A GENERALIZED PROXIMAL LINEARIZED ALGORITHM FOR DC FUNCTIONS WITH APPLICATION TO THE OPTIMAL SIZE OF THE FIRM PROBLEM

JOÃO CARLOS SOUZA, PAULO ROBERTO OLIVEIRA, AND ANTOINE SOUBEYRAN

ABSTRACT. A proximal linearized algorithm with a quasi distance as regularization term for minimizing a DC function (difference of two convex functions) is proposed. If the sequence generated by our algorithm is bounded, it is proved that every cluster point is a critical point of the function under consideration, even if minimizations are performed inexactly at each iteration. A sufficient condition for global convergence is given for a particular case. Finally, an application is given, in a dynamic setting, to determine the limit of the firm, when increasing returns matter in the short run.

Keywords: Proximal point algorithm, generalized algorithm, DC functions, limit of the firm, variational rationality.
MSC[2010]: 90C26, 49M37, 65K10, 49J52, 91E10

1. INTRODUCTION

It is well known that the class of proximal point algorithms is one of the most studied methods for finding zeros of maximal monotone operators, and in particular, it is used to solve convex optimization problems. The proximal point method was introduced into optimization literature by Martinet [1]. It is based on the notion of proximal mapping introduced earlier by Moreau [2]. The proximal point method was popularized by Rockafellar [3], who showed that the following algorithm

$$0 \in c_k T(x^{k+1}) + x^{k+1} - x^k \tag{1}$$

converges to a point satisfying $0 \in T(x^*)$, under some mild conditions, even if each subproblem is performed inexactly, which is an important consideration in practice. In particular, if $T(\cdot) = \partial f(\cdot)$, where f is a convex function, then (1) becomes

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \{f(x) + \frac{1}{2c_k}||x - x^k||^2\}, \tag{2}$$

and the sequence converges to a point $x^* \in \arg\min f(x)$. In this case, the approach used in [3] is only useful for convex problems, because the idea underlying the results is based on the monotonicity of subdifferential operators of convex functions. Therefore, the proximal point method for nonconvex functions has been investigated by many authors in different contexts (see [4, 5, 6, 7, 8, 9] and references therein) as well as its inexact versions; see e.g. [10, 11, 12, 13, 14, 15].

On the other hand, the Euclidean norm in (2) has been changed by an adequate "like-distance" such that some axioms of distance are verified, preserving the nice properties of convexity, continuity and coercivity of the Euclidean norm. Extensions of proximal point method by using "nonlinear" or "generalized" regularizations were considered, for instance, in [7, 10, 14, 15, 16, 17, 18, 19]. The works [14, 15, 16] are devoted to study convergence properties of "generalized" proximal point method where the regularization term is a quasi distance (or quasi metric). Bento and Soubeyran [14] discuss how such "generalized" proximal point method can be a nice tool to modelize the dynamics of human behaviors in the context of the "variational rationality approach", see also Soubeyran [20, 21, 22]. Application of quasi metric spaces to Behavioral Sciences (Psychology, Economics, Management, Game theory,...) and computer theory can be found in [23, 24, 25].

Recently, [26] proposed a powerful algorithm, namely, the proximal alternating linearized minimization (PALM) algorithm, to solve a wide class of nonconvex and nonsmooth problems. Another algorithm used successfully to solve practical problems is the majorization-minimization algorithm or successive upper-bound minimization algorithm; see [27] and [28].

The goal of this paper is two-fold. First, inspired by the methods mentioned before, we propose a "generalized" proximal linearized algorithm for solving the well known class of nonconvex and nonsmooth DC problems. There is a huge literature on DC theory both from a theoretical point of view and for algorithmic purposes; see e.g. [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46]. DC optimization algorithms have been proved to be particularly successful for analyzing and solving a variety of highly structured and practical problems; see, for instance, [45] and references therein. Application of DC theory in game theory can be found in [41], plasma physics and fluid mechanics can be found in [46], and examples of DC functions from various parts of analysis can be found in [42].

As a second contribution, we provide an application, in a dynamic context, to the very important and difficult problem of the limit of the firm, when increasing returns prevail in the short run, using the recent (VR) variationality approach of a lot of stay/stability and change dynamics in Behavioral Sciences (see Soubeyran [20, 21, 22]). Different variants of this example can be found in Soubeyran [20, 21], Bento and Soubeyran [14, 15], and Bao et al. [47]. But none of them examines the very important case of increasing returns, which is the realistic case for production costs, as we do here, as an application of the proximal point method for DC optimization.

The organization of this paper is as follows. In Section 2 some preliminary results in subdifferential theory, DC optimization, and quasi distance are presented. In Section 3 a generalized proximal linearized algorithm is discussed as well as its convergence properties. In Section 4 an inexact version of the algorithm is proposed, and a sufficient condition for global convergence is presented for a particular case. Finally, Section 5 is devoted to determine the limit of the firm, when increasing returns matter in the short run. This is a difficult problem, both for conceptual and technical reasons. Future works are mentioned in the conclusions.

2. Preliminaries

Let $\Gamma_0(\mathbb{R}^n)$ denote the convex cone of all the proper (i.e. not identically equal to $+\infty$) lower semicontinuous convex functions from $\mathbb{R}^n$ to $\mathbb{R}\cup\{+\infty\}$, let $\langle\cdot,\cdot\rangle$ be the canonical inner product, and $||\cdot||$ the corresponding Euclidean norm on $\mathbb{R}^n$. The effective domain of a function f, denoted by $\mathrm{dom}(f)$, is defined as

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n \ : \ f(x) < +\infty\}.$$

2.1. Subdifferential theory.

Let us recall some definitions and properties of the subdifferential theory which can be found, for instance, in [48, 49].

Definition 1. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.*

(1) *The subdifferential of f at x, denoted by $\partial f(x)$, is defined as follows*

$$\partial f(x) = \begin{cases} \{v \in \mathbb{R}^n \ : \ f(y) \geq f(x) + \langle v, y-x\rangle \quad \forall y \in \mathbb{R}^n\}, \textit{ if } x \in \textit{dom}(f); \\ \emptyset, \textit{ if } x \notin \textit{dom}(f). \end{cases}$$

(2) *The Fréchet subdifferential of f at x, denoted by $\partial_F f(x)$, is defined as follows*

$$\partial_F f(x) = \begin{cases} \{v \in \mathbb{R}^n \ : \ \liminf\limits_{\substack{y\to x\\ y\neq x}} \frac{f(y)-f(x)-\langle v,y-x\rangle}{||x-y||} \geq 0\}, \textit{ if } x \in \textit{dom}(f); \\ \emptyset, \textit{ if } x \notin \textit{dom}(f). \end{cases}$$

(3) *The limiting-subdifferential of f at x, denoted by $\partial_L f(x)$, is defined as follows*

$$\partial_L f(x) = \begin{cases} \{v \in \mathbb{R}^n \ : \ \exists x^k \to x, \ f(x^k) \to f(x), \ v^k \in \partial_F f(x^k) \to v\}, \textit{ if } x \in \textit{dom}(f); \\ \emptyset, \textit{ if } x \notin \textit{dom}(f). \end{cases}$$

Recall that when f is a proper, lower semicontinuous and convex function, and $x \in \mathrm{dom}(f)$, then $\partial f(x) = \partial_F f(x) \neq \emptyset$.

In our convergence analysis, the following results will be used.

Proposition 1. *For a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point $\overline{x}$ where f is finite, the subgradient sets $\partial_L f(x)$ and $\partial_F f(x)$ are closed, with $\partial_F f(x)$ convex and $\partial_F f(x) \subset \partial_L f(x)$.*

Proof. See [48, Theorem 8.6]. □

Proposition 2. *If $f = f_1 + f_2$ with f_1 finite at $\overline{x}$ and f_2 continuously differentiable on a neighborhood of $\overline{x}$, then*

$$\partial_F f(x) = \partial_F f_1(x) + \nabla f_2(x), \ \partial_L f(x) = \partial_L f_1(x) + \nabla f_2(x).$$

Proof. See [48, Exercise 8.8]. □

Proposition 3. *If a proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ has a local minimum at x, then $0 \in \partial_F f(x)$ and $0 \in \partial_L f(x)$. In the convex case, these two conditions are both necessary and sufficient for a global minimum. If $f = f_1 + f_2$ with f_2 continuously differentiable, condition $0 \in \partial_F f(x)$ takes the form $-\nabla f_2(x) \in \partial_L f_1(x)$.*

Proof. See [48, Theorem 10.1]. □

A point $x \in \mathbb{R}^n$ such that $0 \in \partial_L f(x)$ is called a critical point of f.

Proposition 4. *If f_1 is locally Lipschitz continuous around x, f_2 is lower semicontinuous and proper with $f_2(x)$ finite, then*

$$\partial_L(f_1 + f_2)(x) \subset \partial_L f_1(x) + \partial_L f_2(x).$$

Proof. See [48, Exercise 10.10]. □

Proposition 5. *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}_+$ be Lipschitz continuous around x, then*

$$\partial_L(f_1.f_2)(x) \subset f_2(x)\partial_L f_1(x) + f_1(x)\partial_L f_2(x).$$

Proof. See [49, Theorem 7.1]. □

Proposition 6. *Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be lower semicontinuous, one of them Lipschitz continuous around $x \in dom(f_1) \cap dom(f_2)$. Then for any $\delta > 0$ and $\gamma > 0$, one has*

$$\partial_F(f_1 + f_2)(x) \subset A + \gamma\overline{B}(0, 1),$$

where $A = \cup \left\{\partial_F f_1(x_1) + \partial_F f_2(x_2) \ : \ x_i \in B(x, \delta), \ |f_i(x_i) - f_i(x)| \leq \delta, \ i = 1, 2\right\}$.

Proof. See [49, Proposition 2.7]. □

Definition 2. *A set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally bounded at a point $x \in \mathbb{R}^n$ if for some neighborhood $V \in \mathcal{N}(x)$ the set $S(V) \subset \mathbb{R}^m$ is bounded, where $\mathcal{N}(x)$ is the set of all neighborhood of x. A set-valued mapping is called locally bounded on $\mathbb{R}^n$ if this holds at every $x \in \mathbb{R}^n$.*

Proposition 7. *A mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally bounded if, and only if, $S(B)$ is bounded for every bounded set B. This is equivalent to the property that whenever $v^k \in S(x^k)$ and the sequence $\{x^k\}$ is bounded, then the sequence $\{v^k\}$ is bounded.*

Proof. See [48, Proposition 5.15]. □

Definition 3. *A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is locally lower semicontinuous at $\hat{x}$, a point where $f(\hat{x})$ is finite, if there is an $\epsilon > 0$ such that all sets of the form $\{x \in B(\hat{x}, \epsilon) \ : \ f(x) \leq \alpha\}$, with $\alpha \leq f(\hat{x}) + \epsilon$, are closed.*

The local lower semicontinuity of f at $\hat{x}$, where $f(\hat{x})$ is finite, can be interpreted as the local closeness of the epigraph of f at $(\hat{x}, f(\hat{x}))$; see [48, Exercise 1.34].

Proposition 8. *Suppose $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is locally lower semicontinuous at x with $f(x)$ finite. Then the following conditions are equivalent:*

(1) *f is locally Lipschitz continuous around x;*
(2) *the mapping $\partial_F f : x \mapsto \partial_F f(x)$ is locally bounded at x;*
(3) *the mapping $\partial_L f : x \mapsto \partial_L f(x)$ is locally bounded at x;*

Moreover, when these conditions hold, $\partial_L f(x)$ is nonempty and compact.

Proof. See [48, Theorem 9.13]. □

2.2. **Difference of convex functions.** A general DC program is of the form

$$\alpha = \inf\{f(x) = g(x) - h(x) \quad : \quad x \in \mathbb{R}^n\},$$

with $g, h \in \Gamma_0(\mathbb{R}^n)$. Such a function f is called a DC function while the convex functions g and h are DC components of f. In DC programming, the convention

$$(+\infty) - (+\infty) = +\infty$$

has been adopted to avoid the ambiguity $(+\infty)-(+\infty)$ that does not present any interest here. Note that the finiteness of α implies that $\mathrm{dom}(g) \subseteq \mathrm{dom}(h)$. Such inclusion will be assumed throughout the paper.

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous DC function (resp. bounded from below). Then, f has lower semicontinuous DC components g and h, with $\inf_{x\in\mathbb{R}^n} h(x) = 0$ (resp. g bounded from below); see [50, Proposition 2.4] and [51, Proposition 3.2]. It is well known that a necessary condition for $x \in \mathrm{dom}(f)$ to be a local minimizer of f is $\partial h(x) \subset \partial g(x)$. In general, this condition is hard to be reached. So, we will focus our attention on finding points such that $\partial h(x) \cap \partial g(x) \neq \emptyset$, namely, critical points of f.

It is worth mentioning the richness of the class of DC functions which is a subspace containing the class of lower-$\mathcal{C}^2$ functions (f is said to be lower-$\mathcal{C}^2$ if f is locally a supremum of a family of $\mathcal{C}^2$ functions). In particular, $\mathcal{DC}(\mathbb{R}^n)$ contains the space $\mathcal{C}^{1,1}$ of functions whose gradient is locally Lipschitz. $\mathcal{DC}(\mathbb{R}^n)$ is closed under the operations usually considered in optimization. For instance, a linear combination, a finite supremum or the product of two DC functions remains DC. Locally DC functions on $\mathbb{R}^n$ are DC functions on $\mathbb{R}^n$ (see [38] and references therein for the details). Under some caution we can say that $\mathcal{DC}(\mathbb{R}^n)$ constitutes a minimal realistic extension of $\Gamma_0(\mathbb{R}^n)$.

2.3. **Quasi distance.**

Definition 4. *A quasi metric space is a pair (X, q) such that X is a nonempty set, and $q : X \times X \to \mathbb{R}_+$, called a quasi metric or quasi distance, is a mapping satisfying:*

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

Therefore, metric spaces are quasi metric spaces satisfying the symmetric property $q(x,y) = q(y,x)$. Quasi distances are not necessarily convex, continuously differentiable or coercive functions. Examples of quasi distances can be found in [16] and references therein. In this paper, we consider quasi distances satisfying the following condition:

Condition 1: There are positive real numbers $\alpha > 0$ and $\beta > 0$ such that

$$\alpha||x-y|| \leq q(x,y) \leq \beta||x-y|| \quad \forall x,y \in \mathbb{R}^n. \tag{3}$$

Remark 1. *This condition was used to prove convergence of the proximal point algorithm for nonconvex and nonsmooth functions that verify the Kurdyka-Łojasiewicz property, see* [14, 15, 16]. *Moreno et al.* [16] *present several examples of quasi distances highlighting two that satisfy Condition 1.*

Proposition 9. *Let $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ be a quasi distance that verifies Condition 1. Then, for each $z \in \mathbb{R}^n$, the functions $q(.,z)$, $q(z,.)$ are Lipschitz, and $q^2(.,z)$, $q^2(z,.)$ are locally Lipschitz functions on $\mathbb{R}^n$.*

Proof. See [16, Propositions 3.6 and 3.7]. □

3. Generalized proximal linearized algorithm

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper DC function bounded from below with DC components g and h, i.e., $f(x) = g(x) - h(x)$ with $g, h \in \Gamma_0(\mathbb{R}^n)$. In this section, we consider a generalized proximal linearized algorithm for finding critical points of a DC function f. At each iteration our algorithm linearizes the function $f(x)$ but never directly minimize it, while it minimizes the function $g(x)$ in conjunction with the linearization of $h(x)$ and a generalized regularization. As mentioned before, this kind of linearized method has proved to be efficient for solving a large number of problems, for instance, the convex problem of minimizing a sum of two convex functions; see [26, 52]. Here, the term "generalized" refers to a quasi distance which does not satisfy all the properties of a distance, while preserving the nice properties of convexity, continuity and coercivity of the Euclidean norm. This kind of generalized method is more appropriate for applications in behavioral science; see [14, 15, 16].

Algorithm 1 (GPLA)

Step 1: Given an initial point $x^0 \in \operatorname{int}(\operatorname{dom}(h))$, and a bounded sequence of positive numbers $\{\lambda_k\}$ such that $\liminf_k \lambda_k > 0$.

Step 2: Calculate

$$w^k \in \partial h(x^k). \tag{4}$$

Step 3: Compute

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} \left\{ g(x) - \langle w^k, x - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, x) \right\}. \tag{5}$$

If $x^{k+1} = x^k$, stop. Otherwise, set $k := k+1$ and return to Step 2.

Remark 2. *The well definition of* $\{w^k\}$ *and* $\{x^k\}$ *is guaranteed if* h *is a convex function, and* g *is a convex and bounded from below function, respectively. As mentioned in the previous section, if* f *is a DC function bounded from below, then* f *admits a DC decomposition with* g *bounded from below. The condition* $\liminf_k \lambda_k > 0$ *is easily satisfied if* $0 < c_1 \leq \lambda_k \leq c_2$*, for all* k*. Note that, if* $h \equiv 0$ *and* $q(x,y) = ||x-y||$*, algorithm GPLA becomes exactly the classical proximal point algorithm* (2) *proposed by Rockafellar* [3].

We emphasize that our algorithm is different of the DCA algorithm considered by Pham and Souad [43]. Our algorithm share the same idea of DCA algorithm, namely, they linearize some component $g(\cdot)$ or $h(\cdot)$; or both of the DC objective function $f(x) = g(x) - h(x)$. However, our algorithm is simpler, because linearization is done directly, and not on the dual components, besides the fact that the proximal point method is more efficient than the subgradient method. Algorithm 1 is closely related to the one proposed by Sun et al. [36], but our algorithm seems to be more appropriate for applications in behavioral science using the "variational rationality approach" (see Soubeyran [20, 21, 22]), because in this "variational rationality approach", costs of being able to change from the current position x^k to x^{k+1}, and costs of being able to stay in the current position x^k are not necessarily symmetric and equal to zero, respectively.

Remark 3. *Note that in* (5)*, instead of minimize* $f(\cdot) = g(\cdot) - h(\cdot)$ *directly, we minimize the linear approximation* $g(\cdot) - \langle w^k, \cdot - x^k\rangle$ *in addiction with the proximal regularization. We mention that, if we replace such approximation by* $g(\cdot) - h(x^k) - \langle w^k, x - x^k\rangle$*, then algorithm GPLA becomes*

$$x^{k+1} \in arg \min_{x\in\mathbb{R}^n} \varphi_k(x), \tag{6}$$

where $\varphi_k(x) = g(x) - h(x^k) - \langle w^k, x - x^k\rangle + \dfrac{1}{2\lambda_k} q^2(x^k, x)$*. In this case, for each* $k \geq 0$*, from the convexity of* h*, we have* $f(x) \leq \varphi_k(x)$*, for all* $x \in dom(f)$*. Therefore, algorithm* (6) *minimizes upper-bounds of the objective function* f*, and each minimization step decreases the value of the objective function. This kind of method is often called majorization-minimization* [27] *or successive upper- bound minimization* [28]*. Majorizing surrogates have been used successfully in large scale problems* [29]*, signal processing literature about sparse estimation* [30, 31]*, linear inverse problems in image processing* [32, 33]*, and matrix factorization* [34, 35].

The following result will be used in the convergence theorem.

Proposition 10. *Let* $\{x^k\}$ *be the sequence generated by algorithm GPLA. Then there exist* $\xi^{k+1} \in \partial g(x^{k+1})$ *and* $\eta^{k+1} \in \partial_L(q(x^k, .))(x^{k+1})$ *satisfying*

$$w^k = \xi^{k+1} + \frac{q(x^k, x^{k+1})}{\lambda_k}\eta^{k+1}. \tag{7}$$

Proof. Combinig (5) with Proposition 3, we have

$$\begin{aligned} 0 &\in \partial_L \left(g(\cdot) - \langle w^k, \cdot - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, \cdot) \right)(x^{k+1}) \\ &\subset \partial_L \left(g(\cdot) - \langle w^k, \cdot - x^k \rangle \right)(x^{k+1}) + \frac{1}{2\lambda_k} \partial_L \left(q^2(x^k, \cdot) \right)(x^{k+1}) \\ &= \partial_L g(x^{k+1}) - w^k + \frac{1}{2\lambda_k} \partial_L \left(q^2(x^k, \cdot) \right)(x^{k+1}), \end{aligned}$$

using Proposition 4 in the second statement and Proposition 3 in the equality. Hence,

$$w^k \in \partial g(x^{k+1}) + \frac{1}{2\lambda_k} \partial_L \left(q^2(x^k, \cdot) \right)(x^{k+1})$$

having in mind that $g \in \Gamma_0(\mathbb{R}^n)$ and $x^{k+1} \in \mathrm{dom}(g)$. From Proposition 5 and 9, we have $\partial_L \left(q^2(x^k, \cdot) \right)(x^{k+1}) \subset 2q(x^k, x^{k+1}) \partial_L \left(q(x^k, \cdot) \right)(x^{k+1})$. Altogether, we obtain

$$w^k \in \partial g(x^{k+1}) + \frac{q(x^k, x^{k+1})}{\lambda_k} \partial_L q(x^k, \cdot)(x^{k+1}),$$

and the proof is concluded. □

Now we establish the convergence of the algorithm. We begin by showing that algorithm GPLA is a descent algorithm.

Theorem 1. *The sequence $\{x^k\}$ generated by algorithm GPLA satisfies:*

(1) *either the algorithm stops at a critical point;*
(2) *or f decreases strictly, i.e., $f(x^{k+1}) < f(x^k) \quad \forall k \geq 0$.*

Proof. If $x^{k+1} = x^k$, then $q(x^k, x^{k+1}) = 0$, and using (7), we have $w^k = \xi^{k+1} \in \partial g(x^k)$. From definition of the algorithm, we know that $w^k \in \partial h(x^k)$. Hence, $w^k \in \partial h(x^k) \cap \partial g(x^k)$, which shows that x^k is a critical point of f. Now, suppose $x^{k+1} \neq x^k$. It follows from (4) that

$$h(x^{k+1}) \geq h(x^k) + \langle w^k, x^{k+1} - x^k \rangle. \tag{8}$$

On the other hand, from (5), we have

$$g(x^k) \geq g(x^{k+1}) - \langle w^k, x^{k+1} - x^k \rangle + \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}). \tag{9}$$

Adding inequalities (8) and (9), we obtain

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) > f(x^{k+1}), \tag{10}$$

where in the last inequality we used the fact that $\lambda_k > 0$ and $q^2(x^k, x^{k+1}) > 0$. □

The following result is a consequence of the last theorem.

Corollary 2. *Let $\{x^k\}$ be the sequence generated by algorithm GPLA. Then $\{f(x^k)\}$ is a convergent sequence. Furthermore, if f is a continuous function and $\{x^k\}$ is bounded, then $\lim_{k\to\infty} f(x^k) = f(\hat{x})$, for any cluster point $\hat{x}$ of $\{x^k\}$.*

The following proposition will be used to prove the main convergence theorem.

Proposition 11. *Consider $\{x^k\}$ generated by algorithm GPLA, then $\lim_{k\to+\infty} q(x^k, x^{k+1}) = 0$.*

Proof. From (10), we can write

$$\frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}),$$

and therefore,

$$\sum_{k=0}^{n-1} \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

Since f is bounded from below and $\liminf_k \lambda_k > 0$, if k goes to $+\infty$, this shows that $\sum_{k=0}^{\infty} q^2(x^k, x^{k+1})$ is convergent. In particular, $\lim_{k\to+\infty} q(x^k, x^{k+1}) = 0$. □

Consider $h \in \Gamma_0(\mathbb{R}^n)$. In view of Proposition 7 and 8, we can easily verify that if a sequence $\{x^k\}$ is bounded and $w^k \in \partial h(x^k)$, then $\{w^k\}$ is also bounded. The following theorem is the main result of this section.

Theorem 3. *Let $\{x^k\}$ be the sequence generated by algorithm GPLA. Then, every cluster point of $\{x^k\}$, if any, is a critical point of f.*

Proof. Let $\hat{x}$ be a cluster point of $\{x^k\}$, and let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ converging to $\hat{x}$. Since $w^k \in \partial h(x^k)$, it follows that $\{w^{k_j}\}$ is also bounded, and we can suppose that $\{w^{k_j}\}$ converges to a point $\hat{w}$ (one can extract other subsequences if necessary). From (5), together with Proposition 3, we have

$$0 \in \partial_F \left(g(\cdot) - \langle w^{k_j}, \cdot - x^{k_j} \rangle + \frac{1}{2\lambda_{k_j}} q^2(x^{k_j}, \cdot) \right) (x^{k_{j+1}}).$$

Using Proposition 6, with $f_1(x) = g(x) - \langle w^{k_j}, x - x^{k_j} \rangle$, $f_2(x) = \frac{1}{2\lambda_{k_j}} q^2(x^{k_j}, x)$ and $\gamma = \delta = \frac{1}{k_j}$, we can write

$$0 \in \partial_F (f_1 + f_2)(x^{k_{j+1}}) \subset A + \frac{1}{k_{j+1}} \overline{B}(0, 1), \tag{11}$$

where $A = \{\partial_F f_1(a_1^{k_{j+1}}) + \partial_F f_2(a_2^{k_{j+1}}) \ : a_i^{k_{j+1}} \in \overline{B}(x^{k_{j+1}}, \frac{1}{k_{j+1}}), \quad |f_i(a_i^{k_{j+1}}) - f_i(x^{k_{j+1}})| < \frac{1}{k_{j+1}}, \ i = 1, 2\}$. Note that, $\partial_F f_1(a_1^{k_{j+1}}) = \partial g(a_1^{k_{j+1}}) - w^{k_{j+1}}$, and

$$\partial_F f_2(a_2^{k_{j+1}}) \subset \partial_L f_2(a_2^{k_{j+1}}) \subset \frac{1}{\lambda_{k_{j+1}}} q(x^{k_j}, a_2^{k_{j+1}}) \partial_L (q(x^{k_j}, \cdot))(a_2^{k_{j+1}}),$$

where the inclusions come from Proposition 1 and 5, respectively. Using these facts in (11), we obtain

$$w^{k_j} \in \partial g(a_1^{k_{j+1}}) + \frac{q(x^{k_j}, a_2^{k_{j+1}})}{\lambda_{k_j}} \partial_L(q(x^{k_j}, \cdot))(a_2^{k_{j+1}}) + \frac{1}{k_{j+1}} \overline{B}(0,1).$$

Hence, there exist subsequences $\xi^{k_{j+1}} \in \partial g(a_1^{k_{j+1}})$, $\eta^{k_{j+1}} \in \partial_L(q(x^{k_j}, \cdot))(a_2^{k_{j+1}})$ and $u^{k_{j+1}} \in \overline{B}(0,1)$ such that

$$w^{k_j} = \xi^{k_{j+1}} + \frac{q(x^{k_j}, a_2^{k_{j+1}})}{\lambda_{k_j}} \eta^{k_{j+1}} + \frac{u^{k_{j+1}}}{k_{j+1}}, \tag{12}$$

where $\{\eta^{k_j}\}$ and $\{u^{k_j}\}$ are bounded sequences, namely, $||\eta^{k_{j+1}}|| \leq M$ and $||u^{k_{j+1}}|| \leq 1$, for some $M > 0$ and for all $j \in \mathbb{N}$.

We will prove next that $\{\xi^{k_j}\}$ converges to $\hat{w}$. From (12), together with triangular inequality, we get

$$0 \leq ||w^{k_j} - \xi^{k_{j+1}}|| \leq \frac{M}{\lambda_{k_j}} q(x^{k_j}, a_2^{k_{j+1}}) + \frac{1}{k_{j+1}}. \tag{13}$$

Now, applying the triangular inequality for a quasi distance, one has

$$\begin{aligned} q(x^{k_j}, a_2^{k_j+1}) &\leq q(x^{k_j}, x^{k_j+1}) + q(x^{k_j+1}, a_2^{k_j+1}) \\ &\leq \beta||x^{k_j} - x^{k_j+1}|| + \beta||x^{k_j+1} - a_2^{k_j+1}|| \\ &\leq \beta||x^{k_j} - x^{k_j+1}|| + \frac{\beta}{k_j+1}, \end{aligned} \tag{14}$$

using Condition 1 in the second inequality and the fact that $a_2^{k_j+1} \in \overline{B}(x^{k_j+1}, \frac{1}{k_j+1})$ in the third. Combining (13) and (14), having in mind that $\{\lambda_k\}$ is bounded, and taking $j \to +\infty$, we obtain

$$\lim_{j \to +\infty} w^{k_j} = \lim_{j \to +\infty} \xi^{k_{j+1}} = \hat{w}.$$

Since $a_1^{k_{j+1}} \in \overline{B}(x^{k_{j+1}}, \frac{1}{k_{j+1}})$, we have $\lim_{j\to+\infty} a_1^{k_{j+1}} = \lim_{j\to+\infty} x^{k_{j+1}} = \hat{x}$. Therefore, using that $\xi^{k_{j+1}} \in \partial g(a_1^{k_{j+1}})$ and $w^{k_j} \in \partial h(x^{k_j})$, which, together with $g, h \in \Gamma_0(\mathbb{R}^n)$, leads to $\hat{w} \in \partial h(\hat{x}) \cap \partial g(\hat{x})$, that is to say, $\hat{x}$ is a critical point of f. □

4. Inexact Version

The proximal method is more practical if we can get approximate solutions for the subproblems. This consideration gives rise to inexact versions of the proximal point algorithm. As mentioned in the introduction, there is a huge literature related to the subject of inexact proximal point method, but a few in the context of DC optimization. Sun et al. [36] proposed an inexact version of the proximal point algorithm for DC functions based on the definition of $\epsilon-$subdifferential. Souza and Oliveira [40] presented an inexact version of the proximal point algorithm on Hadamard manifolds, in which, each subproblem of the algorithm is computed approximately. In this section, we present the following inexact version of the Algorithm GPLA.

Algorithm 2 (IGPLA)

Step 1: Given an initial point $x^0 \in \operatorname{int}(\operatorname{dom}(h))$, $\theta > 0$, $\sigma \in [0,1)$ and a bounded sequence of positive numbers $\{\lambda_k\}$ such that $\liminf_k \lambda_k > 0$.

Step 2: Calculate

$$w^k \in \partial h(x^k). \tag{15}$$

Step 3: Compute $x^{k+1} \in \mathbb{R}^n$ such that:

(16)

$$g(x) - g(x^{k+1}) - \langle w^k, x - x^{k+1}\rangle \geq \frac{(1-\sigma)}{2\lambda_k}\left[q^2(x^k, x^{k+1}) - q^2(x^k, x)\right] \quad \forall x \in \mathbb{R}^n,$$

with

$$||\xi^{k+1} - w^k|| \leq \theta q(x^k, x^{k+1})||\eta^{k+1}||, \tag{17}$$

where

$$\xi^{k+1} \in \partial g(x^{k+1}) \text{ and } \eta^{k+1} \in \partial_L(q(x^k, \cdot))(x^{k+1}). \tag{18}$$

If $x^{k+1} = x^k$, stop. Otherwise, set $k := k+1$ and return to Step 2.

Remark 4. *Note that the exact Algorithm GPLA is a specific case of Algorithm IGPLA (it holds by taking $\sigma = 0$). The algorithm is an inexact version of Algorithm GPLA, because* (16) *implies*

$$g(x^{k+1}) - \langle w^k, x^{k+1} - x^k\rangle + \frac{1}{2\lambda_k} q^2(x^k, x^{k+1}) \leq g(x) - \langle w^k, x - x^k\rangle + \frac{1}{2\lambda_k} q^2(x^k, x) + \epsilon_k,$$

for all $x \in \mathbb{R}^n$, where $\epsilon_k = \frac{\sigma}{2\lambda_k} q^2(x^k, x^{k+1}) \geq 0$. When $h \equiv 0$ and q is the Euclidean norm, Algorithm IGPLA coincides, in the Euclidean setting, with the algorithm proposed by Zaslavki [13] *for a particular choice of the sequence $\{\epsilon_k\}$.*

Next, we prove that Algorithm IGPLA has the same convergence properties relative to the Algorithm GPLA.

Theorem 4. *The sequence $\{x^k\}$ generated by algorithm IGPLA satisfies:*

1. *either the algorithm stops at a critical point;*
2. *or f decreases strictly, i.e., $f(x^{k+1}) < f(x^k) \quad \forall k \geq 0$.*

Proof. If $x^{k+1} = x^k$, it follows from (17) that $\xi^{k+1} = w^k$. Hence, $w^k \in \partial h(x^k) \cap \partial g(x^k)$, because $w^k \in \partial h(x^k)$ and $\xi^{k+1} \in \partial g(x^{k+1})$, from (15) and (18) respectively. Therefore, x^k is a critical point of f. Now, suppose that $x^{k+1} \neq x^k$. From (15) and (16), we have

$$h(x^{k+1}) \geq h(x^k) + \langle w^k, x^{k+1} - x^k\rangle,$$

and

$$g(x^k) \geq g(x^{k+1}) - \langle w^k, x^{k+1} - x^k\rangle + \frac{(1-\sigma)}{2\lambda_k} q^2(x^k, x^{k+1}),$$

respectively. Adding last two inequalities, we obtain

$$f(x^k) \geq f(x^{k+1}) + \frac{(1-\sigma)}{2\lambda_k} q^2(x^k, x^{k+1}), \tag{19}$$

which leads to $f(x^{k+1}) < f(x^k)$. □

Corollary 5. *Let $\{x^k\}$ be the sequence generated by algorithm IGPLA. Then the sequence $\{f(x^k)\}$ is convergent. If f is a continuous function and $\{x^k\}$ is bounded, then $\lim\limits_{k\to\infty} f(x^k) = f(\overline{x})$, for any cluster point $\overline{x}$ of $\{x^k\}$.*

Proof. The proof uses exactly the same argument as the one used to prove Corollary 2. □

Proposition 12. *Consider $\{x^k\}$ generated by algorithm IGPLA. Then $\lim\limits_{k\to+\infty} q(x^k, x^{k+1}) = 0$.*

Proof. From (19), we obtain

$$\frac{(1-\sigma)}{2\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1}),$$

and therefore

$$\frac{(1-\sigma)}{2} \sum_{k=0}^{n-1} \frac{1}{\lambda_k} q^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

Since f is bounded from below and $\liminf_k \lambda_k > 0$, this implies, if k goes to $+\infty$, that $\sum_{k=0}^{\infty} q^2(x^k, x^{k+1}) < \infty$. In particular, we have $\lim_{k\to+\infty} q(x^k, x^{k+1}) = 0$. □

Theorem 6. *Consider $\{x^k\}$ generated by algorithm IGPLA. Then, every cluster point of $\{x^k\}$, if any, is a critical point of f.*

Proof. Let $\hat{x}$ be a cluster point of $\{x^k\}$, and let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ converging to $\hat{x}$. Since $w^{k_j} \in \partial h(x^{k_j})$ and $\{x^{k_j}\}$ is bounded, it follows that $\{w^{k_j}\}$ is also bounded, and we can suppose that $\{w^{k_j}\}$ converges to a point $\hat{w}$ (one can extract other subsequences if necessary). Note that $\{\eta^{k_j}\}$ is bounded, because $\eta^{k_{j+1}} \in \partial_L(q(x^{k_j}, \cdot))(x^{k_{j+1}})$ and $\{x^{k_j}\}$ is bounded. Hence, combining (17) with Proposition 12, we have

$$\lim_{j\to+\infty} \xi^{k_{j+1}} = \lim_{j\to+\infty} w^{k_j} = \hat{w}. \tag{20}$$

On the other hand, it follows from definition of the algorithm that $w^{k_j} \in \partial h(x^{k_j})$ and $\xi^{k_j+1} \in \partial g(x^{k_j+1})$. Thus, letting j goes to $+\infty$ in the last two inclusions, and considering (20), we have $\hat{w} \in \partial h(\hat{x}) \cap \partial g(\hat{x})$, because h and g are lower semicontinuous and convex functions. This is to say, $\hat{x}$ is a critical point of f, and the proof is complete. □

Remark 5. *Dealing with descent methods for convex functions, we can expect that the algorithm will provide globally convergent sequences (i.e., convergence of the whole sequence). When the functions under consideration* *are not convex (or quasiconvex), the method may provide sequences that exhibit highly oscillatory behaviors, and partial convergence results are obtained. As already mentioned in the introduction, the Kurdyka-Lojasiewicz inequality has been successfully used to analyze various types of asymptotic behavior, in particular, proximal point method. It is worth mentioning that algorithm IGPLA (in particular algorithm GPLA) converges globally when* $f(x) = g(x) - h(x)$ *is a proper continuous and bounded from below function, which satisfies the Kurdyka-Lojasiewicz inequality at some cluster point* $\hat{x}$ *of* $\{x^k\}$ *(see the definition of the Kurdyka-Lojasiewicz inequality and other references about this subject in Attouch et al.* [53]*) with* $g, h \in \Gamma_0(\mathbb{R}^n)$*, and* $h \in \mathcal{C}^{1,1}$ *(which means that* h *is continuously differentiable with a Lipschitz gradient). In this context, if* $\{x^k\}$ *is bounded, then* $\{x^k\}$ *converges to* $\hat{x}$*. It is straightforward to check the above claim applying* [53, Theorem 2.9], *together with the following facts:*

(1) *Since* $\{\lambda_k\}$ *is bounded and* $\liminf_k \lambda_k > 0$*, suppose that* $0 < c_1 \leq \lambda_k \leq c_2$*, for all* k*. Hence, for each* k*, combining* (19) *with Condition 1, we can write*

$$f(x^{k+1}) + a||x^{k+1} - x^k||^2 \leq f(x^k),$$

where $a = \dfrac{(1-\sigma)\alpha}{2c_1} > 0$*;*

(2) *From* (17)*, for each* k*, we have*

$$||z^{k+1}|| \leq \theta q(x^k, x^{k+1})||\eta^{k+1}||, \tag{21}$$

where $z^{k+1} = \xi^{k+1} - \nabla h(x^k)$*, with* $\xi^{k+1} \in \partial g(x^{k+1})$*, and* $\eta^{k+1} \in \partial_L(q(x^k, \cdot))(x^{k+1})$*. Therefore,* $z^{k+1} \in \partial_L f(x^{k+1})$*, because* $f(x) = g(x) - h(x)$*, and* $\{\eta^k\}$ *is bounded, since* $\{x^k\}$ *is bounded and* $\eta^{k+1} \in \partial_L(q(x^k, \cdot))(x^{k+1})$*. Taking* $M > 0$ *such that* $||\eta^k|| \leq M$*, for all* k*, together with* (21)*, we obtain*

$$||z^{k+1}|| \leq b||x^{k+1} - x^k||,$$

where $b = \theta\beta M > 0$*;*

(3) *As we supposed,* $\hat{x}$ *is a cluster point of* $\{x^k\}$*. Then there exists a subsequence* $\{x^{k_j}\}$ *converging to* $\hat{x}$*. It follows from the continuity of* f *that* $\{f(x^{k_j})\}$ *converges to* $f(\hat{x})$*. From Corollary 5, we have that* $\{f(x^k)\}$ *converges to* $f(\hat{x})$*.*

5. Application: The optimal size of the firm problem

One of the main topic in Economic and Management Sciences is to determine the optimal size of an organization. This is a difficult problem, both for conceptual and technical reasons. This optimal size can refer to the quantity of the final good produced, the range of different final goods that the multi product firm produces, the number and quality of workers of different types employed, the amount of means used in the production process, as well as the number of intermediate stages in the production process and their different locations in different countries in the globalization process. A huge literature exists and a lot of aspects must be examined.

In this last section we will consider, in a dynamic setting, the simplest static model of the firm we can imagine (the textbook case). Its size refers to the production level, i.e, the number of units of the final good the firm produces. We consider the most realistic but difficult case where this firm is supposed to exhibit increasing returns in the short run, when execution costs of production are concave. The more traditional decreasing returns to scale case is far less difficult, when costs of production are convex. Then, using the recent variational rationality approach (Soubeyran [20, 21, 22]), we will determine the long run optimal size of this firm, when it can, each period, hire, fire and keep again workers. This offers an original and dynamic theory of the limit of the firm.

5.1. **A simple model of the firm with increasing returns in the short run.**

5.1.1. *An example of "to be increased" and "to be decreased" payoffs.* To better see how DC optimization works in applications, let us examine the simplest case we can imagine, which can be generalized to the multidimensional setting. Different variants of this example can be found in Soubeyran [20, 21], Bento and Soubeyran [14, 15] and Bao et al. [47]. But none of them examine the very important case of increasing returns, which is the realistic case for production costs, as we do here, as an application of the proximal point method for DC optimization. Consider a hierarchical firm including an entrepreneur, a profile of workers, and a succession of periods where the entrepreneur can hire, fire or keep working workers in a changing environment. Each period, the entrepreneur chooses how much to produce of the same final good (of a given quality) and sells each unit of this good at the same and fixed price $p > 0$. In the current period, the firm produces $x \in \mathbb{R}_+$ units of a final good and employs $l(x) \in \mathbb{R}_+$ workers. In this simple model the size of the firm refers to x. For simplification, each worker is asked to produce one unit of the final good. Then, $l(x) = x$. The current profit of the entrepreneur, $\pi(x) = r(x) - c(x)$, is the difference between the revenue of the firm $r(x) = px \geq 0$ and the cost of production $c(x) \geq 0$. To produce one unit of the final good, each employed worker must use a given bundle of individual means (tools and ingredients) and a fixed collective mean (say, some given piece of land, a given infrastructure). The entrepreneur rents the durable tools and buys the non durable ingredients. Let $\overline{\pi} = \sup \{\pi(y), y \in X\} < +\infty$ be the highest profit the entrepreneur can hope to achieve. Then, $f(x) = \overline{\pi} - \pi(x) \geq 0$ is the current unrealized profit he can hope to carry out in the current period, or later. The profit function $\pi(\cdot)$ is a "to be increased" payoff, while the unrealized profit function $f(\cdot)$ is a "to be decreased" payoff.

In the mathematical part of the paper, the objective function $f(x) = g(x) - h(x)$ is the difference between two convex functions, $g(x)$ and $h(x)$.

In our behavioral example, $f(x)$ represents the unrealized profit the entrepreneur can hope to achieve, i.e, $f(x) = \overline{\pi} - \pi(x) = \overline{\pi} + c(x) - r(x)$, where $g(x) = \overline{\pi} + c(x)$ and $h(x) = r(x)$. Then, the cost and the revenue functions $r(\cdot)$ and $c(\cdot)$ must be concave to fit with the mathematical part of the paper.

Clearly, in a perfectly competitive market where the price p of the final good is a given, the revenue function $r(\cdot) : x \in \mathbb{R}_+ \longmapsto r(x) = px$ is linear, hence concave with respect to the production level x. What we have to make clear is why a cost function $c(\cdot)$ is usually concave in the short run. But, to escape to mathematical difficulties, textbooks in Economics focus the attention on the less usual case of convex costs of production in the short run. Costs of production are concave when the technology of the firm exhibits increasing returns, coming from economies of scale, economies of specialization, learning by doing several times the same thing, limited capacities, lack of time to be able to change fixed costs in the short run, which become variable costs in the long run. In our standard model of the firm, costs of production $c(x) = wx + hx + K$ are the sum of three different costs, where, i) $w > 0$ is the given wage paid to each employed worker, ii) $h > 0$ is the price paid to suppliers to acquire each bundle of means used by each employed worker to produce one unit of the final good, and, iii) $K > 0$ is the cost to rent a durable, fixed, collective and indivisible mean.

This cost of production exhibits increasing returns to scale because, in the current period, before production takes place, the fixed costs $K > 0$ must be paid even if, later, no worker are required to work, i.e, $c(0) = K > 0$. This implies that the unit cost of production $c(x)/x = w + h + K/x$ decreases when the production level x increases. The cost of production will be strictly concave if, for example, the price $h = h(x)$ of each bundle of means used by each employed worker decreases with the number x of bundles of means the entrepreneur must buy to produce x units of the final good (when suppliers offer discounts).

5.2. **The variational rationality approach: the simplest formulation.**

5.2.1. *Stay/stability and change dynamics.* The (VR) variational rationality approach (Soubeyran [20, 21]) modelizes and unifies a lot of different models of stay and change dynamics which appeared in Behavioral Sciences (Economics, Management Sciences, Psychology, Sociology, Political Sciences, Decision theory, Game theory, Artificial Intelligence, ...). Stays refer to exploitation phases, temporary repetitions of the same action, temporary habits, routines, rules and norms ...., while changes refer to exploration phases, learning and innovations processes, forming and breaking habits and routines, changing doings (actions), havings and beings..... This dynamical approach considers entities (an agent, an organization or several interacting agents), which are, at the beginning of the story, in an undesirable initial position, and are unable to reach immediately a final desired position. The goal of this approach is to examine the transition problem: how such entities can find, build and use an acceptable and feasible transition which is able to overcome a lot of intermediate obstacles, difficulties and resistance to change, with not too much intermediate sacrifices and enough intermediate satisfactions to sustain motivation to change and persevere until reaching the final desired position. This (VR) approach admits a lot of variants, based on the same short list of general principles and concepts. The four main concepts refer to changes and stays, worthwhile changes and stays, worthwhile transitions and variational traps, worthwhile to approach and reach but not worthwhile to leave. A stay and change dynamic refers to a succession of periods, where $k+1$ is the current period and k is the past period, where $x = x^k \in X$ can be a past action (doing), having or being and $y = x^{k+1} \in X$ can be a current action (doing), having or being. A single change from $x = x^k \in X$ to $y = x^{k+1} \in X$ is $x \curvearrowright y, y \neq x$. A single stay at x is $x \curvearrowright y, y = x$.

Let us give, starting from our previous example, the simplest prototype of the (VR) variational rationality approach, to finally show how, at the end of a worthwhile transition, a firm can achieve an optimal size.

5.2.2. *Worthwhile changes.* The (VR) approach starts with the following broad definition of a worthwhile change: a change is worthwhile if motivation to change rather than to stay is "high enough" with respect to resistance to change rather than to stay. This definition allows a lot of variants, as much variants as the definitions of motivation (more than one hundred theories of motivations exist in Psychology), resistance (which includes a lot of different aspects) and "high enough" (see Soubeyran [20, 21]). Let us give a very simple formulation of the worthwhile to change concept.

In the previous example a change refers to a move from having produced $x \in X = \mathbb{R}_+$ units of a final good in the previous period to produce $y \in \mathbb{R}_+$ units of this final good in the current period. A stay is a particular move, from having produced a given quantity $x = x^k$ of the final good in the previous period to produce again the same quantity $y = x^{k+1} = x^k$ of this final good in the current period. The previous and current "to be increased" payoffs of the entrepreneur are the profit $\pi(x)$ and $\pi(y)$. His previous and current "to be decreased" payoffs are his unrealized profits $f(x) = \overline{\pi} - \pi(x) \geq 0$ and $f(y) = \overline{\pi} - \pi(y) \geq 0$.

Advantages to change from x to y, if they exist, represent the difference in profits or unrealized profits, $A(x,y) = \pi(y) - \pi(x) = f(x) - f(y) \geq 0$.

Inconveniences to change from x to y refer to the difference $I(x,y) = C(x,y) - C(x,x) \geq 0$.

$C(x,y) \geq 0$ modelizes the costs of being able to change from x to y. In the present model $C(x,y)$ modelizes costs of hiring, firing and keeping working workers to be able to move from producing x units of the final good, to produce y units of the final good, where y can be higher, lower or the same than x. Costs of hiring $y - x > 0$ workers are $C(x,y) = \rho^+(y-x)$, where

$\rho^+ > 0$ is the cost of hiring one worker. Costs of firing $x - y > 0$ workers are $C(x, y) = \rho^-(x - y)$, where $\rho^- > 0$ is the cost of firing one worker. Costs of keeping working $y = x$ workers are $C(x, x) = \rho^= x$, where $\rho^= \geq 0$ is the cost of keeping working one period more one worker. For simplification (this will require a too long discussion), we will suppose that $\rho^= = 0$. Then, $C(x, x) = 0$, and inconveniences to change are

$$I(x,y) = C(x,y) = \left\{ \begin{array}{l} \rho^+(y-x) \text{ if } y \geq x \\ \rho^-(x-y) \text{ if } y \leq x \end{array} \right\} \geq 0.$$

Motivation to change $M(x, y) = U[A(x, y)]$ is the utility $U[A]$ of advantages to change $A = A(x, y) \geq 0$.

Resistance to change $R(x, y) = D[I(x, y)]$ is the disutility $D[I]$ of inconveniences to change $I = I(x, y) \geq 0$, where the utility function $U[\cdot] : A \in \mathbb{R}_+ \longmapsto U[A] \in \mathbb{R}_+$ and the disutility function $D[\cdot] : I \in R_+ \longmapsto D[I] \in \mathbb{R}_+$ are strictly increasing and zero at zero.

A worthwhile change from x to y is such that motivation to change $M(x, y) \in \mathbb{R}_+$ from x to y is higher than resistance to change $R(x, y)$ from x to y, up to a chosen worthwhile to change satisfaction ratio $\xi > 0$, i.e., such that $M(x, y) \geq \xi R(x, y)$.

In the example, the utility $U[A]$ of advantages to change and the disutility $D[I]$ of inconveniences to change are linear-quadratic, i.e., $M = U[A] = A$, $R = D[I] = I^2$ (see Soubeyran [20, 21] for more general cases). In this context, a change $x \curvearrowright y$ from producing again the quantity x of the final good to produce a different quantity y of this final good is worthwhile if advantages to change are high enough with respect to resistances to change, i.e., $A(x, y) = \pi(y) - \pi(x) = f(x) - f(y) \geq \xi R(x, y) = \xi C(x, y)^2$, where $C(x, x) = 0$. What is "high enough" is defined by the size of $\xi > 0$.

5.2.3. *Worthwhile transitions.* A transition is a succession of single stays and changes $x^0 \curvearrowright x^1 \curvearrowright ....x^k \curvearrowright x^{k+1} \curvearrowright ....$ where $x^{k+1} \neq x^k$ or $x^{k+1} = x^k$ for each $k \in \mathbb{N}$.

A worthwhile transition is a transition such that each stay or change is worthwhile, i.e., $x^{k+1} \in W_{\xi_{k+1}}(x^k)$, $k \in \mathbb{N}$, that is,

$$\begin{aligned} A(x^k, x^{k+1}) &= \pi(x^{k+1}) - \pi(x^k) \\ &= f(x^k) - f(x^{k+1}) \\ &\geq \xi_{k+1} R(x^k, x^{k+1}) \\ &= \xi_{k+1} C(x^k, x^{k+1})^2, \ k \in \mathbb{N}. \end{aligned}$$

5.2.4. *Ends as variational traps.* A (strong) variational trap $x^* \in X$ is both, i) an aspiration point $x^* \in W_{\xi_{k+1}}(x^k)$, $k \in \mathbb{N}$, worthwhile to reach from any position of the transition, ii) a stationary trap $W_{\xi_*}(x^*) = \{x^*\}$, where it is not worthwhile to move to any other position $y \neq x^*$, given that the worthwhile to change ratio tends to a limit, $\xi_{k+1} \to \xi_* > 0, k \to +\infty$, and finally iii) worthwhile to approach, i.e., which converges to the aspiration point. More explicitly, x^* is a variational trap if,

i) $A(x^k, x^*) = \pi(x^*) - \pi(x^k) = f(x^k) - f(x^*) \geq \xi_{k+1} R(x^k, x^*) = \xi_{k+1} C(x^k, x^*)^2$, $k \in \mathbb{N}$;
ii) $A(x^*, y) = \pi(y) - \pi(x^*) = f(x^*) - f(y) < \xi_* R(x^*, y) = \xi_* C(x^*, y)^2$, for all $y \neq x^*$;
iii) It is a limit point of the worthwhile transition, i.e., $x^k \to x^*, k \to +\infty$.

A weak variational trap does not require to be an aspiration point.

5.3. **Proximal algorithms as worthwhile transitions.** To show how an exact or inexact proximal algorithm can be seen as a leading example of a worthwhile transition, the present paper uses a specific formulation of the (VR) approach, where the utility of advantages to change and the disutility of inconveniences to change are linear quadratic, i.e., $M = U[A] = A$ and $R = D[I] = I^2 = C^2$, where $C(x, y) = q(x, y) \geq 0$ is a quasi distance (see Moreno et al. [16] for this linear quadratic case, and Bento and Soubeyran [14, 15] for more general cases).

5.3.1. *The proximal formulation of a worthwhile change.* In a linear quadratic setting, motivation and resistance to change are $M(x, y) = A(x, y) = \pi(y) - \pi(x) = f(x) - f(y)$ and $R(x, y) = q(x, y)^2$. These simplifications allow to define,

1) a proximal "to be increased" payoff $P_\xi(y/x) = \pi(y) - \xi R(x, y)$, which is the difference between the current "to be increased" payoff $\pi(y)$ and the weighted current resistance to change $R(x, y)$, where the weight $\xi > 0$ balances the importance of the current "to be increased" payoff and the current resistance to change.

2) a proximal "to be decreased" payoff $Q_\xi(y/x) = f(y) + \xi R(x, y)$, which is the sum of the current "to be decreased" payoff $f(y)$ and the weighted current resistance to change $R(x, y)$.

Then, a change $x \curvearrowright y \in W_\xi(x)$ is worthwhile if moving from x to y, the proximal "to be increased" payoff increases, $P_\xi(y/x) \geq P_\xi(x/x)$, and the proximal "to be decreased" payoff decreases, $Q_\xi(y/x) \leq Q_\xi(x/x)$. This comes from the following equivalences

$$\begin{aligned} y \in W_\xi(x) &\iff M(x, y) \geq \xi R(x, y) \\ &\iff \pi(y) - \pi(x) = f(x) - f(y) \geq \xi R(x, y) \\ &\iff P_\xi(y/x) \geq P_\xi(x/x) \\ &\iff Q_\xi(y/x) \leq Q_\xi(x/x). \end{aligned}$$

5.3.2. *Exact and inexact proximal algorithms as examples of worthwhile transitions.* A transition is a succession of single stays and changes $x^0 \curvearrowright x^1 \curvearrowright \ldots . x^k \curvearrowright x^{k+1} \curvearrowright \ldots .$ where $x^{k+1} \neq x^k$ or $x^{k+1} = x^k$ for each $k \in \mathbb{N}$.

A worthwhile transition is a transition such that each stay or change is worthwhile, i.e., in term of proximal payoffs to change,

$$x^{k+1} \in W_{\xi_{k+1}}(x^k) = \left\{ \begin{array}{c} y \in X, \text{ such that} \\ P_{\xi_{k+1}}(y/x^k) \geq P_{\xi_{k+1}}(x^k/x^k), \text{ i.e.,} \\ \pi(y) - \xi_{k+1} R(x_n, y) \geq \pi(x^k), \text{ or,} \\ Q_{\xi_{k+1}}(y/x^k) \leq Q_{\xi_{k+1}}(x^k/x^k), \text{i.e.,} \\ f(y) + \xi_{k+1} R(x^k, y) \leq f(x^k) \end{array} \right\},$$

where each $\xi_{k+1} > 0$, $k \in \mathbb{N}$ can be chosen and $R(x^k, y) = q(x^k, y)^2$.

In the context of this paper,

$$x^{k+1} \in W_{\xi_{k+1}}(x^k) \Longleftrightarrow f(x^{k+1}) + \xi_{k+1} q(x^k, x^{k+1})^2 \leq f(x^k),$$

where $\xi_{k+1} = 1/2\lambda_k > 0$.

A worthwhile change is exact if $x^{k+1} \in \arg\max \left\{ P_{\xi_{k+1}}(y/x_n),\ y \in X \right\}$.

An inexact worthwhile change is any worthwhile change "close enough" to an exact worthwhile change, where the term "close enough" can have several different interpretations, depending of chosen reference points and frames. In this paper "close enough" is given by conditions (16), (17), given (15) and (18). The explicit justifications follow Bento and Soubeyran [14].

5.4. **Surrogate proximal algorithms as worthwhile transitions.** Usually the entrepreneur does not know the whole profit function $\pi(\cdot)$. Then, he must perform, each current period $k+1$, an approximate evaluation of this function $\widetilde{\pi}(\cdot/x)$, where $x = x^k$. This requires to consider a more complex formulation of the (VR) approach, where past experience and current evaluations $\widetilde{\pi}(\cdot/x)$ of the payoff functions $\pi(\cdot)$ are included in the worthwhile to change process (see Soubeyran [22]). In the present paper, we will discard the role of past experience to focus our attention on the current evaluation process, when the entrepreneur knows from the very beginning the whole revenue function $r(\cdot) = -g(\cdot)$, but does not know very well the execution cost function $c(\cdot)$. Then, he needs, each period, to make an approximate evaluation of the execution cost function $c(\cdot)$, in term of a simple function $\widetilde{c}(\cdot/x^k)$, which over-estimates globally this cost function $c(\cdot) = -h(\cdot)$, , i.e., $\widetilde{c}(y/x^k) \geq c(y)$, for all $y \in X$ with $\widetilde{c}(x^k/x^k) = c(x^k)$. Then, the surrogate evaluation function $\widetilde{\pi}(./x) : y \in X \longmapsto \widetilde{\pi}(y/x) = r(x) - \widetilde{c}(y/x^k)$ under-estimates the "to be increased" profit function $\pi(\cdot) = r(\cdot) - c(\cdot)$, because $\widetilde{\pi}(y/x) \leq \pi(y)$ for all $y \in X$ and $\widetilde{\pi}(x/x) = \pi(x)$.

Similarly, the surrogate evaluation function $\widetilde{f}(\cdot/x) : y \in X \longmapsto \widetilde{f}(y/x) = g(y) - \widetilde{h}(y/x)$ over-estimates the "to be decreased" profit function $f(\cdot) = g(\cdot) - h(\cdot)$, because $\widetilde{f}(y/x) \geq f(y)$, for all $y \in X$, with $\widetilde{f}(x/x) = f(x)$, where $\widetilde{h}(y/x) \leq h(y)$, for all $y \in X$, with $\widetilde{h}(x/x) = h(x)$.

To fit with the mathematical part of the paper, we will suppose that, in the current period $k+1$, the entrepreneur knows the resistance to change function $R(x, y)$. Then, given this knowledge structure, where, each period, the entrepreneur is allowed to make an under-estimation of his profit function $\widetilde{\pi}(\cdot/x)$, a change $x = x^k \curvearrowright y$ is worthwhile if $\widetilde{A}(x, y) = \widetilde{\pi}(y/x) - \pi(x) = f(x) - \widetilde{f}(y/x) \geq \xi R(x, y)$.

The proximal version of this worthwhile to change condition is

$$\widetilde{P}_\xi(y/x) = \widetilde{\pi}(y/x) - \xi R(x, y) \geq \pi(x) = \widetilde{P}_\xi(x/x),$$

or

$$\widetilde{Q}_\xi(y/x) = \widetilde{f}(y/x) + \xi R(x, y) \leq f(x) = \widetilde{Q}_\xi(x/x).$$

This behavioral evaluation process fits with the mathematical part of the paper, which uses a convex-concave procedure (see Yuille and Rangarajan [54]) in the context of DC programming (see also Horst and Thaoi [55]).

5.5. **Ends.**

5.5.1. *When critical points are variational traps.* A weak variational trap is both a limit point of a worthwhile transition, and a stationary trap not worthwhile to leave. This modelizes the approach, and the end of a worthwhile stay and change process. Usually, a critical point is not a stationary trap. Then, the last question of this paper is: when critical points of the exact and inexact proximal algorithm are variational traps?

Definition 5. *A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be weakly convex if there exists $\rho > 0$ such that for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) + \rho\lambda(1 - \lambda)||x - y||^2. \tag{22}$$

The function f is said to be locally weakly convex at x if there exists $\epsilon > 0$ such that f is weakly convex on $B(x, \epsilon)$. It is locally weakly convex if it is locally weakly convex at every point of its domain.

Proposition 13. *Let f be a weakly convex function. If x^* is a critical point of f, then*

$$f(x^*) \leq f(y) + \frac{\rho}{\alpha^2} q^2(x^*, y) \quad \forall y \in \mathbb{R}^n, \tag{23}$$

where $\alpha > 0$ satisfies Condition 1.

Proof. This comes from [56, Proposition 4.8] and (3). □

Proposition 14. *Let f be a weakly convex function. If x^* is a critical point of f and $\lambda > \frac{\rho}{\alpha^2}$, then $W_\lambda(x^*) = \{x^*\}$.*

Proof. From (23) and $\lambda > \frac{\rho}{\alpha^2}$, we have

$$f(x^*) \leq f(y) + \frac{\rho}{\alpha^2} q^2(x^*, y) < f(y) + \lambda q^2(x^*, y) \quad \forall y \neq x^*.$$

The result follows from last inequality and definition of $W_\lambda(x)$. □

Remark 6. *It is obvious that convex functions are weakly convex. Moreover, $C^{1,1}$ functions and lower-C^2 functions are locally weakly convex and locally Lipschitz weakly convex, respectively. As mentioned in the preliminary section, $C^{1,1}$ functions and lower-C^2 functions are DC functions.*

5.5.2. *The optimal size of the firm.* In the example a variational trap $x^* \in X$ defines an optimal size of the firm where the entrepreneur hires and fires less and less workers to finally stops to hire and fire workers, when resistance to change wins motivation to change. This offers an original theory of the limit of the firm in term of the VR approach, where the entrepreneur optimizes at the end, and satisfies with not too much sacrifices during the transition.

6. Conclusions

We presented a generalized proximal linearized algorithm, which can be performed inexactly, for finding critical points of a DC function (difference of two convex functions). We also provided an application, in a dynamic setting, to determine the limit of the firm, when increasing returns matter in the short run. Future researches will examine the case where the concave revenue function is not well known, while the concave execution cost function is perfectly known to the entrepreneur. The case of a more changing environment can also be considered. Finally multiobjective DC programming must be examined to consider the limit of firms which produce different final products. This is the main realistic case.

Acknowledgments. The work of the first author was supported in part by CNPq-Ciências sem Fronteiras Grant 203360/2014-1. The work of the second author was partially supported by CNPq, Brazil.

References

[1] B. Martinet, Régularisation d'inéquations variationelles par approximations successives, Rev. Française d'Inform. Recherche Oper. 4 (1970) 154-159.
[2] J.J. Moreau, Proximité et dualité dans un espace Hilbertien, Bull. Soc. Math. France 93 (1965) 273-299.
[3] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control. Optim. 14 (1976) 877-898.
[4] A. Kaplan, R. Tichatschke, Proximal point methods and nonconvex optimization, J. Glob. Optim. 13 (1998) 389-406.
[5] W. Hare, C. Sagastizábal, Computing proximal points of nonconvex functions, Math. Program. 116 (2009) 221-258.
[6] G.C. Bento, O.P. Ferreira, P.R. Oliveira, Proximal point method for a special class of nonconvex functions on Hadamard manifolds, Optimization 64 (2015) 289-319.
[7] S. Pan, J.S. Chen, Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming, J. Glob. Optim. 39 (2007) 555-575.
[8] E.A. Papa Quiroz, P.R. Oliveira, An extension of proximal methods for quasiconvex minimization on the nonnegative orthant, European Journal of Operational Research, 216 (2012) 26-32.
[9] S.S. Souza, P.R. Oliveira, J.X. Cruz Neto, A. Soubeyran, A proximal method with separable Bregman distances for quasiconvex minimization over the nonnegative orthant, European Journal of Operational Research, 201 (2010) 365-376.
[10] R.S. Burachik, B.F. Svaiter, A relative error tolerance for a family of generalized proximal point methods, Math. Oper. Res. 26(4) (2001) 816-831.

[11] M.V. Solodov, B.F. Svaiter, Error bounds for proximal point subproblems and associated inexact proximal point algorithms, Math. Program. 88(2) (2000) 371-389.
[12] M.V. Solodov, B.F. Svaiter, A unified framework for some inexact proximal point algorithms, Numer. Funct. Anal. Optim. 22(7-8) (2001) 1013-1035.
[13] A. Zaslavski, Convergence of a proximal point method in the presence of computational errors in Hilbert spaces, SIAM J. Optim. 20(5) (2010) 2413-2421.
[14] G.C. Bento, A. Soubeyran, Generalized inexact proximal algorithms: Routines formation with resistance to change, following worthwhile changes, J. Optim. Theory App. 166 (2015) 172-187.
[15] G.C. Bento, A. Soubeyran, A generalized inexact proximal point method for nonsmooth functions that satisfies Kurdyka-Łojasiewicz inequality, Set-Valued Var. Anal. 23 (2015) 501-517.
[16] F.G. Moreno, P.R. Oliveira, A. Soubeyran, A proximal point algorithm with quasi distance. Application to habit's formation, Optimization 61 (2012) 1383-1403.
[17] G. Chen, M. Teboulle, Convergence analysis of proximal-like optimization algorithm using Bregman functions, SIAM J. Optim. 3 (1993) 538-543.
[18] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, Math. Oper. Res. 18 (1993) 202-226.
[19] K.C. Kiwiel, Proximal minimization methods with generalized Bregman functions, SIAM J. Control Optim. 35 (1997) 1142-1168.
[20] A. Soubeyran, Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors, GREQAM, Aix Marseillle University (2009) Preprint.
[21] A. Soubeyran, Variational rationality and the "unsatisfied man": routines and the course pursuit between aspirations, capabilities and beliefs, GREQAM, Aix Marseillle University (2010) Preprint.
[22] A. Soubeyran, Variational rationality, Part 1. A theory of worthwhile stay and change approach-avoidance transitions ending in traps, GREQAM, Aix Marseillle University (2015) Preprint.
[23] V. Brattka, Recursive quasi-metric spaces, Theoret. Comput. Sci. 305 (2003) 17-42.
[24] H.P.A. Künzi, H. Pajoohesh, M.P. Schellekens, Partial quasi-metrics, Theoret. Comput. Sci. 365 (2006) 237-246.
[25] S. Romaguera, M. Sanchis, Applications of utility functions defined on quasi-metric spaces, J. Math. Anal. Appl. 283 (2003) 219-235.
[26] J. Bolte, S. Sabach, M.Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program. 146 (2014) 459-494.
[27] K. Lange, D.R. Hunter, I. Yang, Optimization transfer using surrogate objective functions, J. Comput. Graph. Stat. 9 (2000) 1-20.
[28] M. Razaviyayn, M. Sanjabi, Z.-Q. Luo, A stochastic successive minimization method for nonsmooth nonconvex optimization, (2013) arXiv:1307.4457v2.
[29] J. Mairal, Incremental majorization-minimization optimization with application to large-scale machine learning, SIAM J. on Optim. 25(2) (2015) 829-855.
[30] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pur. Appl. Math. 57 (2004) 1413-1457.
[31] G. Gasso, A. Rakotomamonjy, S. Canu, Recovering sparse signals with non-convex penalties and DC programming, IEEE T. Signal Process. 57 (2009) 4686-4698.
[32] H. Erdogan, J.A. Fessler, Ordered subsets algorithms for transmission tomography, Phys. Med. Biol. 44 (1999) 2835-2851.
[33] S. Ahn, J.A. Fessler, D. Blatt, A.O. Hero, Convergent incremental optimization transfer algorithms: Application to tomography, IEEE T. Med. Imaging. 25 (2006) 283-296.
[34] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in Adv. NIPS. (2001)

[35] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, J. Mach. Learn. Res. 11 (2010) 19-60.
[36] W. Sun, R.J.B. Sampaio, M.A.B. Candido, Proximal point algorithm for minimization of DC Functions, Journal of Computational Mathematics 21 (2003) 451-462.
[37] P. Hartman, On functions representable as a difference of convex functions, Pac. J. Math. 9 (1959) 707-713.
[38] J.B. Hiriart-Urruty, Generalized differentiability, duality and optimization for problems dealing with diference of convex functions, Convexity and Duality in Optimization, Lecture Notes in Economics and Mathematical Systems 256 (1986) 37-70.
[39] A. Moudafi, P-E. Maingé, On the convergence of an approximate proximal method for d.c. functions, Journal of Computational Mathematics 24 (2006) 475-480.
[40] J.C.O. Souza, P.R. Oliveira, A proximal point algorithm for DC functions on Hadamard manifolds, J. Glob. Optim. (2015) doi 10.1007/s10898-015-0282-7
[41] L.D. Muu, T.D. Quoc, One step from DC optimization to DC mixed variational inequalities, Optimization 59 (2010) 63-76.
[42] M. Bačák, J.M. Borwein, On difference convexity of locally Lipschitz functions, Optimization 60 (2011) 961-978.
[43] D.T. Pham, E.B. Souad, Algorithms for solving a class of nonconvex optimization problems: methods of subgradient, Fermat Days 85: Mathematics for Optimization 129 (1986) 249-271.
[44] D.T. Pham, L.T.H. An, Convex analysis approach to DC programming: Theory, algorithms and applications, ACTA Math. Vietnam. 22 (1997) 289-355.
[45] D.T. Pham, L.T.H. An, F. Akoa, The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems, Annals of Operations Research 133 (2005) 23-46.
[46] E. Fernández Cara, C. Moreno, Critical point approximation through exact regularization, Mathematics of Computation 50 (1988) 139-153.
[47] T. Bao, B.S. Mordukhovich, A. Soubeyran, Variational Analysis in Psychological Modelling, J. Optim. Theory Appl. 164 (2015) 290-315.
[48] R.T. Rockafellar, R.J-B. Wets, Variational Analysis, Springer, Berlin, 1998.
[49] B.S. Mordukhovich, Y. Shao, Nonsmooth Sequential Analysis in Asplund Spaces, Transactions of the American Mathematical Society 348 (1996) 1235-1280.
[50] A. Yassine, N. Alaa, A. Lhilali Alaoui, Convergence of Toland's critical points for sequences of D.C. functions and application to the resolution of semilinear elliptic problems, Control and Cybernetics 30(4) (2001) 405-417.
[51] A. Lhilali Alaoui, Caractérisation des fonctions D.C., Ann. Sci. Math. Quebec 20(1) (1996) 1-13.
[52] D. Goldfarb, S. Ma, K. Scheinberg, Fast alternating linearization methods for minimizing the sum of two convex functions, Math. Program. 141 (2013) 349-382.
[53] H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Math. Program. 137(1-2) (2013) 91-129.
[54] A. Yuille, A. Rangarajan, The concave-convex procedure, Neural Computation 15(4) (2003) 915-936.
[55] R. Horst, N.V. Thoai, DC programming: overview, J. Optim. Theory App. 103(1) (1999) 1-43.
[56] J-P. Vial, Strong and Weak Convexity of Sets and Functions, Math. Oper. Res. 8 (1983) 231-259.

J.C. SOUZA, CEAD, Federal University of Piauí, Teresina, Brazil
E-mail address: joaocos.mat@ufpi.edu.br

P.R. OLIVEIRA, PESC-COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
E-mail address: poliveir@cos.ufrj.br

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