of situated activities $y = (y_1, \dots, y_m) \in X$, where each situated activity $y_j = (a_j, e_j) \in A \times E$ is defined by an activity a_j and the environment e_j where it is executed. This means that, each period, this individual chooses, both, the list of activities $a = (a_1, \dots, a_m) \in A$ he wants to do and the different environments (contexts) $e = (e_1, \dots, e_m) \in E$ where he plans to perform each activity. Following a succession of periods $k = 0, 1, \dots$:

- This individual starts to do an initial bundle of situated activities $x^0 \in X$.
- If this bundle represents a bundle of bad habits, this individual will try to change gradually his initial behavior x^0 , moving from doing x^0 in the first period to doing x^1 in the second period, ..., up to moving from doing the bundle of situated activities x^k in the previous period k to doing the same or a different bundle of situated activities $y = x^{k+1}$ in the current period $k + 1, \dots$. Each period, the individual can change some activities and some contexts (environments), or both, stopping, continuing, and starting to do some activities and stopping, continuing and starting to use some context. This individual will change his bundle of situated activities if $y \neq x$ (change) or will perform the same bundle (as previously) if $y = x$ (stay). A stay can start or prolong an habit.
- In this way, following a succession of periods $0, 1, \dots, k, k + 1, \dots$ and a succession of moves $x^0 \curvearrowright x^1 \curvearrowright \dots \curvearrowright x^k \curvearrowright x^{k+1} \dots$, he can hope to improve gradually his utility from $g(x^0)$ to $g(x^{k+1}) \geq g(x^k) \geq \dots \geq g(x^1) \geq g(x^0)$ until he reaches a better bundle of habits. Notice that his utility function $g : X \rightarrow \mathbb{R}$ defines the utility of performing a bundle of situated activities (a list of activities and their environments).
- Then, in the current period $k + 1$, if he chooses to move, he can benefit from advantages to move (change rather than stay) $A(x^k, y) = g(y) - g(x^k)$,
- These advantages to move will give him some motivation to move $M(x^k, y) = U [A(x^k, y)]$, where the mapping $U[\cdot] : A \in \mathbb{R}_+ \mapsto U[A] \in \mathbb{R}_+$ is a not decreasing value function.
- But moving (change or stay) being costly, this individual will bear costs of moving. These costs include, first, costs of changing $C(x^k, y) \in \mathbb{R}_{++}$ if $y \neq x^k$, and costs of staying $C(x^k, x^k) \in \mathbb{R}_{++}$ if $y = x^k$. Costs of moving (change or stay) are the sum of two different costs: capability costs (learning, maintenance and training costs), i.e., costs to

be able to do y (or x^k), and execution/exploitation costs, that is, costs to do y (or x^k). Each of these costs is strictly positive. Capability costs (costs to change capabilities) include three kinds of costs : costs to be able to stop, continue and start using resources, that is, costs to delete, conserve and acquire physiological, physical, cognitive, affective, social, and financial resources. Costs of moving can also include some fixed costs, like consideration costs of moving $E(x^k) > 0$ which represent exploration and evaluation costs of moving. These consideration costs are chosen and spend in advance, before choosing to move and, then, move in order to be able to choose between change or stay. All these costs model inertia. See Soubeyran [28, 29] for a long list of examples for each of these different costs. A major example of costs of moving are costs to break, prolong and build repetitive and automatic situated activities like habits (costs to change the habit loop including costs to change environments, activities and rewards).

- Inconveniences to move (change rather than stay) are $I(x^k, y) = T(x^k, y) + E(x^k)$, where $T(x^k, y) = C(x^k, y) - C(x^k, x^k) \geq 0$ represent transition costs of moving (change rather than stay), i.e., the excess (if non negative) of costs to change with respect to costs to stay.
- These inconveniences to move lead to some resistance to move $R(x^k, y) = D[I(x^k, y)]$, where $D[\cdot] : I \in \mathbb{R}_+ \mapsto D[I] \in \mathbb{R}_+$ is a not decreasing disutility function.

4.2.5. *Worthwhile moves.* Each period $k + 1$ an individual balances between change or stay (the question is: should I stay or should I go?) and must choose to change or to stay. The balance between motivation and resistance to move (change rather than stay) is

$$B_\xi(x^k, y) = M(x^k, y) - \xi R(x^k, y),$$

where $\xi > 0$ is a weight which bears upon resistance to move. If this balance is non negative, the move $x^k \curvearrowright y = x^{k+1}$ defines, in the current period $k + 1$, a worthwhile move such that motivation to move is high enough with respect to resistance to move when ξ is high enough. This individual can choose to move each time a move is worthwhile enough.

Then, if an individual starts from a bad habit x^0 , he can hope to reach a better one x^* after a succession of worthwhile moves, i.e., a worthwhile transition which modelizes a gradually improving process driven by motivation and resistance to move.

4.3. **The case of a linear quadratic variational rationality structure.** This paper assumes that $M = U[A] = A$ and $R = D[I] = I^2$ for all $A, I \in \mathbb{R}_+$. Then,

$$M(x^k, y) = A(x^k, y) = g(y) - g(x^k)$$

and

$$R(x^k, y) = D[I(x^k, y)] = D[C(x^k, y) - C(x^k, x^k) + E(x^k)] = I(x^k, y)^2.$$

In this linear quadratic context, the balance between motivation and resistance to move (change rather than stay) is $B_\xi(x^k, y) = A(x^k, y) - \xi I(x^k, y)^2$.

Costs of moving and hence inconveniences and resistances to move are very complex dynamic concepts because they have a lot of different aspects. See Soubeyran [24, 25, 26, 27, 28, 29], Moreno, Oliveira and Soubeyran [18], Bento, Bitar and Cruz Neto [8], Bao, Cobzaş and Soubeyran [3] and Bao, Khanh and Soubeyran [4] for further discussions. In this paper where

resistance to move is weak enough, resistance to move is the square of inconveniences to move, while motivation to move is identical to advantages to move.

It is worth to mention that w-distances (Taka, Suzuki and Takahashi [16], Suzuki [31] and Suzuki and Takahashi [32]) can model a lot of different aspects of costs of moving and inconveniences to move. More specifically, consider the case of inconveniences to move

$$I(x, y) = T(x, y) + E(x), \text{ where } T(x, y) = C(x, y) - C(x, x). \tag{4.1}$$

We will assume (see Soubeyran [28, 29] for more details) that transition costs (change rather than stay) $T : X \times X \rightarrow \mathbb{R}_+$ is a pseudo quasi distance; see [5]. This means that, for all $x, y \in X$,

- i) $T(y, x)$ can be different from $T(x, y)$, that is, transition costs are asymmetric;
- ii) If $x = y$, then $T(x, y) = 0$, i.e., transition costs are zero for a stay.

Additionally, we will assume the following assumptions:

- H1. $T(x, y) = C(x, y) - C(x, x) \geq 0$ for all $x, y \in X$;
- H2. $T(x, z) \leq T(x, y) + T(y, z)$ for all $x, y, z \in X$;
- H3. The mapping $E : X \rightarrow \mathbb{R}_+$ is lower semicontinuous.

Remark 4.1. H1 means that costs to change $C(x, y)$, with $y \neq x$, are not lower than costs to stay $C(x, x)$. H2 means that, when moving from x to z , transition costs of making a detour via the intermediate point y , i.e., $T(x, y) + T(y, z)$ are higher than transition costs $T(x, z)$ of directly moving from x to z . Finally, H3 is a regularity condition and from (4.1) one has that $I(x, y) = T(x, y) + E(x) \geq T(x, y)$, for every $x, y \in X$.

Next, we show that inconveniences to move is a w-distance under some mild assumptions.

Proposition 4.1. *Under the previous hypothesis H1, H2 and H3, inconveniences to move $I : X \times X \rightarrow \mathbb{R}_+$ is a w-distance.*

Proof. A slight extension of Suzuki and Takahashi [32, Lemma 1] shows that if $(X, T(\cdot, \cdot))$ is a pseudo quasi-metric space (instead of a metric space), then, if H1, H2 and H3 hold, $I(\cdot, \cdot)$ satisfies the three conditions

- (1) $I(x, z) \leq I(x, y) + I(y, z)$ for all $x, y, z \in X$;
- (2) For all $x \in X$, $I(x, \cdot) : y \in X \mapsto I(x, y) \in \mathbb{R}_+$ is lower semicontinuous;
- (3) $I(x, y) \geq T(x, y)$ for every $x, y \in X$;

then, $I(\cdot, \cdot)$ is a w-distance. In our case, $T(\cdot, \cdot)$ is a pseudo quasi-metric, because $T(x, x) = 0$. But $T(x, y) = 0$ does not imply $y = x$. □

4.4. The gradual emergence of habits. In this linear quadratic context, the balance between motivation and resistance to move (change rather than stay) is $B_\xi(x^k, y) = A(x^k, y) - \xi I(x^k, y)^2$. Then, a succession of worthwhile moves is such that $B_\xi(x^k, x^{k+1}) = A(x^k, x^{k+1}) - \xi I(x^k, x^{k+1})^2 \geq 0, k = 0, 1, \dots$

Let advantages to move be $A(x^k, x^{k+1}) = g(x^{k+1}) - g(x^k) = f(x^k) - f(x^{k+1})$ and let inconveniences to move be $I(x^k, x^{k+1}) = p(x^k, x^{k+1})$ for $k = 0, 1, \dots$. Then a succession of worthwhile moves $x^k \rightsquigarrow x^{k+1}$ verifies the condition, for $k = 0, 1, \dots$,

$$g(x^{k+1}) - g(x^k) = f(x^k) - f(x^{k+1}) \geq \xi p(x^k, x^{k+1})^2 \geq 0$$

which refers to a sufficient descent condition (C1) for each $k = 0, 1, \dots$ where $a = \xi > 0$.

In this context,

- i) This succession of worthwhile moves is an improving process such that utilities do not decrease and dissatisfactions do not increase, because $g(x^{k+1}) - g(x^k) = f(x^k) - f(x^{k+1}) \geq 0$ for $k = 0, 1, \dots$;
- ii) Advantages and inconveniences to move go to zero along this human dynamic driven by motivation and resistance to move;
- iii) Bounded from above utilities and non negative dissatisfactions converge to a limiting utility and limiting dissatisfactions g^* and f^* ;
- iv) Given the definition of a w-distance in a metric space (X, d) , bundles of activities become closer and closer in term of the w-distance p , hence in term of the distance d . The proof is easy. Suzuki [31, Lemma 3] tells us the following: let (X, d) be a metric space, p be a w-distance and $\{x^k\}, \{y^k\}, \{z^k\}$ be three sequences in X . Then, point (iii) of this lemma shows that $\lim_{k \rightarrow +\infty} p(x^k, y^k) = 0$ and $\lim_{k \rightarrow +\infty} p(x^k, z^k) = 0$ imply $\lim_{k \rightarrow +\infty} d(y^k, z^k) = 0$. Taking $y^k = x^k$ and $z^k = x^{k+1}$ shows that $\lim_{k \rightarrow +\infty} d(x^k, x^{k+1}) = 0$. This means that the distance between successive bundles of activities and environments x^k, x^{k+1} goes to zero. As a consequence, the individual repeats more and more similar habits loops, that is, more and more similar bundles of activities, environments and utilities (rewards). A habit forms gradually.
- v) the habit formation process ends when the relative error condition (C2) is satisfied, that is, when $p(x^k, x^{k+1}) \geq (1/b) \|w^{k+1}\|$, for some $w^{k+1} \in \partial f(x^{k+1})$, for each $k = 0, 1, \dots$. This means that inconveniences to move $p(x^k, x^{k+1})$ become higher enough (up to $1/b$) with respect to the norm $\|w^{k+1}\|$ of the marginal unsatisfaction w^{k+1} . This implies that $f(x^k) - f(x^{k+1}) \geq \xi p(x^k, x^{k+1})^2 \geq (\xi/b^2) \|w^{k+1}\|^2$, for some $w^{k+1} \in \partial f(x^{k+1})$. Then, the norm of the subgradient goes to zero with advantages to move. Suppose that the individual stops to be motivated to change as soon as his marginal advantage to move $w^{k+1} \in \partial f(x^{k+1})$ have a norm $\|w^{k+1}\| < \varepsilon$ lower than some given threshold level $\varepsilon > 0$. In this case there exists a period k^* where this individual will choose to stop moving. In all the periods following $k + 1$ a robust habit $x^* = x^{k+1} = x^h, h > k + 1$ is formed because the succession of worthwhile moves stops to be rewarding enough. The individual will choose automatically the same bundle of activities coupled with their environment and will benefit of the same reward.

Acknowledgements

The work of the second author was supported in part by CNPq (Grant 424169/2018-5) and UFPI-PROPEQI-PRPG (Promissão - 10/2018). It was carried out in part during a visit of the second author to Antoine Soubeyran at Aix-Marseille University (Aix-Marseille School of Economics, CNRS). The second author wishes to express his gratitude to Rose and Antoine Soubeyran for their hospitality during the author's visit to Aix-Marseille University (Aix-Marseille School of Economics).

REFERENCES

- [1] B.R. Andrews, Habit, *The American Journal of Psychology* 14 (1903), 121-149.
- [2] H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.* 137 (2013), 91-129.

- [3] T.Q. Bao, S. Cobzaş, A. Soubeyran, Variational principles, completeness and the existence of traps in behavioral sciences, *Ann. Oper. Res.* 269 (2018) 53-79.
- [4] T.Q. Bao, P.Q., Khanh, A. Soubeyran, Variational principles with generalized distances and the modelization of organizational change, *Optimization*, 65 (2016), 2049–2066.
- [5] T.Q. Bao, A. Soubeyran, Variational analysis in cone pseudo-quasimetric spaces and applications to group dynamics, *J. Optim. Appl.* 170 (2016), 458-475.
- [6] G.C. Bento, A. Soubeyran, Generalized inexact proximal algorithms: Routines formation with resistance to change, following worthwhile changes, *J. Optim. Theory App.* 166 (2015), 172-187.
- [7] G.C. Bento, A. Soubeyran, A generalized inexact proximal point method for nonsmooth functions that satisfies Kurdyka-Łojasiewicz inequality, *Set-Valued Var. Anal.* 23 (2015), 501-517.
- [8] G.C. Bento, S.D.B. Bitar, J.X. Cruz Neto, A. Soubeyran, J.C.O. Souza, A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems, *Comput. Optim. Appl.* 75 (2020), 263-290.
- [9] G.C. Bento, J.X. Cruz Neto, G. López, A. Soubeyran, J.C.O. Souza, The proximal point method for locally Lipschitz functions in multiobjective optimization with application to the compromise problem, *SIAM J. Optim.* 28 (2018), 1104-1120.
- [10] G.C. Bento, J.X. Cruz Neto, P.R. Oliveira, A New Approach to the Proximal Point Method: Convergence on General Riemannian Manifolds, *J. Optim. Theory App.* 168 (2016), 743-755.
- [11] J.X. Cruz Neto, P.R. Oliveira, A. Soubeyran, J.C.O. Souza, A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem, *Ann. Oper. Res.* 389 (2020), 313-339.
- [12] C. Duhigg, *The power of habit: why we do what we do in life and business*, Random House, New York, 2012.
- [13] P. Frankel, G. Garrigos, J. Peypouquet, Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates, *J. Optim. Theory App.* 165 (2015), 874-900.
- [14] B. Gardner, A review and analysis of the use of ‘habit’ in understanding, predicting and influencing health-related behaviour, *Health Psychology Rev.* 9 (2015), 277-295.
- [15] M. Joseph, *7 Steps To Make Or Break Habits*, Kindle Edition, 2012.
- [16] O. Kada, T. Suzuki, W. Takahashi, Nonconvex minimization theorems and fixed point theorems in complete metric spaces, *Math. Japonica* 2 (1996), 381-391.
- [17] A. Mazar, W. Wood, *Defining habit in psychology. The psychology of habit*, Springer, Cham, pp. 13-29, 2018.
- [18] F.G. Moreno, P.R. Oliveira, A. Soubeyran, A proximal point algorithm with quasi distance. Application to habit’s formation, *Optimization*, 61 (2012), 1383-1403.
- [19] B.S. Mordukhovich, Y. Shao, Nonsmooth Sequential Analysis in Asplund Spaces, *Trans. Amer. Math. Soc.* 348 (1996), 1235-1280.
- [20] S. Orbell, B. Verplanken, The automatic component of habit in health behavior: Habit as cue-contingent automaticity, *Health Psychology* 29 (2010), 374–38.
- [21] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
- [22] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. control optim.* 14 (1976), 877-898.
- [23] H. Simon, A behavioral model of rational choice, *Quarterly J. Economics* 69 (1955), 99-118.
- [24] A. Soubeyran, *Variational rationality, a theory of individual stability and change: worthwhile and ambidexterity behaviors*, GREQAM, Aix Marseille University, 2009.
- [25] A. Soubeyran, *Variational rationality and the “unsatisfied man”: routines and the course pursuit between aspirations, capabilities and beliefs*, GREQAM, Aix Marseille University, 2010.
- [26] A. Soubeyran, *Variational rationality. A theory of worthwhile stay and change approach-avoidance transitions ending in traps*, GREQAM, Aix Marseille University, 2016.
- [27] A. Soubeyran, *A variational rationality theory of habit formation and change*, AMSE, Aix-Marseille University, 2016.
- [28] A. Soubeyran, *Variational rationality. 1. An adaptive theory of the unsatisfied man*, AMSE, Aix-Marseille University, 2019.

- [29] A. Soubeyran, Variational rationality. 2. A general theory of goals and intentions as satisficing worthwhile moves, AMSE, Aix-Marseille University, 2019.
- [30] A. Soubeyran, J.C.O. Souza, J.X. Cruz Neto, A Generalized Proximal Alternating Linearized Method. In Variational Analysis and Set Optimization, 2019.
- [31] T. Suzuki, Several fixed point theorems in complete metric spaces, Yokohama Math. J. 44 (1997), 61-72.
- [32] T. Suzuki, W. Takahashi, Fixed point theorems and characterizations of metric completeness, Topol. Methods Nonlinear Anal. 8 (1996), 371-382.
- [33] W. Wood, D. Runger, Psychology of Habit, Ann. Rev. Psychology 67 (2016), 289-314.