A Proximal Algorithm with Quasi Distance. Application to Habit's Formation

Felipe Garcia Moreno[*]
PESC, COPPE, Universidade Federal do Rio de Janeiro, Brazil
fgarcia@cos.ufrj.br

Paulo Roberto Oliveira
PESC, COPPE, Universidade Federal do Rio de Janeiro, Brazil
GREQAM-SPORT
poliveir@cos.ufrj.br

Antoine Soubeyran
GREQAM-SPORT, Université d'Aix-Marseille II, France
antoine.soubeyran@univmed.fr

May 20, 2010

Abstract

We consider a proximal algorithm with quasi distance applied to nonconvex and nonsmooth functions involving analytic properties for an unconstrained minimization problem. We show the behavioral importance of this proximal point model for habit's formation in Decision and Making Sciences.

1 Introduction

In this paper, we consider the optimization problem:

$$\min\{f(x) : x \in \mathbb{R}^n\}, \tag{1.1}$$

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

The proximal algorithm has been first introduced by Martinet [17] and Rockafellar [21] as an approximation-regularization method in convex optimization and in the study of variational inequalities associated to maximal monotone operators.

[*]Partially supported by CAPES/CNPq - IEL Nacional - Brasil

In the proximal point method, iterates x^k, $k \geq 1$, are generated by the following rule:

$$x^{k+1} \in \arg\min \left\{ f(u) + \frac{1}{2\lambda_k}\|u - x^k\|^2 \;:\; u \in \mathbb{R}^n \right\}, \tag{1.2}$$

where x^0 is an initial guess for a minimizer, $\{\lambda_k\}_{k\in\mathbb{N}}$ is a positive sequence and $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$ is the usual norm.

In recent years it has been very active research in nonconvex programming and extensions of the proximal point method by using generalized distances. The scheme (1.2) has been studied to change the Euclidean norm by an adequate like distance such that not necesarily all the axioms of distance are verified, in particular the symmetry axiom, for example in the Bregman's distance case, but preserving the nice properties of convexity, continuity and coercivity of the Euclidean norm.

When we don't consider a like distance with such properties, we will have some difficulties in the convergence analysis because the convexity and the differentiablity are generality lost on a generalized distance framework, thus we have to proceed in a different form for the convergence analysis of generalized proximal methods. Literature connected with the analysis and development of proximal point methods in a convex and nonconvex setting includes [14, 13].

By another hand, quasi distances (or quasi-metrics) have been extensively studied in the topology context, see for example [15, 9, 28]. They generalize the distances in the sense that they are not symmetric. We also have that a quasi distance is not necesarily a convex, continuously differentiable and coercive function; so we can't proceed similarly to the Bregman case for the study of the convergence analysis when using the quasi distance for purpose of regularization. In addition to the applications of the quasi distances in the computer theory, see for example [8, 16], they also apply in economy, as consumer choice, utility functions[23] among others.

An important class of functions, in the nonlinear analysis, are the functions that verify the Kurdyka-Łojasiewicz inequality, which is a generalization of the Łojasiewicz gradient inequality [19, 1]. The Kurdyka-Łojasiewicz result has been used in the study of error bounds of analytic inequality systems in optimization [1, 7]. In turn, such error bounds have been used in the convergence analysis of optimization algorithms in the same general spirit as in the present paper; see, e.g., [1, 7, 4]. In [4], the authors used this property to obtain the convergence of the proximal method in a nonconvex and nonsmooth environment as well present an adequate analysis for development of generalized proximal methods, but in the Euclidean context only.

In this paper, we present a generalized proximal method to minimize a nonconvex and nonsmooth function, that verifies the Kurdyka-Łojasewicz property. The procedure has the following form

$$x^{k+1} \in \arg\min \left\{ f(u) + \frac{1}{2\lambda_k} q^2(x^k, u) \;:\; u \in \mathbb{R}^n \right\}, \tag{1.3}$$

where $q(\cdot, \cdot)$ denotes a quasi distance.

The goal of this paper is to establish the convergence of a general proximal point scheme (1.3) to a generalized critical point, and therefore we extend the result of

[4].

Our paper is organized as follows: after this introduction, we present a model of habit's formation which examines the role of inertia and experience in repeated consumption processes to motivate the use of proximal methods with quasi distances. In the section 3 we recall several properties of quasi distance, subdifferential theory, and Kurdyka-Łojasewicz inequality. In the section 4, the proximal algorithm with quasi distance is presented. Its convergence under various conditions is established in the section 5 followed by the conclusions.

2 A Model of Habit's Formation: the Role of Inertia and Experience in Repeated Consumption

Here we give an application to habit's formation of our proximal algorithm which uses as a regularization term the square of a quasi distance. In [2, 3] Attouch-Soubeyran have interpreted regularization terms as costs to change. Our present model of habit's formation describes a consumer who have i) a separable utility function which changes with experience, and ii) costs to change of consumption bundles (which modelize inertia).

The proximal algorithm of [4] is unable to procure such an application to economics and social sciences. It uses as a regularization term the square of an Euclidian distance which cannot represent a cost to move because in this case the cost to change from x to y is equal to the cost to change from y to x, a very restrictive symmetric assumption for a cost to move. In the present application we follow Soubeyran [26, 27] to modelize i) the adaptive aspect of our model (why variable preferences require repeated optimization), ii) costs to change as a quasi distance, and iii) the desutility of cost to change as an increasing convex function of costs to change.

A striking result of our habit's formation model is that, even when an agent ignores partially the variable part of his current utility function, because this variable part changes with experience in a bounded unpredicable way, and if the process starts not too far from the maximum of the static part of the utility function, the process converges to this maximum. Then, an habit forms gradually. Furthermore, when experience improves the variable part of the current utility (a learning process) the habit's formation process is efficient. The model supposes high costs to change in the small as a quasi distance, and a strong desutility of costs to change (a quadratic function of costs to change). The variable part of the utility function $U(y)$ must satisfy four hypothesis: $(\mathcal{H}_1)$ ($U(\cdot)$ is bounded below), $(\mathcal{H}_2)$ (the desutility of costs to change is coercitive, ie very high in the large), $(\mathcal{H}_3)$ (the restriction of the variable part of the utility function to its domain is continuous), $(\mathcal{H}_4)$ (for the Kurdyka-Lojasiewicz inequality). All these hypothesis are not too demanding from an economic point of view.

2.1 One step optimal consumption

Let $X = \mathbb{R}^n$ and $g(\cdot) : y \in X \longmapsto g(y) \in \mathbb{R}$. The global maximum problem is: find $x^* \in X$ such that $g(x^*) \geq g(y)$ for all $y \in X$. In our economic application $g(y) = U(y)$ represents for a consumer the utility of the consunption bundle $y \in X$ and x^* is a maximizing bundle of consumption goods. For the sake of our interpretation suppose that the supremum utility of consumption $\overline{g} = \overline{U} = \sup\{g(y), y \in X\} < +\infty$ is finite. Let $f(\cdot) : y \in X \longmapsto f(y) = \overline{g} - g(y) = \overline{U} - U(y) \geq 0$ be the unsatisfied need of this consumer at y (see Soubeyran,[25]). The global minimum problem is: find $x^* \in X$ such that $f(x^*) \leq f(y)$ for all $u \in X$. Then, x^* minimizes the unsatisfied needs of this consumer. At x^*, unsatisfied needs disappear: $f(x^*) = \overline{g} - g(x^*) = 0$. From a pratical point of view, if the consumer can explore in a first stage the whole space X of consumption bundles to discover the whole graph of utilities, $U(y)$ for all $y \in X$, he can find $x^* \in X$ by direct (brut force) comparison of pairs $(U(x), U(y))$ and elimination of the bundle, x or y, with the lowest utility. This is the essence of the dichotomy principle of substantive rationality (Simon, [24]) where optimization is ideally done in three steps :

i) first exploration of the whole space to discover $\{U(y), y \in X\}$,

ii) comparison by pairs and successive elimination to select x^*,

iii) finally, the bundle of goods x^* is consumed.

In this static consumption process there is no repeated actions (no succession of consumption acts of different bundle of goods). Then, the utility of consumption cannot change with past consumption (experience).

2.2 Repeated consumption and inertia: the impact of experience

Consider now a repeated action process of consumption ("the problem of repeated actions", Soubeyran, [25]). Start from consuming a given bundle of goods $x_0 \in X$. The problem of the consumer is: "should I stay", choosing to consume again x_0, or "should I change", moving from consuming x_0 to consume x_1. In this dynamic context, we will introduce inertia, ie costs to change (see Attouch-Soubeyran [2, 3], Soubeyran, [26, 27]). When passing from the old bundle of goods x to the new bundle y, the consumer can stop consuming some goods, continue to consume others, and start consuming new goods. This process of deletion, repetition, and inclusion generates various costs to change $C(x, y) \geq 0$. They include stopping costs (to decide to stop consuming some goods within the initial bundle x, which break temporary habits), ii) searching costs (to find new goods to include in the bundle of goods y), iii) costs to be able to continue to consume some goods, and finally iv) starting costs (to learn how to consume the new bundle of goods). Let $C_i(x_i, y_i) \in \mathbb{R}_+$ be the cost to change from consuming the quantity of good x_i to the quantity of good y_i. Then $C(x, y) = \sum_{i=1}^{m} C_i(x_i, y_i)$ is the cost to change from being able to consume the bundle of goods $x = (x_1, x_2, ..., x_m) \in X$ to be able to consume the new bundle of goods $y = (y_1, y_2, ..., y_m) \in X$. For simplification we will suppose that costs to be able to consume again a good are zero. $C_i(x_i, y_i) = 0$ if $y_i = x_i$. This means that an agent does not forget how to consume this good (see Dinh The-Soubeyran A, [10], for more on this point). As a consequence $C(x, x) = 0$ for all $x \in X$. Furthermore we will also assume that $C_i(x_i, y_i) = 0$ implies $y_i = x_i$. Then, $C(x, y) = 0 \Longleftrightarrow y = x$. "Costs to change" are different from traditional "costs to consume" $c(y) \geq 0$ which are included in the net utility, say $U(y) = u(y) - c(y)$, where $u(y)$ is the gross utility. Let $D(\cdot) : C \in \mathbb{R}_+ \longmapsto D(C) \in \mathbb{R}_+$ be the desutility of costs to change which is increasing and convex, and zero at zero. We can take $D(C) = C^2$.

Let $X^k = \left\{x^0, x^1, ..., x^k\right\} \subset X$ be the consumption path of the agent at stage k, ie his internal experience, and $\Omega^k = \left\{\omega^0, \omega^1, ..., \omega^k\right\} \subset \Omega$ be the history of the context (external events) until step k. Then, $E^k = (X^k, \Omega^k)$ will be his experience in consumption at this stage. In this context we define an "experience dependent" utility function $U(y/E^k)$. Reference dependent utility functions $U(y/x^k)$ are very special cases where the anchoring effect, the impact of experience on utility, is constant and limited to the most recent consumption bundle $E^k = x^k$ (see Kahnemam-Tversky, [12]).

In this paper we consider the simplest case of a separable experience dependent utility function, $U(y/E^k) = V(E^k)U(y)$, where the impact of experience $\lambda_k = V(E^k) > 0$ is multiplicative. At stage k, the consumer with experience E^k can balance between

i) consuming a new bundle of goods y, which gives him the proximal net utility $P(y/E^k) = U(y/E^k) - D\left[C(x^k, y)\right]$

ii) to consume again the bundle of goods x^k, which gives him the proximal net utility $P(x^k/E^k) = U(x^k/E^k) - D\left[C(x^k, x^k)\right] = U(x^k/E^k)$.

At stage k, the consumer will prefer to change from x^k to y if $U(y/E^k) - D\left[C(x^k, y)\right] \geq U(x^k/E^k)$.

His advantages to change being $A\left[(x^k, y)/E^k\right] = U(y/E^k) - U(x^k/E^k) = V(E^k)(U(y) - U(x^k))$, the consumer will prefer to change if his advantages to change are higher than the desutility of his costs to change $D\left[C(x^k, y)\right]$. This is equivalent to the acceptance criterion $U(y) - U(x^k) \geq (1/\lambda_k)D\left[C(x^k, y)\right]$.

At stage k the consumer will change in an optimal way if he can find a new consumption bundle x^{k+1} such that

$$P(x^{k+1}/E^k) = \sup\left\{P(y/E^k) = V(E^k)U(y) - D\left[C(x^k, y)\right], y \in X\right\} < +\infty.$$

Let $\lambda_k = V(E^k) > 0$ be the variable part of the utility function, which depends of past experience E^k. Then, at each stage, the optimization problem of the consumer is to solve the exact proximal problem:

$$\sup\left\{U(y) - (1/\lambda_k)D\left[C(x^k, y)\right], y \in X\right\}.$$

For the exact proximal algorithm $y = x^{k+1}$ is choosen in an optimal way.

Then, there is, each step, a global optimization problem to solve, which requires, each step, to search again over the whole space X. In this repeated but changing context the consumer can discover very soon, from his first experience, the invariant part of his utility function $\{U(y), y \in X\}$ over the whole space. But, each step, he is unable to known before consuming, what will be his future experience and the variable impact $V(E^k)$ of his experience E^k on the variable part of his utility function. The same is true for his present costs to change $C(x^k, y)$. He only knows the graphs of his past costs to change $\left\{C(x^h, y), y \in X\right\}$, $h = 0, 1, .., k-1$ because if the consumer knows $C(x, y)$ for all $y \in X$, he does not $C(y, z)$ for all $z \in X$.

To sum up, each step, the consumer must repeat exploration of the whole space to discover the new variable part of his utility function and his new costs to change. We must abandon the dichotomy principle which supposes that the consumer explores the whole space at the very beginning of the process and then stops to explore.

2.3 Habit's formation

Our problem will be to show in which circumstances the consumer can finally choose to consume the global optimal consumption bundle $x^* \in X$. We will show how habits form. The convergence result $x^k \longrightarrow x^* \in X$ implies that the distance between successive consumption bundles x^k, x^{k+1} goes to zero, ie $q(x^k, x^{k+1}) \longrightarrow 0, k \longrightarrow 0$. Then, the consumer repeats more and more the same consumption action. An habit forms gradually. In this case, if the consumer reachs x^*, given his experience $E^* = \left\{e^0, e^1, ...e^k = (x^k, \omega^k), ...\right\}$, he will prefer to stay than to move: $U(x^*/E^*) = V(E^*)U(x^*) \geq V(E^*)U(y)$ for all $y \in X$.

The condition $0 < \underline{\lambda} \leq \lambda_k = V(E^k) \leq \overline{\lambda}$ for all k, means that experience has a limited impact of the utility function. This does not require that the impact of experience $\lambda_k = V(E^k)$ converges. It can oscillate, but within bounds (within the bounded interval $\left[\underline{\lambda}, \overline{\lambda}\right]$).

Remark 2.1 *if the impact of experience converges to some limiting impact, ie* $\lambda_k = V(E^k) \longrightarrow \lambda^* = V^* = V(E^*) > 0$*, let* $P(y/E^*) = V^*U(y) - (D\left[C(x^*, y)\right]$ *be the limiting net utility. In this case, it is easy to show that if* $x^k \longrightarrow x^* \in X$*, then,* $P(x^*/E^*) \geq P(y/E^*)$ *for all* $y \in X$*. This means that, once at* x^**, the consumer prefers to stay than to move, taking care of costs to change. In this case* x^* *is an inertial habit. A global maximum requires a stronger condition,* $U(x^*) \geq U(y)$ *for all* $y \in X$*. In this case costs to change at* x^* *are not considered.*

Remark 2.2 *In our process, each step* k*, the consumer chooses to explore again the whole space of bundle of goods to find the global optimal bundle of good* x^{k+1}*. This is not realistic when, i) this space is too complex (includes a huge number of different bundles), ii) the length of each period* $k = 0, 1, .., n, ...$ *is short (lack of time), ...In these case, each step, exploration must be local, around the current consumption bundle. These "local action" aspects are left for future research, using an inexact proximal algorithm at each stage.*

2.4 Costs to change as the square of a quasi distance

Consider an agent who wants to pass from being able to consume a bundle of goods $x = (x_1, x_2, ..., x_m) \in X$ to be able to consume a new bundle of goods $y = (y_1, y_2, ..., y_m) \in X$.

If this agent wants to consume some added quantity of a good he must be able to consume it before consuming it. This means that to be able to consume an added quantity of a good i requires some added specific resource $i = i^+$ (and non specific resources, not modelized here). To be able to stop consuming a good requires to delete some specific resource $i = i^-$, which can be different from the resource $i = i^+$.

Then, if an agent wants to be able to consume a larger quantity of good $i = i^+$, say $v_i = y_i - x_i > 0$ units, he must hire (buy or built) a given added amount of a specific resource i, $K_i^+(v_i)$. Consider the linear case $K_i^+ = k_i^+(y_i - x_i)$ where $k_i^+ > 0$ is the quantity of resource i necessary to be able to consume one unit more of good i. Suppose that the cost to hire one unit of resource i, say $\rho_i^+ > 0$, is constant. Then, costs to change, which are costs to hire an added quantity of resource i, are $C_i(x_i, y_i) = \rho_i^+ K_i^+ = c_i^+(y_i - x_i)$ where $c_i^+ = \rho_i^+ k_i^+ > 0$. They are costs to be able to consume $v_i = y_i - x_i > 0$ units more of good i, starting from being able to consume the quantity x_i. They include search costs and learning costs....

If an agent wants to stop consuming some quantity of good i, say $v_i = x_i - y_i > 0$, he must be able to do it. He must break some habits. This requires to abandon the quantity $K_i^- = k_i^-(x_i - y_i)$ of resource $i = i^-$, where $k_i^- > 0$ is the quantity of resource $i = i^-$ to delete to be able to stop consuming one unit of good i. Suppose that the cost to stop consuming one unit of resource i, say $\rho_i^- > 0$, is constant. Then, costs to change which are costs to fire some quantity of resource i, are $C_i(x_i, y_i) = \rho_i^- K_i^- = c_i^-(x_i - y_i)$ where $c_i^- = \rho_i^- k_i^- > 0$. They are costs to be able to stop consuming $v_i = x_i - y_i > 0$ units of good i, starting from being able to consume the quantity x_i.

If the agent wants to consume again the same quantity of good i, say $y_i = x_i$, he mus tbe able to do that. We will suppose that it is the case. Then $C_i(x_i, y_i) = 0$ if $y_i = x_i$. To sum up, costs to change consumption of good i are, $C_i(x_i, y_i) = c_i^+(y_i - x_i)$, if $y_i - x_i > 0$; $C_i(x_i, y_i) = c_i^-(x_i - y_i)$ if $y_i - x_i < 0$ and $C_i(x_i, y_i) = 0$ if $y_i - x_i = 0$. Then,

$$C_i(x_i, y_i) = \begin{cases} c_i^+(y_i - x_i), & \text{if } y_i - x_i > 0, \\ c_i^-(x_i - y_i), & \text{if } y_i - x_i < 0, \\ 0, & \text{if } y_i - x_i = 0. \end{cases}$$

This is a quasi distance.

Then, $C(x, y) = \sum_{i=1}^m C_i(x_i, y_i)$ is the cost to change of consumption bundle, ie the costs to be able to consume the new bundle of goods $y = (y_1, y_2, ..., y_m) \in X$, starting from being able to consume the old bundle of goods $x = (x_1, x_2, ..., x_m) \in X$. It is as well a quasi distance.

The desutility of costs to change is $D(C) = C^2 = \left[\sum_{i=1}^m C_i(x_i, y_i)\right]^2$.

Remark 2.3 *costs to change can also depend of experience in a separable and bounded way, ie* $C(x, y/E) = \Gamma(E^k)Q(x, y)$, *where* $0 < \underline{\mu} \leq \Gamma(E^k) \leq \overline{\mu} < +\infty$ *for all* k. *In this case there is learning how to change in a less costly way.*

2.5 The litterature on habit's formation

Our model, like the famous "theory of habits, addiction and traditions" (Becker, 1991, Becker-Murphy,1988, Fuhrer, 2000) supposes that choices today are dependent of choices in the past. The current utility function depends on current consumption y and on the stock of past consumption. In our more general case the utility function depends on experience which is the history of past consumptions x^k and external influences ω^k, $E^k = \left\{e^0, e^1, ...e^k = (x^k, \omega^k)\right\}$. The other main differences of our model with the rational addiction model of Becker-Murphy (1988) are the following:

i) First, our model is adaptive. It does not use an intertemporal optimization process to maximize the sum of successive utilities. Using a cumulative objective and an initial optimization procedure made once for all, at the beginnig of the process (the dichtomy principle) is not allowed in our more realistic adaptive model where the variable part $V(E^k)$ of our utility function $U(y/E^k) = V(E^k)U(y)$ changes, each step, depending of the whole experience E^k. Then, the number of feasible experience path to imagine ex ante becomes astronomic as time evolves. Furthermore, this variable part of the utility function changes in a rather unpredicable way, within a given bounded interval of change $V(E^k) \in \left[\underline{\lambda}, \overline{\lambda}\right]$. Then, the consumer discovers each step, after consumption, his new utility function and his new costs to change. Being unable to anticipate his future utilities, he is obliged to adapt. He must repeatedely solve, each step, a new optimization program anchored to the past and revealed by his experience (we call this "the adaptive problem of repeated actions", Soubeyran A, 2009).

ii) Furthermore inertia is modelized in a more general way. Inertia comes from both, iia) a variable part in the utility function which depends of experience (past consumptions as in Becker (1991) and also external influences), iib) costs to change.

The implications of these different hypothesis, and the comparisons with the results of the Becker-Murphy (1988) model are left for future research, using different.formulations for the utility function and costs to change. Just say that in our model, habits and addiction can emerge from hypothesis different from adjacent complementarities (an increase in current consumption will increase future consumption), which is the cornerstone argument of Becker-Murphy (1988).

3 Preliminaries

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function. The effective domain of f is defined by $\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$, f is said to be coercive if $\lim_{\|x\| \to +\infty} f(x) = +\infty$, the symbol $B_r(x)$ denotes the open ball of radius ϵ about $\overline{x}$ and $\overline{B}_r(x)$ its clousure, f is said to be locally Lipschitz continuous around $\overline{x} \in \text{dom}(f)$ if there exist $\epsilon > 0$ and $\ell > 0$ such that $|f(x) - f(y)| \leq \ell\|x - y\| \quad \forall x, y \in B_\epsilon(\overline{x})$, where $B_\epsilon(\overline{x})$. Moreover, f is called locally Lipschitz continuous on the open subset $D \subset \mathbb{R}^n$ if f is locally Lipschitz around each $\overline{x} \in D$.

3.1 Quasi distance

Definition 3.1 *Let X be a set. A mapping $q : X \times X \to \mathbb{R}_+$ is called a quasi distance if*

1. *for all $x, y \in X$, $q(x, y) = q(y, x) = 0 \Leftrightarrow x = y$,*
2. *for all $x, y, z \in X$, $q(x, z) \leq q(x, y) + q(y, z)$.*

If q is also symmetric, that is, for all $x, y \in X$, $q(x, y) := q(y, x)$, then q is a distance. Furthermore, for each quasi distance q, we denote by $\overline{q}$ its conjugate quasi distance, where $\overline{q}(x, y) = q(y, x)$, and we call the distance $\hat{q}$, defined for each $x, y \in X$ by $\hat{q}(x, y) = \max\{q(x, y), \overline{q}(x, y)\}$, its associated distance. Next, we present examples of quase distances.

Example 3.1 *[9] Let $q : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ defined by*

$$q(x, y) = \begin{cases} 0 & \text{if } x \leq y, \\ 1 & \text{otherwise.} \end{cases}$$

generates a quasi distance on $\mathbb{R}$.

Example 3.2 *[20] If $q(x, y) := p(x - y)$ where p is an asymmetric norm.*

Example 3.3 *For each $i = 1, ..., n$, we consider $c_i^-, c_i^+ > 0$ and $q_i : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ defined by*

$$q_i(x_i, y_i) = \begin{cases} c_i^+(y_i - x_i) & \text{if } y_i - x_i > 0, \\ c_i^-(x_i - y_i) & \text{if } y_i - x_i \leq 0, \end{cases}$$

is a quasi distance on $\mathbb{R}$, therefore $q(x, y) = \sum_{i=1}^n q_i(x_i, y_i)$ is a quasi distance on $\mathbb{R}^n$. By another hand, for each $\overline{z} \in \mathbb{R}^n$ we have

$$q(x, \overline{z}) = \sum_{i=1}^n q_i(x_i, \overline{z}_i) = \sum_{i=1}^n \max\left\{c_i^+(\overline{z}_i - x_i), c_i^-(x_i - \overline{z}_i)\right\}, \quad x \in \mathbb{R}^n,$$

thus $q(\,\cdot\,, \overline{z})$ is a convex function. By the same reasoning, $q(\overline{z}, \,\cdot\,)$ is convex.

Example 3.4 *[16] Let $g : X \to \mathbb{R}_+$ be a injective function. Then for $x, y \in X$ we define $q_g(x, y) = \max\{g(x) - g(y), 0\}$, thus q_g generates a quasi distance on X.*

Example 3.5 *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a injective Lipzchitz function with constant $\mathcal{L}$ and $\mu > 0$. Then for $x, y \in X$ we define $q(x, y) = \max\{g(x) - g(y), \mu\|x - y\|\}$, thus q generates a quasi distance on $\mathbb{R}^n$.*

Remark 3.1 *It is not in general true that the quasi distance is convex and coercive in the first argument neither the second. In fact, Let* $q_g(\,\cdot\,,\overline{z}) = \max\{g(\,\cdot\,) - g(\overline{z}), 0\}$ *and* $q_g(\overline{z},\,\cdot\,) = \max\{g(\overline{z}) - g(\,\cdot\,), 0\}$*, where* q_g *is the quasi distance of Example 3.4, thus if* g *is a convex function, then* $q_g(\,\cdot\,,\overline{z})$ *is convex but* $q_g(\overline{z},\,\cdot\,)$ *is concave.*

In this paper, we consider the following condition on the quasi distance q:

There is a continuous function $\alpha : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{++}$ bounded away from zero and $\beta > 0$ such that

$$\alpha(x,y)\|x-y\| \leq q(x,y) \leq \beta\|x-y\|, \qquad x, y \in \mathbb{R}^n. \tag{3.1}$$

We note, that quasi distances of Examples 3.3 and 3.5 verifies (3.1). But the quasi distance of Example 3.4 does not verifies. The next results are important for our study.

Proposition 3.1 *Let* $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ *a quasi distance that verifies (3.1). Then for each* $\overline{z} \in \mathbb{R}^n$ *the functions* $q(\overline{z},\,\cdot\,)$ *and* $q(\,\cdot\,,\overline{z})$ *are Lipschitz.*

Proof: From the axiom (2), for all $x, y \in \mathbb{R}^n$ we obtain

$$|q(\overline{z},x) - q(\overline{z},y)| \leq \max\{q(x,y), q(y,x)\} \leq q(x,y) + q(y,x),$$

from (3.1), we have that exists $\mathcal{L} = 2\beta > 0$ such that

$$|q(\overline{z},x) - q(\overline{z},y)| \leq \mathcal{L}\|x-y\| \quad \forall\ x, y \in \mathbb{R}^n.$$

For the other function, the proof is analogous. ■

Proposition 3.2 *Let* $\overline{z} \in \mathbb{R}^n$*. If* q *verifies (3.1) then* $q^2(\overline{z},\,\cdot\,)$ *and* $q^2(\,\cdot\,,\overline{z})$ *are locally Lipschitz continuous functions on* $\mathbb{R}^n$*.*

Proof: Let $\overline{x} \in \operatorname{dom}(q^2(\overline{z},\,\cdot\,)) = \mathbb{R}^n$ and $\epsilon > 0$. We have that if $w \in B_\epsilon(\overline{x})$ then exists $K_{\overline{x}} > 0$ such that $|q(\overline{z},w)| \leq K_{\overline{x}}$, in fact

$$|q(\overline{z},w)| = |q(\overline{z},w) - q(\overline{z},\overline{z})| \leq \mathcal{L}\|w - \overline{x} + \overline{x} - \overline{z}\| \leq \mathcal{L}\left(\|w-\overline{x}\| + \|\overline{z}-\overline{x}\|\right)$$

and we consider $K_{\overline{x}} = \mathcal{L}\left(\epsilon + \|\overline{z}-\overline{x}\|\right)$. Thus for $x,\ y \in B_\epsilon(\overline{x})$ we have again from Proposition 3.1

$$|q^2(\overline{z},x) - q^2(\overline{z},y)| = |q(\overline{z},x) + q(\overline{z},y)|\,|q(\overline{z},x) - q(\overline{z},y)| \leq 2K_{\overline{x}}\mathcal{L}\|x-y\|,$$

because $\overline{x}$ is arbitrary point, $q^2(\overline{z},\,\cdot\,)$ is locally Lipschitz on $\mathbb{R}^n$. For the other function, the proof is analogous. ■

Remark 3.2 *We observe, if a quasi distance* q *verifies (3.1), then, for each* $\overline{z} \in \mathbb{R}^n$*, the functions* $q(\overline{z},\,\cdot\,)$*,* $q^2(\overline{z},\,\cdot\,)$*,* $q(\,\cdot\,,\overline{z})$ *and* $q^2(\,\cdot\,,\overline{z})$ *are coercives.*

Lemma 3.1 *Let $\overline{z} \in \mathbb{R}^n$ and $\lambda > 0$. If $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is bounded below and $q^2(\overline{z}, \cdot)$ is a coercive function, then the function $f + \frac{1}{\lambda} q^2(\overline{z}, \cdot) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is coercive.*

From Proposition 3.1 we have that $q(\overline{z}, \cdot)$ is a Lipschitz function when (3.1) holds, thus we have that $f + \frac{1}{\lambda} q^2(\overline{z}, \cdot)$ is a lower semicontinuous function. The next result shows when the minimizer set of a coercive function is nonempty, see for example [22, Theorem 1.9].

Theorem 3.1 *Suppose $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous, coercive and proper. Then the value $\inf f$ is finite and the set $\arg\min f$ is nonempty and compact.*

3.2 Subdifferential theory

Let us recall a few definitions concerning subdifferential calculus.

Definition 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function and $x \in \mathbb{R}^n$.*

1. *The Fréchet subdifferential of f at x, $\widehat{\partial} f(x)$, is defined as follows*

$$\widehat{\partial} f(x) := \begin{cases} \left\{ x^* \in \mathbb{R}^n : \liminf\limits_{\substack{y \neq x \\ y \to x}} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|x - y\|} \geq 0 \right\}, & \text{if } x \in \operatorname{dom}(f), \\ \emptyset, & \text{if } x \notin \operatorname{dom}(f); \end{cases}$$

2. *The limiting-subdifferential of f at $x \in \mathbb{R}^n$, $\partial f(x)$, is defined as follows*

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

3. *The remoteness of the subdifferential ∂f at $x \in \mathbb{R}^n$ as follows*

$$\|\partial f(x)\|_- = \inf_{p \in \partial f(x)} \|p\| = \operatorname{dist}(0, \partial f(x)).$$

Remark 3.3 *From the definition above implies that the set $\widehat{\partial} f(x)$ is closed and convex, $\partial f(x)$ is closed and $\widehat{\partial} f(x) \subset \partial f(x)$.*

Let write down the optimality condition, see for example [22, Theorem 10.1].

Theorem 3.2 *If a proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ has a local minimum at $\overline{x}$, then $0 \in \widehat{\partial} f(\overline{x})$, $0 \in \partial f(\overline{x})$.*

Unless f is convex the condition $0 \in \partial f(\overline{x})$ is not a sufficient for $\overline{x}$ be a local minimum point. In the remainder, a point $\overline{x} \in \mathbb{R}^n$ that satisfies its is called critical-limiting. The set of critical-limiting points of f is denoted by $\operatorname{crit}(f)$.

Definition 3.3 *A mapping $S : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$ is locally bounded at a point $\overline{x} \in \mathbb{R}^n$ if for some neighborhood $V \in \mathcal{N}(\overline{x})$ the set $S(V) \subset \mathbb{R}^m$ is bounded. It is called locally bounded on $\mathbb{R}^n$ if this holds at every $\overline{x} \in \mathbb{R}^n$.*

Proposition 3.3 *A mapping $S : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$ is locally bounded if and only if $S(B)$ is bounded for every bounded set B.*

Proof: See [22, Proposition 5.15]. ∎

The following subdifferential characterization of local Lipschitz continuity, see [22, Theorem 9.13], we will used in this work.

Theorem 3.3 *Suppose $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is locally lsc at $\overline{x}$ with $f(\overline{x})$ finite. Then the following conditions are equivalent:*

1. *f is locally Lipschitz continuous at $\overline{x}$,*
2. *the mapping $\widehat{\partial} f : x \mapsto \widehat{\partial} f(x)$ is locally bounded at $\overline{x}$,*
3. *the mapping $\partial f : x \mapsto \partial f(x)$ is locally bounded at $\overline{x}$.*

Moreover, when these conditions hold, $\partial f(\overline{x})$ is nonempty and compact.

Next, we present a formula for the limiting-subdifferential for a sum of functions, see for example [22, Exercise 10.10].

Theorem 3.4 *If f_1 locally Lipschitz continuous at $\overline{x}$, f_2 is lower semicontinuous and proper with $f_2(\overline{x})$ finite, then*

$$\partial\left(f_1 + f_2\right)(\overline{x}) \subset \partial f_1(\overline{x}) + \partial f_2(\overline{x}).$$

For the product of two functions, we have the following calculus rule, see for example [18, Theorem 7.1].

Theorem 3.5 *Let $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2$, be Lipschitz continuous around $\overline{x}$. If $f_i \geq 0$, $i = 1, 2$, then*

$$\partial(f_1.f_2)(\overline{x}) \subset f_2(\overline{x})\,\partial f_1(\overline{x}) + f_1(\overline{x})\,\partial f_2(\overline{x}). \tag{3.2}$$

We need the following version of the "fuzzy sum rule" for the Fréchet subdifferential applied to the finite dimensional case, see [18, Proposition 2.7].

Theorem 3.6 *Let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $i = 1, 2$, be lower semicontinuous functions, one of which is Lipschitz continuous around $\overline{x} \in \operatorname{dom}(f_1) \cap \operatorname{dom}(f_2)$. Then for any $\delta > 0$, and $\gamma > 0$ one has*

$$\widehat{\partial}\left(f_1 + f_2\right)(\overline{x}) \subset A + \gamma \overline{B}_1(0).$$

where

$$A = \bigcup \left\{ \widehat{\partial} f_1(x_1) + \widehat{\partial} f_2(x_2) \;:\; x_i \in B_\delta(\overline{x}),\, |f_i(x_i) - f_i(\overline{x})| \leq \delta,\, i = 1, 2 \right\}$$

We denote $[0 < f < \alpha] := \{x \in \mathbb{R}^n : 0 < f(x) < \alpha\}$ and $[0 < f \leq \alpha] := \{x \in \mathbb{R}^n : 0 < f(x) \leq \alpha\}$. The next result is a chain rule in the limiting-subdifferential context, see [7, Lema 43].

Lemma 3.2 *Let $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ be a proper lower semicontinuous function and $\varphi : (0, \alpha) \to \mathbb{R}$ be a C^1 function. Then*

$$\partial (\varphi \circ f)(x) = \varphi' (f(x)) \, \partial f(x), \qquad \forall\, x \in [0 < f < \alpha] .$$

3.3 Kurdyka-Łojasiewicz inequality

In this section, we present a generalization of the Kurdyka-Łojasiewicz property, on a limiting-subdifferential context. This characterization has been considered in [1, 4]. Given $\overline{r} \in (0, +\infty]$, we set

$$\mathcal{K}(0, \overline{r}) := \left\{ \phi \in C([0, \overline{r})) \cap C^1(0, \overline{r}) : \phi(0) = 0 \text{ and } \phi'(r) > 0, \forall r \in (0, \overline{r}) \right\}$$

where $C[0, \overline{r})$ (respectively, $C^1(0, \overline{r})$) denotes the set of continuous functions on $[0, \overline{r})$ (respectively, C^1 functions on $(0, \overline{r})$). Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function and $\overline{x} \in \{x \in \mathbb{R}^n : f(x) = 0\} \cap \mathrm{crit}(f)$. We assume

$$(\mathcal{H}_1) \quad \exists\, \overline{\epsilon}, \overline{r} > 0 \text{ such that it holds: } x \in \overline{B}_{\overline{\epsilon}}(\overline{x}) \cap [0 < f \leq \overline{r}] \;\Rightarrow\; 0 \notin \partial f(x)$$

This condition can be seen as a Sard-type condition, i.e. 0 is a locally upper isolated critical value.

We consider the following characterization of the **Kurdyka-Łojasiewicz inequality** in terms of the limiting-subdifferential:

There exist $r_0 \in (0, \overline{r})$, $\epsilon \in (0, \overline{\epsilon})$ and a concave function $\varphi \in \mathcal{K}(0, r_0)$ such that

$$\|\partial(\varphi \circ f)(x)\|_{-} \geq 1, \qquad \forall\, x \in \overline{B}_{\epsilon}(\overline{x}) \cap [0 < f \leq r_0] . \tag{3.3}$$

Remark 3.4 *As a particular case of (3.3) we have the Łojasiewicz inequality. In fact consider the concave function $\varphi(s) := s^{1-\theta} \in \mathcal{K}(0, r_0)$ where $\theta \in [\frac{1}{2}, 1)$ and the restriction of f to its domain is a continuous function. We have that there exist $r_0 \in (0, \overline{r})$, $\epsilon \in (0, \overline{\epsilon})$ such that (3.3) is verified and by Lemma 3.2, we obtain*

$$1 \leq \|\varphi' (f(x))\, \xi\| = |\varphi' (f(x))| \, \|\xi\| \qquad \forall \xi \in \partial f(x),\ \forall\, x \in \overline{B}_{\epsilon}(\overline{x}) \cap [0 < f < r_0] ,$$

and because $1 \leq \left|(1-\theta) f(x)^{-\theta}\right| \|\xi\| \leq \left|f(x)^{-\theta}\right| \|\xi\|$ and $f(\overline{x}) = 0$ we have

$$|f(x)|^{\theta} = |f(x) - f(\overline{x})|^{\theta} \leq \|\xi\| \qquad \forall \xi \in \partial f(x),\ \forall\, x \in \overline{B}_{\epsilon}(\overline{x}) \cap [0 < f < r_0]$$

This last characterization has been considered in [4].

Example 3.6 *[19] Real-analytic functions and Semialgebraic functions verifies the Kurdyka-Łojasiewicz inequality.*

Example 3.7 *[4] Convex functions with the following growth condition:* $\forall\, \widehat{x} \in \arg\min(f)$, $\exists\, C > 0$, $r \geq 1$, $\epsilon > 0$ *such that*

$$f(x) \geq f(\widehat{x}) + C\,\mathrm{dist}\,(x, \arg\min(f))^r\,, \qquad x \in \overline{B}_\epsilon(\widehat{x}),$$

verify the Kurdyka-Lojasiewicz inequality, in particular the strongly convex functions.

For further imformation about the Kurdyka-Łojasiewicz property and characterizations see [19, 1, 7].

4 Proximal algorithm with quasi distance

The following hypotheses will be assumed in our study:

$(\mathcal{H}_1)$ $-\infty < \inf\limits_{x\in\mathbb{R}^n} f(x)$;

$(\mathcal{H}_2)$ q verifies the condition (3.1);

$(\mathcal{H}_3)$ The restriction of f to its domain is a continuous function;

$(\mathcal{H}_4)$ for each $\overline{x} \in \mathrm{crit}(f) \cap \{x \in \mathbb{R}^n \,:\, f(x) = 0\}$, $\exists\, \overline{\epsilon}, \overline{r} > 0$ such that it holds:

$$x \in \overline{B}_{\overline{\epsilon}}(\overline{x}) \cap [0 < f \leq \overline{r}] \;\Rightarrow\; 0 \notin \partial f(x).$$

The method generates a sequence $\{x^k\}_{k\in\mathbb{N}}$ in the following way:

1. Take $x^0 \in \mathrm{dom}(f)$.
2. Given x^k, find x^{k+1} such that

$$x^{k+1} \in \arg\min\left\{ f(u) + \frac{1}{2\lambda_k} q^2(x^k, u) \,:\, u \in \mathbb{R}^n \right\}, \tag{4.1}$$

   where $\{\lambda_k\}_{k\in\mathbb{N}}$ is positive sequence such that $\lambda_k \in (\lambda_-, \lambda_+)$.
3. If $x^{k+1} = x^k$, **STOP**.

The following result gathers a few elementary facts concerning (4.1).

Proposition 4.1 *If* $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$ *are verifies, then sequence* $\{x^k\}$ *is well defined.*

Proof: It consequence From Theorem 3.1. ∎

Proposition 4.2 *Let* $\{x^k\}_{k\in\mathbb{N}}$ *be a sequence which complies with (4.1). Then* $\{f(x^k)\}_{k\in\mathbb{N}}$ *is nonincreasing. Moreover* $\{f(x^k)\}_{k\in\mathbb{N}}$ *is bounded.*

Proof: By definition, (4.1) implies that for all $k \geq 0$ we have

$$f(x^{k+1}) + \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^k), \quad \forall\, k, \tag{4.2}$$

thus $f(x^{k+1}) \leq f(x^k)\ \forall\, k$. By another hand,

$$-\infty < \inf_{x \in \mathrm{R}^n} f(x) \leq f(x^{k+1}) \leq f(x^0) < +\infty \qquad \forall\, k$$

therefore $\{f(x^k)\}_{k\in\mathrm{N}}$ is bounded. ■

Proposition 4.3 *We assume $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ are verifies. If $\{x^k\}_{k\in\mathrm{N}}$ be a sequence which complies with (4.1) then $\sum_{k=0}^{+\infty} q^2(x^k, x^{k+1}) < +\infty$, in particular*

$$\lim_{k\to+\infty} q^2(x^k, x^{k+1}) = 0.$$

Proof: From (4.2) and $\lambda_k \in (\lambda_-, \lambda_+)$, we have

$$f(x^{k+1}) + \frac{1}{2\lambda_+} q^2(x^k, x^{k+1}) \leq f(x^k) \qquad \forall\, k, \tag{4.3}$$

and by summing the inequalities (4.3) from 0 to $n \geq 0$ we obtain that

$$\sum_{k=0}^{n} q^2(x^k, x^{k+1}) \leq 2\lambda_+ \Big[f(x^0) - f(x^{n+1})\Big] \leq 2\lambda_+ \Big[f(x^0) - \inf_{\mathrm{R}^n} f\Big] < \infty.$$

From the last result we have that $\lim_{k\to+\infty} q^2(x^k, x^{k+1}) = 0$. ■

Now we give a characterization of (4.1) on term of limiting-subdifferential.

Proposition 4.4 *If $(\mathcal{H}_1)$ and $(\mathcal{H}_2)$ are verifies, then there exists $\xi^{k+1} \in \partial f(x^{k+1})$ and $\zeta^{k+1} \in \partial \Big(q(x^k, \cdot)\Big)(x^{k+1})$ such that*

$$0 = \xi^{k+1} + \frac{q(x^k, x^{k+1})}{\lambda_k} \zeta^{k+1}. \tag{4.4}$$

Proof: From Proposition 3.1, in particular for $\overline{z} = x^k$ we have that $q^2(x^k, \cdot)$ is locally Lipschitz at x^{k+1}. Regarding the algorithm, we can use the result of Theorem 3.2 with f substituted by $f_1 + f_2$ where $f_1 := \frac{1}{2\lambda_k} q^2(x^k, \cdot)$ and $f_2 := f$. Now apply Theorem 3.4, going

$$0 \in \partial\left(f(\cdot) + \frac{1}{2\lambda_k} q^2(x^k, \cdot)\right)(x^{k+1}) \subset \partial f(x^{k+1}) + \frac{1}{2\lambda_k} \partial\left(q^2(x^k, \cdot)\right)(x^{k+1}),$$

moreover, from Theorem 3.5 applied to $\varphi_1 = \varphi_2 = q(x^k, \cdot)$ we obtain

$$\partial\left(q^2(x^k, \cdot)\right)(x^{k+1}) \subset 2q(x^k, x^{k+1}) \partial\left(q(x^k, \cdot)\right)(x^{k+1}),$$

thus

$$0 \in \partial f(x^{k+1}) + \frac{1}{\lambda_k} q(x^k, x^{k+1}) \partial\left(q(x^k, \cdot)\right)(x^{k+1})$$

therefore there exists $\xi^{k+1} \in \partial f(x^{k+1})$ and $\zeta^{k+1} \in \partial\left(q(x^k, \cdot)\right)(x^{k+1})$ such that (4.4) is verified. ■

Proposition 4.5 *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence which complies with (4.1) and denote by $\omega(x^0)$ the set of its limit points. If f satisfies $(\mathcal{H}_3)$ then*

1. *f is finite and constant on $\omega(x^0)$,*
2. *If $\{x^k\}_{k\in\mathbb{N}}$ is bounded then $\omega(x^0)$ is a nonempty compact connected set, and*

$$\lim_{k\to+\infty} \operatorname{dist}\left(x^k, \omega(x^0)\right) = 0. \tag{4.5}$$

Proof: (1) The sequence $\{f(x^k)\}_{k\in\mathbb{N}}$ is nonincreasing and bounded from below, then it converges to a point β. Now, let $\overline{x}$ any cluster point of $\{x^k\}_{k\in\mathbb{N}}$, then there exists $\{x^{k_j}\}_{j\in\mathbb{N}}$ such that $\lim_{j\to+\infty} x^{k_j} = \overline{x}$. From continuity of f we obtain $f(\overline{x}) = \lim_{j\to+\infty} f(x^{k_j}) = \beta$. As $\overline{x}$ is arbitrary the prove is concluded.

(2) The connenteness and compactness are trivial, we only prove (refdistlim). By contradiction suppose that there exists $\beta > 0$ and $n_0 \in \mathbb{N}$ such that:

$$\operatorname{dist}\left(x^k, \omega(x^0)\right) \geq 2\beta > 0, \quad \forall\, k \geq n_0.$$

As $\{x^k\}_{k\in\mathbb{N}}$ is bounded, then there exists $\overline{x} \in \mathbb{R}^n$ and a subsequence, denoted by $\{x^{k_j}\}_{j\in\mathbb{N}}$, such that $\lim_{j\to+\infty} x^{k_j} = \overline{x}$, so $\lim_{j\to+\infty} \operatorname{dist}(x^{k_j}, \overline{x}) = 0$. Therefore $\lim_{j\to+\infty} \operatorname{dist}\left(x^{k_j}, \omega(x^0)\right) = 0$. From the definition of convergent subsequence, there exists $n_1 := n_1(\beta) \in \mathbb{N}$ such that for all $j \geq n_1$ we have $\operatorname{dist}\left(x^{k_j}, \omega(x^0)\right) < \beta$. Now, considere $j \geq n_1$, such that $k_j \geq n_0$ (such j exists because $\{x^{k_j}\}_{j\in\mathbb{N}}$ is a subsequence of $\{x^k\}_{k\in\mathbb{N}}$) then $2\beta \leq \operatorname{dist}\left(x^{k_j}, \omega(x^0)\right) < \beta$, which is a contradiction. Therefore we obtain the aimed result. ■

Remark 4.1 *Even when $\{x^k\}_{k\in\mathbb{N}}$ is bounded, the convergence of the whole sequence $\{x^k\}_{k\in\mathbb{N}}$ may fail even for a finite-valued smooth function f, see [1].*

5 Convergence Result

Lemma 5.1 *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence generated by the proximal algorithm. If $\{x^k\}_{k\in\mathbb{N}}$ is bounded, then the set $\partial\left(q(x^k,\,\cdot\,)\right)(x^{k+1})$ is bounded for all k.*

Proof: We apply $f := q(\overline{z},\,\cdot\,)$ in Theorem 3.3, obtaining that $\partial(q(\overline{z},\,\cdot\,))(\overline{x})$ is locally bounded $\forall\,\overline{x} \in \operatorname{dom}(q(\overline{z},\,\cdot\,))$ and $\overline{z} \in \mathbb{R}^n$. Then by using Proposition 3.3 with $S = \partial(q(\overline{z},\,\cdot\,))$ and $B = \{x^k\}_{k\in\mathbb{N}}$

$$\partial(q(\overline{z},\,\cdot\,))(B) := \bigcup_{x\in B} \partial(q(\overline{z},\,\cdot\,))(x)$$

is bounded for each $\overline{z} \in \mathbb{R}^n$ fixed, thus if $\overline{z} \in B$, $\exists\, M > 0$:

$$\|y\| \leq M \quad \forall\, y \in \bigcup_{\overline{z}, x\in B} \partial(q(\overline{z},\,\cdot\,))(x)$$

in particular, $\partial\left(q(x^k,\,\cdot\,)\right)(x^{k+1})$ is bounded by M. ■

Proposition 5.1 *If $(\mathcal{H}_3)$ is verified, then any cluster point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is a limiting-critical point of f*

Proof: Suppose that there exists $\overline{x}$ and a subsequence $\{x^{k_j}\}_{j\in\mathbb{N}}$ of $\{x^k\}_{k\in\mathbb{N}}$ converging to $\overline{x}$. We will prove that $0 \in \partial f(\overline{x})$. From (4.1) and Theorem 3.2 we have

$$0 \in \widehat{\partial}\left(f(\cdot) + \frac{1}{2\lambda_{k_j-1}} q^2(x^{k_j-1}, \cdot)\right)(x^{k_j}).$$

Applying Theorem 3.6, with $f_1 := f$, $f_2 := \frac{1}{2\lambda_{k_j-1}} q^2(x^{k_j-1}, \cdot)$, $\overline{x} = x^{k_j}$ and $\delta = \gamma = \frac{1}{k_j}$ we obtain $0 \in \widehat{\partial}(f_1 + f_2)(x^{k_j}) \subset A + \frac{1}{k_j}\overline{B}_1(0)$ where

$$A = \left\{\widehat{\partial} f_1(x_1^{k_j}) + \widehat{\partial} f_2(x_2^{k_j}) \;:\; x_i^{k_j} \in \overline{B}_{\frac{1}{k_j}}(x^{k_j}),\, |f_i(x_i^{k_j}) - f_i(x^{k_j})| \le \frac{1}{k_j},\, i = 1, 2\right\}$$

Thus there exists $x_1^{k_j}$, $x_2^{k_j} \in \overline{B}_{\frac{1}{k_j}}(x^{k_j})$, $\xi^{k_j} \in \widehat{\partial} f(x_1^{k_j})$, $\zeta^{k_j} \in \widehat{\partial}\left(q^2(x^{k_j-1}, \cdot)\right)(x_2^{k_j})$ and $u^{k_j} \in \overline{B}_1(0)$ such that

$$0 = \xi^{k_j} + \frac{1}{2\lambda_{k_j-1}}\zeta^{k_j} + \frac{1}{k_j}u^{k_j}. \tag{5.1}$$

Notice that

$$\begin{aligned}
0 \le q(x^{k_j-1}, x_2^{k_j}) &\le q(x^{k_j-1}, x^{k_j}) + q(x^{k_j}, x_2^{k_j}) \\
&= q(x^{k_j-1}, x^{k_j}) + q(x^{k_j}, x_2^{k_j}) - q(x^{k_j}, x^{k_j}) \\
&\le \mathcal{L}\|x^{k_j-1} - x^{k_j}\| + \mathcal{L}\|x_2^{k_j} - x^{k_j}\|
\end{aligned}$$

therefore

$$\lim_{j\to+\infty} q(x^{k_j-1}, x_2^{k_j}) = 0.$$

Now, as $\widehat{\partial}\left(q^2(x^{k_j-1}, \cdot)\right)(x_2^{k_j}) \subset \partial\left(q^2(x^{k_j-1}, \cdot)\right)(x_2^{k_j})$ we have

$$\zeta^{k_j} \in \partial\left(q^2(x^{k_j-1}, \cdot)\right)(x_2^{k_j}) \subset 2q(x^{k_j-1}, x_2^{k_j})\partial\left(q(x^{k_j-1}, \cdot)\right)(x_2^{k_j})$$

and using the boundenes of the last set, we have from (5.1)

$$\|\xi^{k_j}\| \le \frac{1}{2\lambda_{k_j-1}}\|\zeta^{k_j}\| + \frac{1}{k_j} \le \frac{1}{2\lambda_-}\|\zeta^{k_j}\| + \frac{1}{k_j} \le \frac{1}{\lambda_-} q(x_2^{k_j}, x^{k_j-1})M + \frac{1}{k_j}.$$

Therefore, there are sequences $\{x_1^{k_j}\}_{j\in\mathbb{N}}$, $\{f(x_1^{k_j})\}_{j\in\mathbb{N}}$ and $\{\xi^{k_j}\}_{j\in\mathbb{N}}$ with $\xi^{k_j} \in \widehat{\partial} f(x_1^{k_j})$ such that $\lim_{j\to+\infty} x_1^{k_j} = \overline{x}$, $\lim_{j\to+\infty} f(x_1^{k_j}) = f(\overline{x})$ and $\lim_{j\to+\infty} \xi^{k_j} = 0$, so $0 \in \partial f(\overline{x})$. ∎

Next, we state and prove our main convergence result, the proof that we develop here, is motivated by [4].

Theorem 5.1 *Assume that f satisfies $(\mathcal{H}_1)$-$(\mathcal{H}_4)$, and verifies the Kurdyka-Lojasiewicz inequality (3.3). Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence generated by the proximal algorithm. If the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded, then*

$$\sum_{k=0}^{+\infty} q(x^k, x^{k+1}) < +\infty, \tag{5.2}$$

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

Proof: Without loss of generality, we can assume that $f(x) := f(x) - \inf_{k\geq 0} f(x^k)$, thus we have that $\lim_{k\to+\infty} f(x^k) = 0$.

The case when $x^{k+1} = x^k$ for some $k \geq 1$ has no consequence on the asymptotic analysis, so that we may suppose that $q(x^k, x^{k+1}) > 0$ for all $k \geq 0$. In view of (4.2), we obtain also that $f(x^k)$ is positive and decreases (strictly) to 0. Thus for $\overline{x} \in \operatorname{crit}(f)$, we have $f(\overline{x}) = 0$ and from $(\mathcal{H}_4)$ exists $\epsilon, \overline{r} > 0$.

From the definition of the set $\mathcal{K}(0, \overline{r})$ the function φ is strictly concave on $(0, \overline{r})$. Then by using the differentiable convex inequality for the function $-\varphi(s)$ for $s, t \in (0, \overline{r})$, where $s := f(x^{k+1})$ and $t := f(x^k)$, obtaining

$$\begin{aligned} \varphi\left(f(x^k)\right) - \varphi\left(f(x^{k+1})\right) &\geq \varphi'\left(f(x^k)\right)\left(f(x^k) - f(x^{k+1})\right) \\ &\geq \varphi'\left(f(x^k)\right)\frac{1}{\lambda_k} q^2(x^k, x^{k+1}), \end{aligned} \tag{5.3}$$

where the last inequality comes from (4.2) and because $\varphi'\left(f(x^k)\right) > 0$. By another hand, from the Kurdyka-Lojasiewicz inequality we have that exist $r_0 \in (0, \overline{r})$ and $\epsilon \in (0, \overline{\epsilon})$ such that (3.3) is verified, and from Lemma 3.2 we have for $\alpha = r_0$ that $\varphi'(f(x))\,\partial f(x) \subset \partial(\varphi \circ f)(x)$ then

$$\operatorname{dist}\left(0, \varphi'(f(x))\,\partial f(x)\right) \geq 1, \qquad \forall\, x \in \overline{B}_\epsilon(\overline{x}) \cap [0 < f < r_0],$$

which implies $\|p\| \geq 1$ $\forall p \in \varphi'(f(x))\,\partial f(x)$, $\forall\, x \in \overline{B}_\epsilon(\overline{x}) \cap [0 < f < r_0]$. Thus we obtain for $x^k \in \overline{B}_\epsilon(\overline{x}) \cap [0 < f < r_0]$ and $\xi^k \in \partial f(x^k)$: $1 \leq \|\varphi'(f(x^k))\xi^k\| = \varphi'(f(x^k))\|\xi^k\|$, then

$$\frac{1}{\|\xi^k\|} \leq \varphi'(f(x^k)) \qquad \forall\, x^k \in \overline{B}_\epsilon(\overline{x}) \cap [0 < f < r_0]\ , \forall\, \xi^k \in \partial f(x^k). \tag{5.4}$$

By another hand, from (4.4) we have for all $k \geq 0$

$$\|\xi^k\| = \frac{q(x^{k-1}, x^k)}{\lambda_{k-1}}\|\zeta^k\|$$

where $\zeta^k \in \partial\left(q(x^{k-1}, \cdot)\right)(x^k)$ and because $\lambda_k \in (\lambda_-, \lambda_+)$ we get

$$\frac{\lambda_-}{q(x^{k-1}, x^k)\|\zeta^k\|} \leq \frac{1}{\|\xi^k\|} \qquad \forall\, k \geq 0. \tag{5.5}$$

Now from (5.4) and (5.5) we obtain that there exists $n_0 = n_0(\epsilon) \in \mathbb{N}$ such that both inequalities are verified for all $k \geq n_0$. Then

$$\frac{\lambda_-}{q(x^{k-1}, x^k)\|\zeta^k\|} \leq \varphi'\left(f(x^k)\right) \qquad \forall\, k \geq n_0 \tag{5.6}$$

where $\zeta^k \in \partial\left(q(x^{k-1}, \cdot)\right)(x^k)$. Combining (5.6) with (5.3) (recall that $\lambda_+ > \lambda_k > 0$) we have

$$\frac{q^2(x^k, x^{k+1})}{q(x^{k-1}, x^k)\|\zeta^k\|} \leq \frac{\lambda_+}{\lambda_-}\left(\varphi\left(f(x^k)\right) - \varphi\left(f(x^{k+1})\right)\right) \quad \forall\, k \geq n_0. \tag{5.7}$$

Fix $r \in (0,1)$ and take $k \geq n_0$. We are going to consider two cases $q(x^k, x^{k+1}) \geq rq(x^{k-1}, x^k)$ or $q(x^k, x^{k+1}) < rq(x^{k-1}, x^k)$. Let $q(x^k, x^{k+1}) \geq rq(x^{k-1}, x^k)$, by Lemma 5.1, we know that $\partial\left(q(x^{k-1}, \cdot)\right)(x^k)$ is bounded by some M, thus $\|\zeta^k\| \leq M\ \forall\, k$, it follows though the starting hypothesis:

$$\frac{r}{M} q(x^k, x^{k+1}) \leq \frac{q^2(x^k, x^{k+1})}{q(x^{k-1}, x^k)M} \leq \frac{q^2(x^k, x^{k+1})}{q(x^{k-1}, x^k)\|\zeta^k\|}.$$

With (5.7) we achieve:

$$q(x^k, x^{k+1}) \leq \frac{M\lambda_+}{r\lambda_-}\left(\varphi\left(f(x^k)\right) - \varphi\left(f(x^{k+1})\right)\right).$$

Now, let $q(x^k, x^{k+1}) < rq(x^{k-1}, x^k)$. Then

$$q(x^k, x^{k+1}) \leq rq(x^{k-1}, x^k) + \frac{M\lambda_+}{r\lambda_-}\left(\varphi\left(f(x^k)\right) - \varphi\left(f(x^{k+1})\right)\right) \quad \forall\, k \geq n_0. \tag{5.8}$$

Summing over $k = n_0, \cdots, n$, and set $L = \frac{M\lambda_+}{r\lambda_-}$ we obtain

$$\begin{aligned}
\sum_{k=n_0}^{n} q(x^k, x^{k+1}) \quad &\leq \quad r\sum_{k=n_0}^{n-1} q(x^k, x^{k+1}) + rq(x^{n_0-1}, x^{n_0}) \\
&\qquad + \frac{M\lambda_+}{r\lambda_-}\left(\varphi\left(f(x^{n_0})\right) - \varphi\left(f(x^{n+1})\right)\right) \\
&\leq \quad r\sum_{k=n_0}^{n} q(x^k, x^{k+1}) + rq(x^{n_0-1}, x^{n_0}) \\
&\qquad + \frac{M\lambda_+}{r\lambda_-}\left(\varphi\left(f(x^{n_0})\right) - \varphi\left(f(x^{n+1})\right)\right)
\end{aligned}$$

Thus

$$(1-r)\sum_{k=n_0}^{n} q(x^k, x^{k+1}) \leq rq(x^{n_0-1}, x^{n_0}) + \frac{M\lambda_+}{r\lambda_-}\left(\varphi\left(f(x^{n_0})\right) - \varphi\left(f(x^{n+1})\right)\right).$$

or equivalentely

$$\sum_{k=n_0}^{n} q(x^k, x^{k+1}) \leq \frac{r}{1-r} q(x^{n_0-1}, x^{n_0}) + \frac{M\lambda_+}{r(1-r)\lambda_-}\left(\varphi\left(f(x^{n_0})\right) - \varphi\left(f(x^{n+1})\right)\right).$$

Since f is bounded from below and φ is strictly increasing we conclude (5.2). ∎

6 Conclusions

We analyze the proximal point algorithm associated to quasi distances for minimizing nondifferentiable functions that verify the Kurdyka-Łojasiewicz inequality. Apart from the convergence results, we show a decision model that relates proximal methods and quasi distances, as an interesting motivation for that setting.

References

[1] P.A. ABSIL, R. MAHONY and B. ANDREWS, *Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim. 16 (2005), pp. 531–547.

[2] H. ATTOUCH and A. SOUBEYRAN, *Inertia and reactivity in decision making as cognitive variational inequalities*, J. Convex Analysis 13 (2) (2006), pp. 207–224.

[3] H. ATTOUCH and A. SOUBEYRAN, *Worthwhile-to-move behaviors as temporary satisficing without too many sacrificing processes*, preprint (2008).

[4] H. ATTOUCH and J. BOLTE, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., Ser. B 116 (2009), pp. 5–16.

[5] G. BECKER, Habits, addictions and traditions, *Nancy Schwartz Lecture*, Northwestern University (1991).

[6] G. BECKER and K. MURPHY, *A theory of rational addiction*, Journal of Political Economy 96 (4) (1988), pp. 675–700.

[7] J. BOLTE, A. DANIILIDIS, O. LEY and L. MAZET, *Characterizations of Łojasiewicz inequalities and applications*, Preprint (2009), pp. 1–47.

[8] V. BRATTKA, *Recursive quasi-metric spaces*, Theoretical Computer Science 305 (2003), pp. 17–42.

[9] A. DI CONCILIO and G. GERLA, *Quasi-metric spaces and point-free geometry*, Math. Struct. in Comp. Science 16 (2006), pp. 115–137.

[10] L. DINH-THE and A. SOUBEYRAN, *Evolutionary preferences and the muddling through problem*, Preprint, Avignon (2009).

[11] J. FUHRER, *Habit formation in consumption and its implications for monetary-policy models*, The American Economic Review 90(3) (2000), pp. 367–390.

[12] D. KAHNEMAN and A. TVERSKY, *Prospect theory: an analysis of decision under risk*, Econometrica 47 (1979), pp. 313–327.

[13] A. KAPLAN and R. TICHATSCHKE, *Proximal point methods and nonconvex optimization*, J. Global Optim. 13(4) (1998), pp. 389–406.

[14] K.C. KIWIEL, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optimization 35(4) (1997), pp. 1142–1168.

[15] H.-P.A. KÜNZI, *Nonsymmetric Topology*, Bolyai Society in Mathematical Studies 4, Topology with Applications (1993), pp. 303–338.

[16] H.-P.A. KÜNZI, H. PAJOOHESH and M.P. SCHELLEKENS, *Partial quasi-metrics*, Theoretical Computer Science 365(2006), pp. 237–246.

[17] B. MARTINET, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle 4 (1970), pp. 154–158.

[18] B.S. MORDUKHOVICH and Y. SHAO, *Nonsmooth Sequential Analysis in Asplund Spaces*, Transactions of the American Mathematical Society 348(4) (1996), pp. 1235–1280.

[19] S. ŁOJASIEWICZ, *Ensembles Semi-Analytiques*, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette (Seine-et-Oise) France, (1965).

[20] L.M. GARCIA-RAFFI, *Compactness and finite dimension in asymmetric normed linear spaces*, Topology and its Applications 153 (2005), pp. 844–853.

[21] R.T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optimization 14(5) (1976), pp. 877–898.

[22] R.T. ROCKAFELLAR and R.J-B. WETS, *Variational analysis*, Springer, Berlin, 1998.

[23] S. ROMAGUERA and M. SANCHIS, *Applications of utility functions defined on quasi-metric spaces*, J. Math. Anal. Appl. 283(1) (2003), pp. 219–235.

[24] H. SIMON, From substantive to procedural rationality, *Methods and Appraisal in Economics*, (1967), Latsis, editor.

[25] A. SOUBEYRAN, *Satisficing processes by rejection*, Preprint, Aix Marseille (2007).

[26] A. SOUBEYRAN, *Individual variational rationality: improving by combining worthwhile repetitive and variational activities*, Preprint (2009).

[27] A. SOUBEYRAN, *Gradual satisficing and individual variational rationality*, Preprint (2009).

[28] A. STOJMIROVIĆ, *Quasi-metric Spaces with Measure*, Topology Proceedings 28(2)(2004), pp. 655–671.