Dual Descent Methods as Tension Reduction Systems

Glaydston de Carvalho Bento, João Xavier da Cruz Neto, Antoine Soubeyran, Valdinês Leite de Sousa Júnior

Dual Descent Methods as Tension Reduction Systems

Glaydston de Carvalho Bento[1] · **João Xavier da Cruz Neto**[2] · **Antoine Soubeyran**[3] · **Valdinês Leite de Sousa Júnior**[1]

Abstract In this paper, driven by applications in Behavioral Sciences, wherein the speed of convergence matters considerably, we compare the speed of convergence of two descent methods for functions that satisfy the well-known Kurdyka–Lojasiewicz property in a quasi-metric space. This includes the extensions to a quasi-metric space of both the primal and dual descent methods. While the primal descent method requires the current step to be more or less half of the size of the previous step, the dual approach considers more or less half of the previous decrease in the objective function to be minimized. We provide applications to the famous "Tension systems approach" in Psychology.

Keywords Dual descent · Inexact proximal · Worthwhile change · Kurdyka–Lojasiewicz property · Tension systems · Variational rationality

Mathematics Subject Classification 90C30 · 49M29

Communicated by Boris S. Mordukhovich.

✉ Valdinês Leite de Sousa Júnior
valdinesldsj@gmail.com

Glaydston de Carvalho Bento
glaydstonc@gmail.com

João Xavier da Cruz Neto
jxavier@ufpi.edu.br

Antoine Soubeyran
antoine.soubeyran@gmail.com

1 IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil

2 DM, Universidade Federal do Piauí, Teresina, PI 64049-550, Brazil

3 Aix-Marseille School of Economics, CNRS and EHESS, Aix-Marseille University, Marseille, France

1 Introduction

Driven by applications in Behavioral Sciences, wherein the speed of convergence matters considerably, in this study we compare the speed of convergence of two descent methods for functions that satisfy the well-known Kurdyka–Lojasiewicz property in a quasi-metric space. This includes the extension to a quasi-metric space of the primal descent method in Attouch and Bolte [1] and that of the dual descent method given by Alaa and Pierre [2]. The two main conditions (1) and (2) given in this paper are weaker than conditions H1 and H2 considered in Attouch et al. [3]. Our approach also retrieves the exact proximal point method presented by Moreno et al. [4], as well as the inexact proximal point method studied by Fuentes et al. [5].

Dual descent methods consider two types of steps. In [1], the current step must be more or less half of the size of the previous step. In contrast, the dual approach in [2] considers more or less half of the previous decrease in the objective function to be minimized. Before comparing these dual algorithms, we need to extend them to quasi-metric spaces. This is because in Behavioral Sciences (see Soubeyran [6–8] for the introduction of quasi-distances in such areas), distances cannot represent specific instances of costs of being able to change from one position to another. They are, most of the time, quasi-distances, where the costs of being able to change from one position to another are different from those of being able to change to go back to the first position. We find that in a large set of cases, dual resolution converges faster than primal resolution. This is caused by behavioral considerations, which imply the different ways in which an agent can choose to reduce the discrepancy between where he is initially (away from an optimal position) and where he wants to be in the near or distal future (the goal, at an optimal position if possible). The primal resolution chooses to reduce the discrepancy, defined as the quasi-distance between where the agent is in the current period and where he wants to be (an optimal position) in the near or distal future. The dual resolution chooses to reduce the discrepancy, defined as the difference between his current dissatisfaction (a disutility, a negative feeling) and his minimal dissatisfaction (tension-free state).

Let us be more explicit on the motivation of this paper. Lewis [9] is termed the "father of tension systems" in Psychology. He advocates that life is a constant interplay between tension reduction and tension production phases. In tension reduction phases, when agents have some goals in mind, they try to reduce the discrepancy between where they are (the current status quo) and where they want to be (their goals) in the near or distal future. These periods refer to goal-striving phases. When they succeed in reaching their goals, tensions release. This is a temporary equilibrium phase. Then, in tension production phases, agents need new challenges. They set new goals, creating a discrepancy between where they are in the current period and where they want to be in a near or more distal future. This refers to goal-setting phases. These powerful general ideas give rise to the (not formalized) famous theory of self-regulation in Psychology (goal setting, goal striving, goal revision, goal disengagement; see Bandura [10], Carver and Scheier [11], Oettingen and Gollwitzer [12], De Ridder and de Wit [13]).

This paper, being strongly mathematically oriented, will not consider tension production processes for the following reason. The optimization community emphasizes much more one side of self-regulation processes, the tension reduction processes. They consider a given function to be minimized and a given discrepancy between where an agent is and where he wants to be and try to find a position that minimizes this discrepancy. Possible obstacles to reduce this discrepancy refer to constraints. This is a traditional minimization problem with or without constraints. Mathematicians use step-by-step algorithms to reach such minimization solutions (tension-free positions). Such algorithms abound: descent methods, exact and inexact proximal algorithms, Newton methods, local models of approximation like trust region methods and rejection processes like branch and bound algorithms.

A recent variational rationality (VR) approach of human behaviors (Soubeyran [6–8]) offers a general model for tension systems, where tension reduction phases interplay with tension production phases. This model unifies different approaches of human behaviors in Psychology, Economics, Management Sciences, Political Sciences, Game theory, Artificial Intelligence and Mathematics. These include habit-routine formation and breaking processes at the individual level and exploration–exploitation adaptive learning dynamics in organizations. As a direct application, it helps to link the tension reduction aspect of the self-regulation theory in Psychology with traditional optimization problems in Mathematics. The first idea is that tension reduction is a progressive process that cannot reduce discrepancy in one step. Then VR approach poses the following question: How to fill such gaps in an acceptable manner? Because of not being able to do this in a single step, the agent will have to follow a transition defined as a succession of acceptable single steps (change or stay). Then, he will accept to entering into such a transition if acceptable.

The core of the VR approach is to define and modelize an acceptable single change if it is sufficiently worthwhile. This is the case when the motivation to change instead of stay is greater than the resistance to change instead of stay. Motivation to change is the utility of advantages to change, and resistance to change is the disutility of inconveniences to change. Advantages to change refer to the difference between where the agent wants to be or what he wants to have (an intention or its related payoff) and where he is or what he has (the status quo) in the current period. They modelize how tension production (creating a discrepancy) reduces tension, being a descent in Mathematics, bringing the agent closer to the minimum (tension-free state). Inconveniences to change represent the difference between the costs of being able to change and being able to stay. Under mild hypothesis, they are quasi-distances. Then, the VR approach shows how, in a quasi-metric space, a primal descent method [1, 14, 15] is worthwhile tension reduction process which reduces inconveniences to change. However, while the dual descent method ([2] and this study) reduces the advantages to change from where the agent is to where he wants to be, the tension-free position, this shows that descent methods belong to tension reduction systems.

Note that in terms of our VR approach, our descent process is also dual to the descent method considered in [3] because of the added reason that our stopping rule majorizes the norm of the subgradient by a function of advantages to change, while [3] majorizes the norm of the subgradient by a function of the inconveniences to change.

The mathematical side of the VR approach proposed by Soubeyran [6,7] has been further studied in a number of papers with applications to behavioral sciences including psychological modeling in quasi-metric space settings and minimal points, variational principles and variable preferences in set optimization; see Bao et al. [16,17]. To further clarify how these developments are related to those presented in this paper, we provide, among others, a short list of comparisons. The main point is that the variational rationality approach unifies the majority of algorithms and variational principles used in optimization in terms of descent methods, approximate linear quadratic model-based methods (trust region methods and other variants) and as tension reduction–tension production systems in Psychology. It helps operate different broad classifications among them:

(i) Weak resistance to change aspects (proximal algorithms; see Bento and Soubeyran [14,15]) and strong resistance to change aspects (Ekeland theorem, Caristi theorem and equivalent variational principles; see Bao et al. [16,17]);
(ii) Tension reduction processes with fixed ends (a given optimum for proximal algorithms) and those with moving ends (variational traps and desires for Ekeland and Caristi theorems);
(iii) Satisficing processes as tension reduction systems, where each satisficing step refers to a sufficient epsilon reduction in quasi-distance or sufficient epsilon reduction in payoff;
(iv) Equivalent proximal formulations in terms of proximal payoffs (profit or utility, minus costs of being able to change) and Ekeland-like worthwhile to change conditions (descent condition, that is, descent conditions where advantages to change, where advantages to change are higher than inconveniences to change);
(v) Tension reduction processes with variable and multidimensional preferences; see Cruz and Allende [18] for a steepest-like descent method and Bao et al. [16,17] for Ekeland theorem. In this case, all things seem to move, each period. Then, how is it possible for something to stay in the end?

The organization of the paper is as follows. In Sect. 2, the basic definitions used in the paper are presented. In Sect. 3, the main results are stated and proved. In Sect. 4, we present a short VR structure. In Sect. 5, we compare the speed of convergence of the two dual descent methods in a quasi-metric space. The conclusions are presented in Sect. 6.

2 Preliminaries on Nonsmooth Analysis

In this section, some elements of nonsmooth analysis are presented; see, for instance, [19,20]. Let us consider $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, a lower semicontinuous function. The domain of f, denoted by $\operatorname{dom} f$, is the subset of $\mathbb{R}^n$ on which f is finite valued. The function f is said to be proper when $\operatorname{dom} f \neq \emptyset$. If T denotes a multivalued mapping on $\mathbb{R}^n$, the domain of T is the set of elements $x \in \mathbb{R}^n$ such that $T(x) \neq \emptyset$.

Definition 2.1 (i) The Fréchet subdifferential (also known as regular subdifferential) of f at $x \in \mathbb{R}^n$ is defined by

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

(ii) The (Mordukhovich or limiting) subdifferential of f at $x \in \mathbb{R}^n$ is defined by

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

Note that $\hat{\partial} f(x) \subset \partial f(x)$. In the particular case where f is differentiable at x (resp. continuously differentiable around x), then $\hat{\partial} f(x) = \{\nabla f(x)\}$ (resp. $\partial f(x) = \{\nabla f(x)\}$). If f is a convex function, both subdifferentials $\hat{\partial} f(x)$ and $\partial f(x)$ coincide with the usual subdifferential for each $x \in \operatorname{dom} f$.

Denote Graph $\partial f := \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n : v \in \partial f(x)\}$. Checking the following closedness property of ∂f from the definition is straightforward. Let $(x^k, v^k)_{k \in \mathbb{N}} \subset \mathbb{R}^n \times \mathbb{R}^n$ such that $(x^k, v^k) \in$ Graph ∂f for all $k \in \mathbb{N}$. If $(x^k, v^k)_{k \in \mathbb{N}}$ converges to (x, v), and $f(x^k)$ converges to $f(x)$, then $(x, v) \in$ Graph ∂f; see [3] for more details. A necessary (but not sufficient) condition for $x \in \mathbb{R}^n$ to be a local minimum of f is $0 \in \partial f(x)$. A point $x \in \mathbb{R}^n$ satisfying the last inclusion is said to be limiting critical or simply critical.

3 Dual Descent Methods

In this section, we propose and study an inexact descent method whose full convergence is assured for cost functions that satisfy the Kurdyka–Lojasiewicz property.

Definition 3.1 A mapping $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is said to be a quasi-distance iff, for all $x, y, z \in \mathbb{R}^n$,

(i) $q(x, y) = q(y, x) = 0$ if, only if, $x = y$;
(ii) $q(x, y) \leq q(x, z) + q(z, y)$.

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

Throughout this paper, q represents a quasi-distance that satisfies the following assumption:

Assumption 3.1 There exist $\beta_1, \beta_2 \in \mathbb{R}_{++}$ such that $\beta_1 \|x - y\| \leq q(x, y) \leq \beta_2 \|x - y\|$, $x, y \in \mathbb{R}^n$.

In [4], the authors present some examples of quasi-distances that satisfy Assumption 3.1.

In the sequel, we consider sequences $\{x^k\}$ satisfying the following conditions (1), (2), which, for convenience, we refer to as follows:

Method 3.1 *Take $x^0 \in \mathbb{R}^n$, $0 < \bar{\lambda} \leq \tilde{\lambda} < +\infty$, and let a, b, be positive constants. For each $k = 0, 1, \ldots$, choose $\lambda_k \in [\bar{\lambda}, \tilde{\lambda}]$ and find $\left(x^{k+1}, w^{k+1}\right) \in \mathbb{R}^n \times \mathbb{R}^n$ such that*

$$(a/\lambda_k)\, q^2\left(x^k, x^{k+1}\right) \leq f(x^k) - f(x^{k+1}), \quad (\textit{worthwhile to change condition}) \tag{1}$$

$$w^{k+1} \in \partial f(x^{k+1}), \quad b\lambda_k \left\| w^{k+1} \right\|^2 \leq f(x^k) - f(x^{k+1}), \quad (\textit{curvature condition on the payoff function } f) \tag{2}$$

where $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ represents a quasi-distance.

Remark 3.1 Note that any sequence generated by the proximal point methods proposed in [1,4,21], as well as those complying with conditions (2.2) and (2.3) in [2], satisfies conditions (1) and (2) particularly when quasi-distance is the Euclidean distance and f is a differentiable function. Moreover, (1) and (2) can be easily seen to imply the following condition:

$$\exists\, w^{k+1} \in \partial f\left(x^{k+1}\right) : \quad f\left(x^k\right) - f\left(x^{k+1}\right) \geq \sqrt{ab} \left\| w^{k+1} \right\| q\left(x^k, x^{k+1}\right), \quad k = 0, 1, \ldots, \tag{3}$$

which is a natural extension to the quasi-metric setting of the discrete implicit angle condition (1.6) in [2] considering the nondifferentiability of f (an explicit version of this condition, in the differentiable case, was considered in Absil et al. [22]). It is important to note that conditions (1) and (2) are less restrictive than conditions H1 and H2 considered in [3], since (1) is equivalent to H1, and H1 and H2 together imply (2) (in the particular case where q is a Euclidean distance).

We can assume, without loss of generality, that $x^0 \in \operatorname{dom} f$. Note that if $\{x^k\}$ is a finite sequence, there exists $k_0 \in \mathbb{N}$ such that $f(x^{k+1}) = f(x^k)$, $k \geq k_0$. Hence, (2) shows that $w^{k+1} = 0$, $k \geq k_0$. Then, $\{x^k\}$ terminates at a critical point. In view of this, we can assume that $\{x^k\}$ is an infinite sequence. Unless stated otherwise, in the remainder of this paper, we assume that f is a proper and lower semicontinuous function bounded below and is continuous on dom f, and $\{x^k\}$ is an infinite sequence satisfying the conditions of Method 3.1.

3.1 Convergence Analysis

In this section, assuming that f satisfies the Kurdyka–Lojasiewicz property, we show full convergence of the sequence $\{x^k\}$ to a critical point. First, we present a partial convergence result.

Proposition 3.1 *The following statements hold.*

(i) *The sequence* $\{f(x^k)\}$ *is strictly decreasing;*
(ii) $\sum_{k=0}^{+\infty} q^2\left(x^k, x^{k+1}\right) < +\infty$*;*
(iii) $\lim_{k\to+\infty} q\left(x^k, x^{k+1}\right) = 0$*;*
(iv) *Each accumulation point of the sequence* $\{x^k\}$*, if any, is a critical point of* f*.*

Proof The proofs of items (i), (ii) and (iii) are simple verification. Let us deal with item (iv). Suppose that $\bar{x} \in \mathbb{R}^n$ is an accumulation point of $\{x^k\}$, and let $\{x^{k_j}\}$ be a subsequence converging to $\bar{x}$. It follows from (i) that $x^k \in \mathrm{dom}\, f$. Since f is a proper and lower semicontinuous function, $\bar{x} \in \mathrm{dom}\, f$, and as f is bounded below and is continuous on dom f, we have from (i) that $\{f(x^k)\}$ converges to $f(\bar{x})$. However, as $\{x^k\}$ is a sequence that satisfies the conditions of Method 3.1, there exists a sequence $\{w^{k+1}\}$ such that $w^{k+1} \in \partial f(x^{k+1})$, satisfying (2). In view of (2), $\{w^{k+1}\}$ converges to zero, because $0 < \bar{\lambda} \leq \lambda_k \leq \tilde{\lambda}$, $k \in \mathbb{N}$, and $\{f(x^k) - f(x^{k+1})\}$ converges to zero. As Graph ∂f is closed, it follows that $0 \in \partial f(\bar{x})$, which concludes the proof. □

As in [1,3,4], our main convergence result is restricted to functions that satisfy the Kurdyka–Lojasiewicz property. This was first introduced by Lojasiewicz [23] to real analytic functions and extended by Kurdyka [24] to differentiable definable functions in an o-minimal structure (for a detailed discussion on o-minimal structures, see Dries and Miller [25]). For extensions of the Kurdyka–Lojasiewicz property, in the Euclidian context, to the class of nonsmooth functions, see Bolte et al. [26], Bolte et al. [27] and Attouch et al. [21]. For extensions of the Kurdyka–Lojasiewicz property to functions defined on nonlinear spaces, see Kurdyka et al. [28], Lageman [29], Bolte et al. [30], Bento et al. [31] and Hosseini [32]. The next formal definition of the Kurdyka–Lojasiewicz property can be found in [21], where finding several examples and a good discussion on important classes of functions that satisfy the mentioned inequality is possible.

Definition 3.2 A proper and lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to have the Kurdyka–Lojasiewicz property at $\tilde{x} \in \mathrm{dom}\, \partial f$ iff there exist $\eta \in]0, +\infty]$, a neighborhood U of $\tilde{x}$, and continuous concave function $\varphi : [0, \eta[\to \mathbb{R}_+$ such that

$$\varphi(0) = 0, \quad \varphi \in C^1\left(]0, \eta[\right), \quad \varphi'(s) > 0, \quad s \in]0, \eta[; \tag{4}$$

$$\varphi'\left(f(x) - f(\tilde{x})\right) \mathrm{dist}\left(0, \partial f(x)\right) \geq 1, \quad x \in U \cap [f(\tilde{x}) < f < f(\tilde{x}) + \eta], \tag{5}$$

– $\mathit{dist}(0, \partial f(x)) := \mathit{inf}\{\|v\| : v \in \partial f(x)\}$,
– $[\eta_1 < f < \eta_2] := \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\}, \quad \eta_1 < \eta_2$.

Remark 3.2 f is known to have the Kurdyka–Lojasiewicz property at any noncritical point; see [21]. In [1] the authors considered that f has the Lojasiewicz property at some point $\tilde{x}$ if there exist $C > 0$ and $\delta > 0$ such that

$$|f(x) - f(\tilde{x})|^{\theta} \leq C\|w\|, \quad x \in B(\tilde{x}, \delta), \quad w \in \partial f(x). \tag{6}$$

Note that if f has the Lojasiewicz property at some point $\tilde{x}$, then it has the Kurdyka–Lojasiewicz property at $\tilde{x}$ with

$$\eta := \delta/2, \quad U := B(\tilde{x}, \delta/2), \quad \varphi(s) := (C/(1-\theta))s^{1-\theta}, \quad \theta \in [0, 1[. \tag{7}$$

Indeed, assume that there exist $C > 0$ and $\delta > 0$ satisfying (6). Taking η, U and φ as defined in (7), we can easily see that φ satisfies all the requirements in (4) and there holds

$$\varphi'\left(f(x) - f(\tilde{x})\right) = C\left(f(x) - f(\tilde{x})\right)^{-\theta} \geq \|w\|^{-1}, \quad w \in \partial f(x),$$
$$x \in U \cap [f(\tilde{x}) < f < f(\tilde{x}) + \eta],$$

from which it follows (5) and, in particular, φ defined (7) satisfies all the conditions of Definition 3.2.

Next, we present the main convergence result.

Theorem 3.1 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function that satisfies the Kurdyka–Lojasiewicz property at some $\tilde{x} \in \mathbb{R}^n$, and assume that Assumption 3.1 holds. Let $U \subset \mathbb{R}^n$ be a neighborhood of $\tilde{x}$, $\eta \in]0, +\infty]$ and $\varphi : [0, \eta[\to \mathbb{R}_+$ the objects appearing in Definition 3.2. Let $\varepsilon > 0$ be such that $B(\tilde{x}, \varepsilon) \subset U$, and take $0 < \rho < \varepsilon\beta_1$, $C_1 = \left(\sqrt{ab}/2\right)^{-1}$ and $C_2 = \sqrt{2\tilde{\lambda}}/\sqrt{a}\left(\sqrt{2} - 1\right)$ satisfying*

$$f(\tilde{x}) < f(x^0) < f(\tilde{x}) + \eta, \tag{8}$$

$$q\left(\tilde{x}, x^0\right) + \left(\left(\tilde{\lambda}/a\right)^{1/2} + C_2\right)\left(f(x^0) - f(\tilde{x})\right)^{1/2} + C_1\varphi\left(f(x^0) - f(\tilde{x})\right) < \rho. \tag{9}$$

$$x^k \in B_q(\tilde{x}, \rho) \Rightarrow x^{k+1} \in B_q(\tilde{x}, \varepsilon\beta_1) \ \text{with} \ f(x^{k+1}) \geq f(\tilde{x}), \quad k = 0, 1, \ldots. \tag{10}$$

Then, the sequence $\{x^k\}$ satisfies

$$x^k \in B_q(\tilde{x}, \rho), \quad k = 0, 1, 2, \ldots, \tag{11}$$

$$\sum_{k=0}^{+\infty} q\left(x^k, x^{k+1}\right) < +\infty, \tag{12}$$

and converges to a point $\bar{x}$. Moreover, $\bar{x}$ is a critical point of f, $f(\bar{x}) = f(\tilde{x})$ and $f(x^k) \to f(\tilde{x})$.

Proof From now on, we can assume, without loss of generality, that $f(\tilde{x}) = 0$, since we can change f to $f - f(\tilde{x})$. First, we need to prove that for each fixed $k_0 \in \mathbb{N}$ with $x^{k_0} \in B_q(\tilde{x}, \rho)$, we have

$$q\left(x^{k_0}, x^{k_0+1}\right) \leq C_1\left(\varphi\left(f(x^{k_0})/2\right) - \varphi\left(f(x^{k_0+1})/2\right)\right)$$
$$+ C_2\left(f(x^{k_0})^{1/2} - f(x^{k_0+1})^{1/2}\right). \tag{13}$$

Fix any $k_0 \in \mathbb{N}$ with $x^{k_0} \in B_q(\tilde{x}, \rho)$. In view of (8) and (10), we have

$$0 \leq f(x^{k_0+1}) \leq f(x^{k_0}) \leq f(x^0) < \eta. \tag{14}$$

Hence, the quantities $\varphi\left(f(x^{k_0})/2\right)$ and $\varphi\left(f(x^{k_0+1})/2\right)$ appearing in (13) make sense.

Now, we have two possibilities to consider:

(a) $f\left(x^{k_0+1}\right) > f\left(x^{k_0}\right)/2$;
(b) $f\left(x^{k_0+1}\right) \leq f\left(x^{k_0}\right)/2$.

Suppose that (a) holds. From concavity of the function φ in (4), we obtain the following inequalities:

$$\varphi\left(f(x^{k_0})/2\right) - \varphi\left(f(x^{k_0+1})/2\right) \geq (1/2)\varphi'\left(f(x^{k_0})/2\right)\left(f(x^{k_0}) - f(x^{k_0+1})\right), \tag{15}$$

$$\varphi'\left(f(x^{k_0})/2\right) \geq \varphi'\left(f(x^{k_0+1})\right). \tag{16}$$

Considering that (3) holds with $k = k_0$, from (15) and (16), we have

$$\begin{aligned} &\varphi\left(f(x^{k_0})/2\right) - \varphi\left(f(x^{k_0+1})/2\right) \\ &\quad \geq \left(\sqrt{ab}/2\right)\varphi'\left(f(x^{k_0+1})\right)\left\|w^{k_0+1}\right\| q\left(x^{k_0}, x^{k_0+1}\right), \end{aligned} \tag{17}$$

where $w^{k_0+1} \in \partial f\left(x^{k_0+1}\right)$. In view of (1), we have

$$q\left(x^{k_0}, x^{k_0+1}\right) \leq \left(\frac{\tilde{\lambda}}{a} f(x^{k_0})\right)^{1/2}. \tag{18}$$

The inequality (18) and triangle inequality give us

$$q\left(\tilde{x}, x^{k_0+1}\right) \leq q\left(\tilde{x}, x^{k_0}\right) + q\left(x^{k_0}, x^{k_0+1}\right) \leq \left(\frac{\tilde{\lambda}}{a} f(x^{k_0})\right)^{1/2} + q\left(\tilde{x}, x^{k_0}\right).$$

As $x^{k_0} \in B_q(\tilde{x}, \rho)$, (10) implies $x^{k_0+1} \in B_q(\tilde{x}, \varepsilon\beta_1)$. However, from Assumption 3.1 there exists $\beta_1 > 0$ such that $\beta_1\|\tilde{x} - x^{k_0+1}\| \leq q(\tilde{x}, x^{k_0+1})$. This tells us that $x^{k_0+1} \in B(\tilde{x}, \varepsilon) \subset U$. In view of (14),

$$x^{k_0+1} \in U \cap [0 < f < \eta].$$

Hence, as f satisfies the Kurdyka–Lojasiewicz inequality at $\tilde{x}$, we have $0 \notin \partial f(x^{k_0+1})$. Moreover, combining (5) with the definition of dist $(0, \partial f(x))$, we obtain

$$\varphi'\left(f(x^{k_0+1})\right)\left\|w^{k_0+1}\right\| \geq \varphi'\left(f(x^{k_0+1})\right) \operatorname{dist}\left(0, \partial f(x^{k_0+1})\right) \geq 1.$$

Hence, inequality (17) becomes

$$\left(\sqrt{ab}/2\right)^{-1}\left(\varphi\left(f(x^{k_0})/2\right)-\varphi\left(f(x^{k_0+1})/2\right)\right)\geq q\left(x^{k_0},x^{k_0+1}\right). \tag{19}$$

As $C_2\left(f\left(x^{k_0}\right)^{1/2}-f\left(x^{k_0+1}\right)^{1/2}\right)$ is positive, inequality (13) follows immediately from (19).

Now, let us assume that (b) holds. This fact gives us the following inequality:

$$f\left(x^{k_0}\right)^{1/2}\leq\sqrt{2}/\left(\sqrt{2}-1\right)\left(f\left(x^{k_0}\right)^{1/2}-f\left(x^{k_0+1}\right)^{1/2}\right).$$

Hence, from (18),

$$q\left(x^{k_0},x^{k_0+1}\right)\leq\sqrt{2\tilde{\lambda}}/\sqrt{a}\left(\sqrt{2}-1\right)\left(f\left(x^{k_0}\right)^{1/2}-f\left(x^{k_0+1}\right)^{1/2}\right).$$

Because $C_1\left(\varphi\left(f\left(x^{k_0}\right)/2\right)-\varphi\left(f\left(x^{k_0+1}\right)/2\right)\right)$ is positive, (13) follows immediately from the last inequality, which completes the proof of the statement.

Let us prove (11) by induction on k. First, we are going to prove (11) for $k=0,1$. From (9), we have that $x^0\in B(\tilde{x},\rho)$, and in view of (10), $x^1\in B(\tilde{x},\varepsilon\beta_1)$ and $f(x^1)\geq 0$. Using (1) with $k=0$, we have

$$q^2\left(x^0,x^1\right)\leq\frac{\tilde{\lambda}}{a}\left(f(x^0)-f(x^1)\right)\leq\frac{\tilde{\lambda}}{a}f(x^0). \tag{20}$$

A combination of the last inequality and the triangle inequality gives us

$$q\left(\tilde{x},x^1\right)\leq q\left(\tilde{x},x^0\right)+q\left(x^0,x^1\right)\leq\left(\frac{\tilde{\lambda}}{a}f(x^0)\right)^{1/2}+q\left(\tilde{x},x^0\right).$$

Combining the last inequality with (9) implies that x^1 belongs to $B_q(\tilde{x},\rho)$. Now, take $j>1$, and assume that (11) holds for all $k=1,\ldots,j-1$. In this case, (13) holds for $k=1,\ldots,j-1$. Then, we obtain

$$\sum_{i=1}^{j-1}q\left(x^i,x^{i+1}\right)\leq C_1\left(\varphi(f(x^0))-\varphi(f(x^j))\right)+C_2\left(f\left(x^0\right)^{1/2}-f\left(x^j\right)^{1/2}\right). \tag{21}$$

From the triangle inequality,

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

Thus, combining the last two inequalities with (20), we have

$$q\left(\tilde{x}, x^j\right) \leq q\left(\tilde{x}, x^0\right) + \left(\frac{\tilde{\lambda}}{a} f(x^0)\right)^{1/2} + C_1\varphi(f(x^0)) + C_2 f\left(x^0\right)^{1/2},$$

which, from (9), conclude the induction proof. Now, inequality (13) holds for all $k \geq 0$. Then, we have

$$\sum_{j=0}^{N} q\left(x^k, x^{k+1}\right) \leq C_1\varphi\left(f(x^0)\right) + C_2 f\left(x^0\right)^{1/2}, \quad N \geq 0. \tag{22}$$

Note that (12) follows immediately from (22). Now, combining the first inequality presented in Assumption 3.1 with (12), we conclude that $\{x^k\}$ is a Cauchy sequence and, consequently, converges to some point $\bar{x}$. Now, it follows from *vi)* of Proposition 3.1, that $\bar{x}$ is a critical point of f. Lastly, let us prove $f(\bar{x}) = f(\tilde{x})$ and $f(x^k) \to f(\tilde{x})$. As $\{f(x^k)\}$ is strictly decreasing and f is bounded below, we have $x^k \in \operatorname{dom} f$ for all k, and $\{f(x^k)\}$ converges to $\bar{f} = \inf_{k\geq 0} f(x^k)$. From (10), we have $f(x^{k+1}) \geq 0$ for all k, which implies that $\bar{f} \geq 0$. Suppose that $\bar{f} > 0$. As φ is a concave function and $x^k \in U$ for all k, we obtain

$$\varphi'\left(\bar{f}\right)\left\|w^k\right\| \geq \varphi'\left(f\left(x^k\right)\right)\left\|w^k\right\| \geq 1, \quad k = 0, 1, \ldots,$$

which is impossible, because $w^k \to 0$. Hence, $\bar{f} = 0$. As f is continuous on dom f, we have $\bar{x} \in \operatorname{dom} f$ and $f(x^k) \to f(\bar{x})$. Therefore $f(\bar{x}) = f(\tilde{x})$. □

Assumptions (8), (9) and (10) were used in [3], in the particular case where q was the Euclidean distance. Besides, the authors in [3] noted that if f satisfies the Kurdyka–Lojasiewicz property at an accumulation point of the sequence $\{x^k\}$, then assumptions (8), (9) and (10) are dispensable, since they are naturally verified for some x^{k_0} as a new initial point instead of x^0. Some papers have dealt with the case where the Kurdyka–Lojasiewicz property is assumed directly at an accumulation point of the sequence $\{x^k\}$; see, for instance, [2,3,14]. In [21] the authors considered sequences that naturally satisfied the conditions of the abstract model investigated in [3] and noted that several standard assumptions automatically guarantee the boundedness of the sequence $\{x^k\}$, hence its convergence. Next, we present a convergence result following the idea presented in [21].

Theorem 3.2 *Assume that f satisfies the Kurdyka–Lojasiewicz property* (5) *at each point in the domain of f, and suppose that Assumption 3.1 holds. Then*

(i) either $\left\|x^k\right\|$ tends to infinity;
(ii) or $\left\{q\left(x^k, x^{k+1}\right)\right\}$ is l^1, i. e.,

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

and as a consequence, $\{x^k\}$ converges to a critical point of f.

Proof Assume that (i) does not happen, and let $\tilde{x}$ be a limit point of $\{x^k\}$ and U, ρ, η, φ the associated objects, as defined in (5). Note that Proposition 3.1 implies that $\tilde{x}$ is critical of f. As $\{f(x^k)\}$ is a nonincreasing sequence, we deduce that $f(x^k)$ converges to $f(\tilde{x})$ and $f(x^k) \geq f(\tilde{x})$ for all $k \geq 0$. Hence, the sequence

$$b_k := q\left(\tilde{x}, x^k\right) + \left(\left(\tilde{\lambda}/a\right)^{1/2} + C_2\right)\left(f(x^k) - f(\tilde{x})\right)^{1/2} + C_1\varphi\left(f(x^k) - f(\tilde{x})\right)$$

admits 0 as a cluster point; we obtain the existence of $k_0 \in \mathbb{N}$ such that (8) and (9) are fulfilled with x^{k_0} as a new initial point. Now, let us prove Assumption (10). Take $x^k \in B_q(\tilde{x}, \rho)$. If necessary, shrink η so that $\eta < a/\tilde{\lambda}(\varepsilon\beta_1 - \rho)^2$. From (1), we have

$$q\left(x^k, x^{k+1}\right) \leq \sqrt{\frac{\lambda_k}{a}\left(f(x^k) - f(x^{k+1})\right)}. \tag{23}$$

Taking into account that

$$f(x^{k+1}) - f(\tilde{x}) \leq f(x^k) - f(\tilde{x}) \leq f(x^{k_0}) - f(\tilde{x}) < \eta, \quad k \geq k_0,$$

we deduce from (23) that $q(x^k, x^{k+1}) < \varepsilon\beta_1 - \rho$, $k \geq k_0$. Hence, from the triangle inequality,

$$q(\tilde{x}, x^{k+1}) \leq q(\tilde{x}, x^k) + q(x^k, x^{k+1}) < \rho + (\varepsilon\beta_1 - \rho) = \varepsilon\beta_1.$$

Therefore, Assumption (10) holds for all $k \geq k_0$. The conclusion is a consequence of Theorem 3.1. □

Remark 3.3 The stopping rule (2) limits the size of the norm of the subgradient as an incentive to stop to change, compared to advantages to change $f(x^k) - f(x^{k+1})$. Let us perform a comparison with the stopping rule $H2$ in [3], using a quasi-distance instead of a distance, $\|w^{k+1}\| \leq bq(x^k, x^{k+1})$. It shows that the two descent processes are dual in the following two senses. First, our stopping rule majorizes the norm of the subgradient $\|w^{k+1}\|$ by a function of advantages to change, namely, $\left(\left(f(x^k) - f(x^{k+1})\right)/b\lambda_k\right)^{1/2}$, while [3] majorizes the norm of the subgradient $\|w^{k+1}\|$ by a function of inconveniences to change, namely $b\left(q(x^k, x^{k+1})^2\right)^{1/2} = bq(x^k, x^{k+1})$. Second, duality works also on $f(x^{k+1}) \geq f(x^k)/2$ or the reverse, instead of $q(x^k, x^{k+1}) \geq q(x^{k-1}, x^k)/2$ or the reverse. This shows a great advantage in using concepts of the VR approach (see [6,7]) to further understand different descent dynamics in a unified manner. The dual descent method works on advantages to change to limit the norm of the gradient, while the primal descent method [3] approach works on inconveniences to change.

To establish the rate of convergence of Method 3.1, we assume that φ is given as in (7) and, for simplification, we consider that $(C/(1-\theta)) = 1$.

Corollary 3.1 *Under the assumptions of Theorem 3.1, there exists $\bar{x} \in \operatorname{dom} f$ such that*

$$q\left(x^k, \bar{x}\right) \leq \frac{\beta_2}{\beta_1}\left(\frac{C_1}{2^{1-\theta}} f\left(x^k\right)^{1-\theta} + C_2 f\left(x^k\right)^{1/2}\right), \quad k \geq 0.$$

Proof As we argued in Theorem 3.1, let us suppose that $f(\tilde{x}) = 0$. In view of (11), $x^k \in B_q(\tilde{x}, \rho)$ for all $k \geq 0$, which means (13) holds for all $k \geq 0$. Using inequality (13), considering that $\varphi(s) = s^{1-\theta}$ and $-f(x^{N+1}) < 0$, we have

$$q\left(x^k, x^{N+1}\right) \leq \sum_{j=k}^{N} q\left(x^j, x^{j+1}\right) \leq \frac{C_1}{2^{1-\theta}} f\left(x^k\right)^{1-\theta} + C_2 f\left(x^k\right)^{1/2}, \quad N \geq k \geq 0,$$

where the first inequality follows from the triangular inequality. As Assumption 3.1 holds, the last inequality becomes

$$\beta_1 \left\|x^k - x^{N+1}\right\| \leq \frac{C_1}{2^{1-\theta}} f\left(x^k\right)^{1-\theta} + C_2 f\left(x^k\right)^{1/2}, \quad N \geq k \geq 0.$$

From Theorem 3.1, there exists $\bar{x} \in \operatorname{dom} f$ such that $\lim_{k\to+\infty} x^k = \bar{x}$. Letting N goes to infinity in the last inequality and using again Assumption 3.1, we obtain

$$\frac{\beta_1}{\beta_2} q\left(x^k, \bar{x}\right) \leq \frac{C_1}{2^{1-\theta}} f\left(x^k\right)^{1-\theta} + C_2 f\left(x^k\right)^{1/2}, \quad k \geq 0,$$

which proves the desired result. □

Now, we present an analysis of the rate of convergence.

Theorem 3.3 *Under the assumptions of Theorem 3.1, there exists $\bar{x} \in \operatorname{dom} f$ satisfying the following statements:*

(i) *If $\theta \in [0, 1/2]$, there exists $\bar{k} \in \mathbb{N}$ such that*

$$q\left(x^k, \bar{x}\right) \leq \mu Q^k, \quad k \geq \bar{k}, \tag{24}$$

with $\mu = f(x^0)^{1/2}\beta_2/\beta_1 \left(C_1/2^{1-\theta} + C_2\right)$ and $Q = \left(1/(1 + b\bar{\lambda}/(1-\theta)^2)\right)^{1/2}$;

(ii) *If $\theta \in]1/2, 1[$, there exists $C > 0$ and $\bar{k} \in \mathbb{N}$ such that*

$$q\left(x^k, \bar{x}\right) \leq C \left(\sum_{j=\bar{k}}^{k-1} \lambda_j\right)^{(\theta-1)/(2\theta-1)}, \quad k \geq \bar{k}. \tag{25}$$

with $C = \beta_2/\beta_1 \left(C_1/2^{1-\theta} + C_2\right) \tau^{1-2\theta}$ and $\tau = \min\left\{2^{-2\theta}(2\theta - 1)b, (2^{2\theta-1} - 1)/\tilde{\lambda}\right\}$.

Proof As we argued before, let us assume that $f(\tilde{x}) = 0$. From Theorem 3.1 we have $x^k \in B(\tilde{x}, \rho)$ for all $k \geq 0$ and $f(x^k) \to 0$. This implies that there exists $\bar{k} \in \mathbb{N}$ such that

$$x^k \in B_q(\tilde{x}, \rho), \quad 0 \leq f(x^k) \leq \min\{\eta, 1\}, \quad k \geq \bar{k}.$$

Considering that $B_q(\tilde{x}, \rho) \subset B(\tilde{x}, \varepsilon)$, for $k \geq \bar{k}$, we have from (5),

$$(1-\theta) f\left(x^{k+1}\right)^{-\theta} \left\|w^{k+1}\right\| \geq 1, \quad k \geq \bar{k}. \tag{26}$$

Now, let us prove (24). We estimate the rate of decay of the sequence $\{f(x^k)\}$. From (2), we have

$$f\left(x^k\right) \geq f\left(x^{k+1}\right) + b\lambda_k \left\|w^{k+1}\right\|^2, \quad k \geq 0,$$

with $w^{k+1} \in \partial f\left(x^{k+1}\right)$. As $\lambda_k > \bar{\lambda}$, (26), along with the last inequality, gives us

$$f\left(x^k\right) \geq f\left(x^{k+1}\right) + b\bar{\lambda}/(1-\theta)^2 f\left(x^{k+1}\right)^{2\theta}, \quad k \geq \bar{k}. \tag{27}$$

As $0 \leq \theta \leq 1/2$, and $f(x^k) \leq 1$, we obtain $f\left(x^{k+1}\right)^{2\theta} \geq f\left(x^{k+1}\right)$, $k \geq \bar{k}$. Hence, (27) implies

$$f\left(x^{k+1}\right) \leq 1/(1 + b\bar{\lambda}/(1-\theta)^2) f\left(x^k\right), \quad k \geq \bar{k}.$$

Thus, it is easy to see that

$$f\left(x^k\right) \leq \left(1/(1 + b\bar{\lambda}/(1-\theta)^2)\right)^k f\left(x^0\right), \quad k \geq \bar{k}. \tag{28}$$

However, from Corollary 3.1, there exists $\bar{x} \in \operatorname{dom} f$ such that

$$q\left(x^k, \bar{x}\right) \leq \frac{\beta_2}{\beta_1}\left(\frac{C_1}{2^{1-\theta}} f\left(x^k\right)^{1-\theta} + C_2 f\left(x^k\right)^{1/2}\right), \quad k \geq \bar{k}. \tag{29}$$

As $f(x^k)^{1-\theta} \leq f(x^k)^{1/2}$, $k \geq \bar{k}$, in view of (29), we have

$$q\left(x^k, \bar{x}\right) \leq \frac{\beta_2}{\beta_1}\left(\frac{C_1}{2^{1-\theta}} + C_2\right) f\left(x^k\right)^{1/2}, \quad k \geq \bar{k}. \tag{30}$$

Combining (28) with (30), we obtain

$$q\left(x^k, \bar{x}\right) \leq f(x^0)^{1/2} \frac{\beta_2}{\beta_1}\left(\frac{C_1}{2^{1-\theta}} + C_2\right)\left[\left(\frac{1}{1 + b\bar{\lambda}/(1-\theta)^2}\right)^{1/2}\right]^k, \quad k \geq \bar{k},$$

which proves (24). Now, let us prove (25). For this purpose, we need to prove that

$$f\left(x^{k+1}\right)^{1-2\theta} - f\left(x^{k}\right)^{1-2\theta} \geq \tau\lambda_k, \quad k \geq \bar{k}, \tag{31}$$

where $\tau = \min\left\{2^{-2\theta}(2\theta-1)b/(1-\theta)^2, (2^{2\theta-1}-1)/\tilde{\lambda}\right\}$. For $k \geq \bar{k}$ fixed, we have two possibilities:

(a) $f(x^{k+1}) > f(x^k)/2$;
(b) $f(x^{k+1}) \leq f(x^k)/2$.

Suppose that (a) holds. As the function $[0, +\infty) \ni t \mapsto t^{1-2\theta}$ is convex,

$$f\left(x^{k+1}\right)^{1-2\theta} - f\left(x^{k}\right)^{1-2\theta} \geq (2\theta-1)\, f\left(x^{k}\right)^{-2\theta}\left(f\left(x^{k}\right) - f\left(x^{k+1}\right)\right).$$

As $\theta \in]1/2, 1[$, $2^{-2\theta} f(x^{k+1})^{-2\theta} < f(x^k)^{-2\theta}$. Combining (2) with the last inequality, we obtain

$$f\left(x^{k+1}\right)^{1-2\theta} - f\left(x^{k}\right)^{1-2\theta} \geq 2^{-2\theta}\,(2\theta-1)\, b\lambda_k f\left(x^{k+1}\right)^{-2\theta}\left\|w^{k+1}\right\|^2, \tag{32}$$

where $w^{k+1} \in \partial f(x^{k+1})$. Taking into account that inequality (26) holds, (32) becomes:

$$f\left(x^{k+1}\right)^{1-2\theta} - f\left(x^{k}\right)^{1-2\theta} \geq 2^{-2\theta}\,(2\theta-1)\, b/(1-\theta)^2\lambda_k \geq \tau\lambda_k.$$

However, if (b) holds, $2^{1-2\theta} f(x^{k+1})^{1-2\theta} \geq f(x^k)^{1-2\theta}$; hence,

$$f\left(x^{k+1}\right)^{1-2\theta} - f\left(x^{k}\right)^{1-2\theta} \geq \left(2^{2\theta-1}-1\right) f\left(x^{k}\right)^{1-2\theta}.$$

As $1/2 < \theta < 1$ and $f(x^k) \leq 1$, we have $f(x^k)^{1-2\theta} \geq 1$. Then,

$$f\left(x^{k+1}\right)^{1-2\theta} - f\left(x^{k}\right)^{1-2\theta} \geq \left(2^{2\theta-1}-1\right) \geq \tau\tilde{\lambda} \geq \tau\lambda_k.$$

Hence, (31) holds. Take $k \in \mathbb{N}$ greater that $\bar{k}$. Summing inequality (31) from $\bar{k}$ to $k-1$, we obtain

$$\begin{aligned} f\left(x^{k}\right)^{1-2\theta} - f\left(x^{\bar{k}}\right)^{1-2\theta} &= \sum_{j=\bar{k}}^{k-1}\left(f\left(x^{j+1}\right)^{1-2\theta} - f\left(x^{j}\right)^{1-2\theta}\right) \\ &\geq \tau\sum_{j=\bar{k}}^{k-1}\lambda_j, \quad k \geq \bar{k}. \end{aligned}$$

Taking into account that $f(x^{\bar{k}})^{1-2\theta} > 0$ and $1 - 2\theta < 0$, last inequality becomes

$$f\left(x^k\right) \leq \left(\tau \sum_{j=\bar{k}}^{k-1} \lambda_j\right)^{1-2\theta}, \qquad k \geq \bar{k}. \tag{33}$$

As $f(x^k)^{1/2} \leq f(x^k)^{1-\theta}$, we can combine (29) with (33) to obtain

$$q\left(x^k, \bar{x}\right) \leq \frac{\beta_2}{\beta_1}\left(\frac{C_1}{2^{1-\theta}} + C_2\right)\tau^{1-2\theta}\left(\sum_{j=\bar{k}}^{k-1} \lambda_j\right)^{(\theta-1)/(2\theta-1)}, \quad k \geq \bar{k},$$

and the proof is complete. □

4 Variational Rationality and Discrepancy Reduction

In this section we justify: (i) Why these two descent methods are dual in terms of Behavioral Sciences, and (ii) why we introduce quasi-distances.

The structure of the VR approach [6–8] defines each period a previous position $x = x^k$ and current position $y = x^{k+1}$, where a position is an activity, having something or being somewhere. A change $x = x^k \curvearrowright y = x^{k+1}$ is worthwhile if the motivation to change $M(x, y) = U[A(x, y)] \in \mathbb{R}_+$ is sufficiently high compared to the resistance to change $R(x, y) = D[I(x, y)] \in \mathbb{R}_+$, that is if $M(x, y) \geq \xi R(x, y)$, $\xi > 0$, where,

(a) ξ refers to how worthwhile the change is: The higher ξ, the more worthwhile the change is;
(b) $U[A(x, y)]$ refers to the utility $U[A]$ of advantages to change $A(x, y)$;
(c) $A(x, y) = f(x) - f(y)$ is the difference between the dissatisfaction $f(x) \in \mathbb{R}$ of being again in position x and the possibly reduced dissatisfaction $f(y)$ to be in the new position y;
(d) $M(x, y)$ is the positive tension generated by the discrepancy $A(x, y) = f(x) - f(y) \geq 0$;
(e) $D[I(x, y)]$ refers to the disutility $D[I]$ of inconveniences to change $I(x, y)$;
(f) $I(x, y) = C(x, y) - C(x, x) \geq 0$ is the difference between costs $C(x, y) \in \mathbb{R}_+$ of being able to change from x to y and costs $C(x, x) \in \mathbb{R}_+$ of being able to stay at x.

Our Method 3.1 is a specific instance of a worthwhile change, where $U[A] = A$, $D[I] = I^2$, $I(x, y) = q(x, y)$, $\xi_{k+1} = a/\lambda_k > 0$. Condition (1) reads $M(x^k, x^{k+1}) = f(x^k) - f(x^{k+1}) \geq \xi_{k+1} q^2(x^k, x^{k+1}) = \xi_{k+1} R(x^k, x^{k+1})$. It means that the change $x^k \curvearrowright x^{k+1}$ is worthwhile. Then, our dual descent must be worthwhile. Condition (2) gives a condition on the curvature of the unsatisfaction function at x^{k+1}.

Costs of being able to stay are supposed to be zero, $C(x, x) = 0$ for all $x \in X$. Costs of being able to change, $C(x, y) = q(x, y)$, are quasi-distances (see [8]) for a precise justification. They are not symmetric. The costs of being able to change from x to y are different from those of being able to change from y to x.

The primal descent method [1,14] considers a descent greater than half of the previous distance and quasi-distance, $q(x^k, x^{k+1}) < (1/2)q(x^{k-1}, x^k)$, while the dual descent method ([2] and this paper) considers a descent greater than half of the previous discrepancy $f(x^{k+1}) > (1/2)f(x^k)$.

5 Comparison of the Dual Speeds of Convergence

In this last section, we compare the speed of convergence of our dual descent method (this paper, Theorem 3.3) with the speed of convergence of the primal descent method [14, Theorem 3.12].

MAIN RESULT. The primal and dual descent methods have the same rate of convergence when $0 < \theta < 1/2$. In the opposite case, where $1/2 < \theta < 1$, the dual descent method converges faster than the primal descent methods if $\lambda_k \geq \lambda > 1$ for $k \geq k^*$.

Proof There are two cases,

Case I $0 < \theta < 1/2$. Then, the two methods give the same rate of convergence $q(x^k, \overline{x}) \leq cQ^k$ for $k \geq k_0$.

Case II $1/2 < \theta < 1$. Let $\sigma = (1-\theta)/(2\theta - 1) > 0$ and $L(k) = \sum_{j=\overline{k}}^{k-1} \lambda_j$. Then,

- the primal descent method gives $q_P(x^k, \overline{x}) \leq C_P k^{-\sigma}$,
- the dual descent method gives $q_D(x^k, \overline{x}) \leq C_D L(k)^{-\sigma}$.

This shows that the quasi-distance from the current position x^k to the limit position $\overline{x}$ decreases more in the dual descent method than in the primal descent method if $L(k)^{-\sigma} < k^{-\sigma}$, that is if $L(k) = \sum_{j=\overline{k}}^{k-1} \lambda_j > k$. This is the case when $\lambda_j \geq \lambda > 1$ for all $j = \overline{k}, \ldots, k-1$. In this situation, $(k - 1 - \overline{k})\lambda > k$ is satisfied if $k > k^* = (\lambda/(\lambda - 1))(1 + \overline{k})$. □

Theorem 3.3 gives us an algebraic interpretation of the speed of convergence of the sequence that satisfies the conditions of Method 3.1. Let us define the speed of convergence as proportional to the size of the quasi-distance between the current point x^k and critical point $\bar{x}$, $q(x^k, \bar{x})$. Now, let us look more closely to analyze conditions (24) and (25) accurately from the viewpoint of the recent VR approach; see Soubeyran [6–8]. In item (i), expressions of the constants μ and Q in (24) show that μ decreases when a and b increase, while Q decreases when only b increases. The speed of convergence increases in three situations:

(i) First, if a increases, i.e., when being worthwhile to change requires more advantages to change compared to inconveniences to change (see inequality (1));

(ii) Second, if b increases, because a higher b decreases both μ and Q, it decreases Q^k even more. A higher b requires, for the same advantage to change $f(x^k) - f(x^{k+1})$, to get a lower marginal advantage to change. This pushes the agent to stop changing earlier, each period, because, marginally (making one more incremental very small unit change), he will have a lower marginal advantage to change, despite being given the same advantage to change. Then, the size of each step will probably decrease. The effect of a higher b seems to be the most important for increase in the speed of convergence. Then, the stopping rule, which determines the size of each step, plays a major role in the speed of convergence;

(iii) Third, the lower the $f(x^0)$, the higher is the speed of convergence (the time spent to converge). This is very intuitive, because the gap to fill, $f(x^0) - \inf f \geq 0$, is smaller. Similarly, the expression of C in (25) shows us that C will decrease if a or b increases. In both cases, the same comments as above can be applied.

6 Conclusions

In this paper, we present an abstract descent method in the quasi-metric setting, where the descent condition works on payoffs (half the payoff, each time). The main convergence result is restricted to functions that satisfy the Kurdyka–Lojasiewicz property. An analysis of the rate of convergence is also presented. We also compare the rate of convergence of the primal and dual descent methods, showing that the dual descent method can converge faster in a large domain of cases. The motivation comes from Psychology, where tension reduction processes (goal striving) play a major role. In future research, we intend to investigate our descent method in more general contexts, for example, to the Riemannian context. It is worth noting that in recent years there has been an increasing number of studies proposing extensions of concepts and techniques, as well as methods of mathematical programming pertaining to the linear setting to the Riemannian context. Particularly, Li et al. [33] extended the notion of weak sharp minima and its characterization to the Riemannian setting. As in the linear setting, in the Riemannian context this notion is very instrumental for descent methods. In fact, introducing these notions has been motivated by numerical applications; see, for instance, Bento and Cruz Neto [34]. We foresee further progress in this topic in the near future.

Acknowledgments The work was supported by CAPES and CNPq.

References

1. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. Ser. B **116**(1–2), 5–16 (2009)
2. Alaa, N.E., Pierre, M.: Convergence to equilibrium for discretized gradient-like systems with analytic features. IMA J. Numer. Anal. **33**(4), 1291–1321 (2013)
3. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebric and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. Math. Program. Ser. A **137**, 91–129 (2013)
4. Moreno, F.G., Oliveira, P.R., Soubeyran, A.: A proximal algorithm with quasi distance. Application to habit's formation. Optimization **61**(12), 1383–1403 (2012)
5. Fuentes, M., Malick, J., Lemaréchal, C.: Descentwise inexact proximal algorithms for smooth optimization. Comput. Optim. Appl. **53**(3), 755–769 (2012)
6. Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. GREQAM, Aix Marseillle University (2009, (pre-print))
7. Soubeyran, A.: Variational rationality and the nsatisfied man routines and the course pursuit between aspirations, capabilities and beliefs. GREQAM, Aix Marseillle University (2010, (pre-print))
8. Soubeyran, A.: Variational rationality. Part 1. Worthwhile stay and change approach-avoidance human dynamics ending in traps. GREQAM, Aix Marseillle University (2016, (pre-print))

9. Lewis, K.: Field theory in Social Science. Harper, New York (1951)
10. Bandura, A.: Failures in self-regulation: Energy depletion or selective disengagement? Psychological Inquiry **7**(1), 20–24 (1996)
11. Carver, C. S., Scheier, M. F.: On the structure of behavioral self-regulation. In: M. Boekaerts, P. R. Pintrich, and M. Zeidner, Handbook of Self-Regulation., pp. 42–80. San Diego: Academic Press (2000)
12. Gabriele Oettingen, G., Gollwitzer, P.: Strategies of setting and implementing goals. In: Social psychological foundations of clinical psychology ed. by James E. Maddux (Eds.). New York : Guilford, pp. 114–135 (2010)
13. De Ridder, D.T.D., de Wit, J.B.F.: Self-regulation in health behavior. Wiley, London (2006)
14. Bento, G.C., Soubeyran, A.: A generalized inexact proximal point method for nonsmooth functions that satisfies kurdyka-Lojasiewicz inequality. Set-Valued Var. Anal. **23**(3), 501–517 (2015)
15. Bento, G.C., Soubeyran, A.: Generalized inexact proximal algorithms: Routine formation with resistance to change, following worthwhile changes. J. Optim. Theory Appl. **166**(1), 172–187 (2015)
16. Bao, T.Q., Mordukhovich, B.S., Soubeyran, A.: Variational analysis in psychological modeling. J. Optim. Theory Appl. **164**(1), 290–315 (2015)
17. Bao, T.Q., Mordukhovich, B.S., Soubeyran, A.: Minimal points, variational principles, and variable preferences in set optimization. J. Nonlinear Convex Anal. **16**, 1511–1537 (2015)
18. Cruz, Bello: J.Y., Bouza Allende, G.: A Steepest Descent-Like Method for Variable Order Vector Optimization Problems. J. Optim. Theory Appl. **162**(2), 371–391 (2014)
19. Mordukhovich, B. S.: Variational analysis and generalized differentiation. I. Basic theory, Grundlehren der Mathematischen Wissenschaften, **330**. Springer-Verlag, Berlin (2006)
20. Rockafellar, R.T., Wets, R.J.B.: Variational analysis, Grundlehren der Mathematischen Wissenschaften, 317. Springer-Verlag, Berlin (1998)
21. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
22. Absil, P.A., Mahony, R., Andrews, B.: Convergence of the iterates of descent methods for analytic cost functions. SIAM J. Optim. **16**(2), 531–547 (2005)
23. Lojasiewicz, S.: Une propriété topologique dessous-ensembles analytiques réels. Les Équations aux DérivéesPartielles, Éditions du centre National de la RechercheScientifique **87–89**, (1963)
24. Kurdyka, K.: On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier (Grenoble) **48**(3), 769–783 (1998)
25. van den Dries, L., Miller, C.: Geometric categories and o-minimal structures. Duke Math. J. **84**(2), 497–540 (1996)
26. Bolte, J., Daniilidis, A., Lewis, A.: The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2006)
27. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
28. Kurdyka, K., Mostowski, T., Parusiński, A.: Proof of the gradient conjecture of R. Thom. Ann. of Math. (2) **152**(3), 763–792 (2000)
29. Lageman, C.: Convergence of gradient-like dynamical systems and optimization algorithms, Ph.D. Thesis, Universität Würzburg (2007)
30. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Amer. Math. Soc. **362**(6), 3319–3363 (2010)
31. Bento, G.C.: Cruz Neto, J.X., Oliveira, P.R.: A New Approach to the Proximal Point Method: Convergence on General Riemannian Manifolds. J. Optim. Theory Appl. **3**, 743–755 (2016)
32. Hosseini, S.: Convergence of nonsmooth descent methods via Kurdyka-Lojasiewicz inequality on Riemannian manifolds. Hausdorff Center for Mathematics and Institute for Numerical Simulation, University of Bonn (2015, (INS Preprint No. 1523))
33. Li, C., Mordukhovich, B.S., Wang, J., Yao, J.C.: Weak sharp minima on Riemannian manifolds. SIAM J. Optim. **21**(4), 1523–1560 (2011)
34. Bento, G.C., Cruz Neto, J.X.: Finite termination of the proximalpoint method for convex functions on Hadamard manifolds. Optim **63**(9), 1281–1288 (2014)