Global convergence of a proximal linearized algorithm for difference of convex functions

João Carlos O. Souza, Paulo Roberto Oliveira, Antoine Soubeyran

Global convergence of a proximal linearized algorithm for difference of convex functions

João Carlos O. Souza[1,2] ·
Paulo Roberto Oliveira[1] · **Antoine Soubeyran**[3]

Abstract A proximal linearized algorithm for minimizing difference of two convex functions is proposed. If the sequence generated by the algorithm is bounded it is proved that every cluster point is a critical point of the function under consideration, even if the auxiliary minimizations are performed inexactly at each iteration. Linear convergence of the sequence is established under suitable additional assumptions.

Keywords Global optimization · Proximal linearized algorithm · DC functions · Rate of convergence

---

The research of the J. C. Souza was supported in part by CNPq-Ciências sem Fronteiras Grant 203360/2014-1. The P. R. Oliveira was partially supported by CNPq, Brazil.

✉ João Carlos O. Souza
joaocos.mat@ufpi.edu.br

Paulo Roberto Oliveira
poliveir@cos.ufrj.br

Antoine Soubeyran
antoine.soubeyran@gmail.com

1 COPPE-PESC, Federal University of Rio de Janeiro, Postal Code 21945-970 Rio de Janeiro, RJ, Brazil

2 CEAD, Federal University of Piauí, Teresina, PI, Brazil

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

1 Introduction

It is well known that the class of Proximal Point Algorithm (PPA) 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 classical PPA was introduced into optimization literature by Martinet [1]. It is based on the notion of proximal mapping introduced earlier by Moreau [2]. The PPA was popularized by Rockafellar [3], who showed that, in the context of maximal monotone operators, 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 mild assumptions, even if the auxiliary minimizations are 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 $x^* \in \arg\min f(x)$. The algorithm is useful, however, only for convex problems, because the idea underlying the results is based on the monotonicity of subdifferential operators of convex functions. Therefore, PPA for nonmonotone operators (or nonconvex functions) has been investigated by many authors in different contexts; see, for instance [4–9] and references therein. As remarked in Rockafellar [3], for a proximal point method to be practical, it is also important that it should work with approximate solutions of the subproblems. Since then, there has been a growing interest in inexact versions of proximal methods and many works appeared, treating the problem under different perspectives; see [7–13] and references therein. Recently, Bento and Soubeyran [14] discussed how ”generalized” proximal point algorithms can be a nice tools to modelize the dynamics of human behaviours on the context of the “variational rationality approach”, Soubeyran [31].

Sun et al. [15] proposed a proximal point algorithm for minimization of a class of nonconvex functions called DC functions, i.e., difference of two convex functions, which use convex properties of the two convex functions separately. Moudafi and Maingé [16] proposed an alternative proof of the main result of [15]. It is also considered in [16] an interesting result in the case where the second component of the DC function is differentiable. Souza and Oliveira [17] extended this algorithm in the context of Hadamard manifolds considering inexact computations of each proximal iteration. DC functions defined on $\mathbb{R}^n$, briefly $\mathcal{DC}(\mathbb{R}^n)$, form an important class of functions, both from a theoretical point of view (see [18–21]) and for algorithmic purposes (see [15,22,23]). Thus, the interest in the theory of DC functions has much increased in the last years. 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 [24–26].

The aim of the paper is to study global convergence properties of a proximal linearized algorithm for minimizing a nonsmooth DC function. Our algorithm seems to be the first one which considers a convex linear approximation in each proximal subproblem in the context of DC functions.

The organization of this paper is as follows. In Sect. 2 we give some basic definitions and properties of DC functions and nonsmooth analysis. In Sect. 3 we describe our method for minimizing DC functions which has the property that every cluster point of the sequence is a critical point of the DC function. In Sect. 4 we study the convergence of the whole sequence and its convergence rate for a special case. Finally, in Sect. 5 we establish the convergence results for an inexact algorithm.

2 Preliminaries

In this section we recall some concepts and basic results from convex analysis and DC programming. For more details about these subjects we refer to [20,21,27,28]. The following notations will be used throughout the paper. Let $\langle\cdot,\cdot\rangle$ be the canonical inner product and $||\cdot||$ the corresponding Euclidean norm on $\mathbb{R}^n$. The open ball with center $x \in \mathbb{R}^n$ and radius $\epsilon > 0$ is denoted by $B(x,\epsilon) = \{y \in \mathbb{R}^n \ : \ ||x-y|| < \epsilon\}$. Let $\Gamma_0(\mathbb{R}^n)$ denote the convex cone of all the lower semicontinuous proper (i.e. not identically equal to $+\infty$) convex functions from $\mathbb{R}^n$ to $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. The effective domain of a function f, denoted $\mathrm{dom}(f)$, is defined by $\mathrm{dom}(f) = \{x \in \mathbb{R}^n \ : \ f(x) < +\infty\}$.

A general DC program is of the form $\alpha = \inf_{x\in\mathbb{R}^n}\{f(x) = g(x) - h(x)\}$, 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. Actually, we are concerned with the following case $\alpha = \inf_{x\in\mathrm{dom}(h)}\{f(x) = g(x)-h(x)\}$, which is equivalent to the last one under the above convention. Moreover, we suppose that $\mathrm{dom}(g) \cap \mathrm{dom}(h) \neq \emptyset$.

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, where ∂f denotes the subdifferential of f. We will denote by S the set of all critical points of f and throughout the paper $S \neq \emptyset$.

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. Properties which help to recognize a DC function can be found, for instance in [18,29]. $\mathcal{DC}(\mathbb{R}^n)$ is closed under the operations usually considered in optimization. For instance, a linear combination, a finite supremum or the product of two DC functions remain DC. Locally DC functions on $\mathbb{R}^n$ are DC functions on $\mathbb{R}^n$ (see [21] and references therein for the details). It is also known that the set of DC functions defined on a compact convex set of $\mathbb{R}^n$ is dense in the set of continuous functions on this set. Under some caution we can say that $\mathcal{DC}(\mathbb{R}^n)$ constitutes a minimal realistic extension of $\Gamma_0(\mathbb{R}^n)$.

Recall that a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be a *convex* (resp. *strongly convex* with modulus $\rho > 0$) function, if for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

$$(\text{resp. } f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\rho}{2}\lambda(1-\lambda)||x-y||^2).$$

We say that f is *locally convex* at a point $x \in \mathbb{R}^n$ if there exists a neighborhood U of x such that the restriction of f to U is a convex function.

Let $F : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ be a multivalued operator with domain $dom(F) = \{x \in \mathbb{R}^n : F(x) \neq \emptyset\}$. The operator F is called *monotone* (resp. *strongly monotone* with modulus $\rho > 0$), if for any $x, y \in \mathbb{R}^n$, $u \in F(x)$ and $v \in F(y)$ we have

$$\langle v - u, y - x\rangle \geq 0 \ \ (\text{resp. } \langle v - u, y - x\rangle \geq \rho||x - y||^2).$$

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *Lipschitz continuous*, if there exists a constant $L > 0$ such that

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

A function is called *locally Lipschitz continuous* at a point $z \in \mathbb{R}^n$ if the above inequality is satisfied for all $x, y \in B(z, \epsilon)$ for some $L > 0$ and $\epsilon > 0$. Since convex functions are locally Lipschitz continuous on $\mathbb{R}^n$, they are differentiable almost everywhere. The *subdifferential* of f at a point $x \in \mathbb{R}^n$ is the set

$$\partial f(x) = \{w \in \mathbb{R}^n \ : \ f(y) \geq f(x) + \langle w, y - x\rangle \quad \forall y \in \mathbb{R}^n\},$$

if $x \in \text{dom}(f)$ and $\partial f(x) = \emptyset$, if $x \notin \text{dom}(f)$. The subdifferential of a convex function f at a point $x \in \text{dom}(f)$ is a nonempty, convex and compact set. When f is both convex and differentiable at some point $x \in \text{dom}(f)$, then the subgradient is unique and equals to the gradient, i.e., $\partial f(x) = \{\nabla f(x)\}$. Furthermore, a lower semicontinuous function f is convex (resp. strongly convex) if and only if $\partial f(\cdot)$ is a monotone operator (resp. strongly monotone operator).

A sequence $\{y^k\}$ is called *Fejér convergent* to a nonempty set $U \subset \mathbb{R}^n$ if

$$||y^{k+1} - u|| \leq ||y^k - u||,$$

for all $u \in U$ and $k \in \mathbb{N}$.

Proposition 1 *If $\{y^k\}$ is Fejér convergent to a nonempty set $U \subset \mathbb{R}^n$, then $\{y^k\}$ is bounded. Furthermore, if every cluster point y of $\{y^k\}$ belongs to U, then* $\lim_{k\to\infty} y^k = y$.

Proof See [30]. □

3 Proximal linearized Algorithm

Consider $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ a DC function bounded from below, i.e., $f(x) = g(x) - h(x)$, with $g, h \in \Gamma_0(\mathbb{R}^n)$. In this section, we present the following proximal linearized algorithm for finding critical points of a DC function f, which considers in each proximal subproblem a linear approximation of f.

Algorithm 1

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

Step 2: Calculate

$$w^k \in \partial h(x^k). \tag{3}$$

Step 3: Compute

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{ g(x) - \langle w^k, x - x^k \rangle + \frac{1}{2\lambda_k} ||x - x^k||^2 \right\}. \tag{4}$$

Step 4: If $x^{k+1} = x^k$, stop. Otherwise, set $k := k + 1$ and return to Step 2.

The well definition of $\{w^k\}$ and $\{x^k\}$ is guaranteed if $h, g \in \Gamma_0(\mathbb{R}^n)$. Note that if $h(x) \equiv 0$, then Algorithm 1 becomes exactly the classical proximal point algorithm for convex functions (2).

Remark 1 It is worth mentioning that our algorithm and DCA algorithm (see [22]) share the same idea, namely, they linearize some component $g(\cdot)$ or $h(\cdot)$; or both of the DC objective function $f(x) = g(x) - h(x)$. However, Algorithm 1 is simpler because linearization is done directly, and not on the dual components, besides the fact that proximal algorithms are more efficient than subgradient algorithms. Recently, Bento and Soubeyran [14] discussed how ”generalized” proximal point algorithms can be a nice tools to modelize the dynamics of human behaviours on the context of the “variational rationality approach” (see Soubeyran [31]), where the term ”generalized” refers to a quasi distance such that not necessarily all the axioms of distance are verified, but preserving the nice properties of convexity, continuity and coercivity of the Euclidean norm. In this “variational rationality approach”, costs to be able to change from the current position x^k to x^{k+1} and costs to be able to stay in the current position x^k are not necessarily symmetric and equal to zero, respectively. In this context, Algorithm 1 seems to be more appropriate than the algorithm proposed in [15], because in our algorithm there is no auxiliar sequence working in the quadratic term in (4) while the algorithm in [15] has.

Now we shall establish the convergence of the algorithm. We begin by showing that Algorithm 1 is a descent algorithm.

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

1. *either the algorithm stops at a critical point;*
2. *or f decreases strictly, i.e., $f(x^{k+1}) < f(x^k)$, for all $k \in \mathbb{N}$.*

Proof From (3) and (4) we have $w^k \in \partial h(x^k)$ and

$$w^k \in \partial g(x^{k+1}) + \frac{1}{\lambda_k}(x^{k+1} - x^k), \tag{5}$$

respectively. If $x^{k+1} = x^k$, then $w^k \in \partial h(x^k) \cap \partial g(x^k)$, which means that x^k is a critical point of f. Now, suppose $x^{k+1} \neq x^k$. It follows from (3) and $h \in \Gamma_0(\mathbb{R}^n)$ that

$$h(x^{k+1}) \geq h(x^k) + \langle w^k, x^{k+1} - x^k \rangle. \tag{6}$$

On the other hand, from (4), we have

$$g(x^k) \geq g(x^{k+1}) - \langle w^k, x^{k+1} - x^k \rangle + \frac{1}{2\lambda_k}||x^{k+1} - x^k||^2. \tag{7}$$

Adding inequalities (6) and (7), we obtain

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{2\lambda_k}||x^{k+1} - x^k||^2. \tag{8}$$

Since $\{\lambda_k\}$ is a positive sequence and $x^{k+1} \neq x^k$, then $f(x^{k+1}) < f(x^k)$. □

Next result is a consequence of Theorem 1 and the lower boundedness of f.

Corollary 1 *Consider $\{x^k\}$ generated by Algorithm 1. Then the sequence $\{f(x^k)\}$ is convergent. Furthermore, if f is a continuous function and $\{x^k\}$ is bounded, then $\lim_{k\to\infty} f(x^k) = f(\overline{x})$, for some cluster point $\overline{x}$ of $\{x^k\}$.*

The following proposition will be useful to prove the convergence theorem.

Proposition 2 *Let $\{x^k\}$ be generated by Algorithm 1. Then* $\lim_{k\to+\infty} ||x^{k+1} - x^k|| = 0$.

Proof From (8), we have that $\sum_{k=0}^{n-1} \frac{1}{2\lambda_k}||x^{k+1} - x^k||^2 \leq f(x^0) - f(x^n)$. Since f is bounded from below and $\liminf_k \lambda_k > 0$, we obtain $\sum_{k=0}^{\infty} ||x^{k+1} - x^k||^2 < \infty$, and it follows that $\lim_{k\to+\infty} ||x^{k+1} - x^k|| = 0$. □

Note that from (3) and convexity of h, if $\{x^k\}$ is bounded, then $\{w^k\}$ is also bounded.

Theorem 2 *Suppose that $\{x^k\}$ is bounded. Then every cluster-point of $\{x^k\}$ is a critical point of the function f.*

Proof Let x^* and w^* be cluster points of $\{x^k\}$ and $\{w^k\}$, respectively. Then, there exist two subsequences x^{k_j} and w^{k_j} converging respectively to x^* and w^*, i.e., $x^{k_j} \to x^*$ and $w^{k_j} \to w^*$. Since h is convex and lower semicontinuous, it follows from (3) and the definition of subdifferential, taking j goes to $+\infty$, that $w^* \in \partial h(x^*)$. Now, we claim that $w^* \in \partial g(x^*)$. From (5), there exists $z^{k_j+1} \in \partial g(x^{k_j+1})$ such that

$$||w^{k_j} - z^{k_j+1}|| = \frac{1}{\lambda_{k_j}}||x^{k_j+1} - x^{k_j}||. \tag{9}$$

Since $\{\lambda_k\}$ is bounded, combining (9) and Proposition 2, we get that $\lim_{j\to+\infty} w^{k_j} = \lim_{j\to+\infty} z^{k_j+1} = w^*$. Therefore, since $g \in \Gamma_0(\mathbb{R}^n)$ and combining the definition of the subdifferential and the fact that $z^{k_j+1} \in \partial g(x^{k_j+1})$, taking j goes to $+\infty$, we have $w^* \in \partial g(x^*)$. Thus, $w^* \in \partial h(x^*) \cap \partial g(x^*)$, which means that x^* is a critical point of f. □

Remark 2 If f is continuous and $\{x^k\}$ is bounded it can be easily proved that the whole sequence $\{x^k\}$ converges to some critical point $x^* \in S$, as long as S satisfies the sharp minima condition; see [32]. Recently, finite termination and convergence rate of the proximal point method has been studied under the concept of weak sharp minima, Kurdyka-Lojasiewicz property and subregularity property; see [9,33,34]. We hope that this paper may stimulate further research involving proximal linearized algorithm for DC functions and these concepts. The results presented in this section are still true if we consider an approximate version obtained by replacing the exact subdifferential by the approximate one such as in [16].

4 Global convergence

Let us consider now the special case discussed in [16] where f is a DC function with $f(x) = g(x) - h(x)$ and the function h is differentiable. In this context (3) and (4) reduce to

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{ g(x) - \langle \nabla h(x^k), x - x^k \rangle + \frac{1}{2\lambda_k} ||x - x^k||^2 \right\} \tag{10}$$

and x^* is a critical point of f if $\nabla h(x^*) \in \partial g(x^*)$.

Theorem 3 *Consider Algorithm 1 with (3) and (4) replaced by (10). Suppose that g is a strongly convex function (with constant $\rho > 0$) and $\nabla h(x)$ a Lipschitz continuous function (with constant $L > 0$). If $\rho > 2L$, then there exists a constant $0 < r < 1$ such that*

$$||x^{k+1} - x^*|| \leq r||x^k - x^*|| \quad \forall x^* \in S. \tag{11}$$

Therefore, the whole sequence $\{x^k\}$ converges linearly to a point $x^* \in S$.

Proof From (10), there exists $z^{k+1} \in \partial g(x^{k+1})$ such that

$$\nabla h(x^k) = z^{k+1} + \frac{1}{\lambda_k}(x^{k+1} - x^k). \tag{12}$$

Let $x^* \in S$ be a critical point of f, namely, $\nabla h(x^*) \in \partial g(x^*)$. By strong monotonicity of $\partial g(\cdot)$ and (12), we have

$$\begin{aligned} 0 &\leq \langle x^{k+1} - x^*, z^{k+1} - \nabla h(x^*) \rangle - \rho ||x^{k+1} - x^*||^2 \\ &= \langle x^{k+1} - x^*, \frac{x^k - x^{k+1}}{\lambda_k} + \nabla h(x^k) - \nabla h(x^*) \rangle - \rho ||x^{k+1} - x^*||^2. \end{aligned}$$

Thus, since $\lambda_k > 0$, using the Cauchy-Schwarz inequality and the fact that $\nabla h(\cdot)$ is a Lipschitz function, we have

$$
\begin{aligned}
0 &\leq 2\lambda_k \langle x^{k+1} - x^*, \nabla h(x^k) - \nabla h(x^*)\rangle - 2\langle x^{k+1} - x^*, x^{k+1} - x^k\rangle \\
&\quad -2\lambda_k \rho ||x^{k+1} - x^*||^2 \\
&\leq 2\lambda_k L||x^{k+1} - x^*||||x^k - x^*|| - ||x^{k+1} - x^*||^2 - ||x^k - x^{k+1}||^2 \\
&\quad +||x^k - x^*||^2 - 2\lambda_k \rho ||x^{k+1} - x^*||^2 \\
&\leq 2\lambda_k L(||x^{k+1} - x^*||^2 + ||x^k - x^*||^2) - ||x^{k+1} - x^*||^2 - ||x^k - x^{k+1}||^2 \\
&\quad +||x^k - x^*||^2 - 2\lambda_k \rho ||x^{k+1} - x^*||^2 \\
&= (1 + 2\lambda_k L)||x^k - x^*||^2 - [1 + 2\lambda_k(\rho - L)]||x^{k+1} - x^*||^2 - ||x^k - x^{k+1}||^2.
\end{aligned}
$$

It immediately follows that

$$
[1 + 2\lambda_k(\rho - L)]||x^{k+1} - x^*||^2 \leq (1 + 2\lambda_k L)||x^k - x^*||^2.
$$

The first assertion is proved setting $0 < r := \sqrt{\dfrac{(1 + 2\lambda_k L)}{1 + 2\lambda_k(\rho - L)}} < 1$ in the last inequality. The second one follows from Proposition 1 and Theorem 2. □

Remark 3 If all the assumptions of Theorem 3 are satisfied and f is lower semicontinuous and locally convex at the limit point x^* of $\{x^k\}$, then there exist a neighborhood U of x^* and $k_0 \in \mathbb{N}$ such that $x^k \in U$, for all $k \geq k_0$ and the restriction of f to U is a convex function. From (5), we have

$$
\frac{1}{\lambda_k}(x^k - x^{k+1}) + \nabla h(x^k) - \nabla h(x^{k+1}) \in \partial g(x^{k+1}) - \nabla h(x^{k+1}) \subset \partial f(x^{k+1})
$$

that means, $f(x) \geq f(x^{k+1}) + \langle \frac{1}{\lambda_k}(x^k - x^{k+1}) + \nabla h(x^k) - \nabla h(x^{k+1}), x - x^{k+1}\rangle$, for all $x \in U$. Having in mind that $\nabla h(\cdot)$ is Lipschitz continuous and Proposition 2 it follows from last inequality, taking the lower limit, that x^* is a local minimizer of f. This provided a sufficient condition for a limit point of $\{x^k\}$ to be a local minimizer of f.

5 Inexact version

For the method to be practical, it is important to handle approximate solutions of subproblems. This consideration gives rise to inexact versions of the proximal point algorithm. Now, we present the following inexact algorithm for DC functions.

Algorithm 2

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

Step 2: Calculate

$$
w^k \in \partial h(x^k). \tag{13}
$$

Step 3: Compute $(x^{k+1}, \xi^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^n$ such that:

$$g(x^k) - g(x^{k+1}) - \langle w^k, x^k - x^{k+1}\rangle \geq \frac{(1-\sigma)}{2\lambda_k}||x^{k+1} - x^k||^2, \tag{14}$$

with

$$||\xi^{k+1} - w^k|| \leq \theta||x^{k+1} - x^k||, \tag{15}$$

where

$$\xi^{k+1} \in \partial g(x^{k+1}). \tag{16}$$

Step 4: If $x^{k+1} = x^k$, stop. Otherwise, set $k := k+1$ and return to Step 2.

Note that when $h(x) = 0$ Algorithm 2 becomes exactly the algorithm proposed in [9]. It follows from (14) that

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

where $0 \leq \epsilon_k = \frac{\sigma}{2\lambda_k}||x^{k+1} - x^k||^2$. When $\epsilon_k = 0$ (it holds by taking $\sigma = 0$) (4) implies the weaker condition (14).

Theorem 4 *The sequence $\{x^k\}$ generated by Algorithm 2 satisfies:*

1. *either the algorithm stops at a critical point;*
2. *or f decreases strictly, i.e., $f(x^{k+1}) < f(x^k)$, for all $k \in \mathbb{N}$.*

Furthermore, $\lim_{k\to+\infty} ||x^{k+1} - x^k|| = 0$.

Proof The proof uses exactly the same argument as the one used to prove Theorem 1 and Proposition 2 having in mind that $\sigma \in [0, 1)$. □

Corollary 2 *Consider $\{x^k\}$ generated by Algorithm 2. Then the sequence $\{f(x^k)\}$ is convergent. Additionally, if f is a continuous function and $\{x^k\}$ is bounded, then $\lim\limits_{k\to\infty} f(x^k) = f(\overline{x})$, for some cluster point $\overline{x}$ of $\{x^k\}$.*

Proof The proof uses the same argument as the one used to prove Corollary 1. □

Theorem 5 *Let $\{x^k\}$ generated by Algorithm 2. Then every cluster point of $\{x^k\}$, if any, is a critical point of f.*

Proof Let x^* be a cluster point of $\{x^k\}$. So, consider a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to x^*. As mentioned before, we know that $\{w^k\}$ is also bounded. Consider a subsequence $\{w^{k_j}\}$ of $\{w^k\}$ converging to w^*. From (15) and Theorem 4, we have

$$\lim_{j\to+\infty} \xi^{k_j+1} = \lim_{j\to+\infty} w^{k_j} = w^*. \tag{17}$$

From definition of the algorithm $w^{k_j} \in \partial h(x^{k_j})$ and $\xi^{k_j+1} \in \partial g(x^{k_j+1})$. Therefore,

$$h(y) \geq h(x^{k_j}) + \langle w^{k_j}, y - x^{k_j}\rangle \quad \forall y \in \mathbb{R}^n$$

and

$$g(y) \geq g(x^{k_j+1}) + \langle \xi^{k_j+1}, y - x^{k_j+1} \rangle \quad \forall y \in \mathbb{R}^n,$$

respectively. Taking j goes to $+\infty$ in last two inequalities, since $h, g \in \Gamma_0(\mathbb{R}^n)$, we have $w^* \in \partial h(x^*) \cap \partial g(x^*)$ and the proof is complete. □

6 Conclusion

Future researches will examine the case where the Euclidean norm in the regularization term is replaced by a generalized quasi distance which is more appropriate to interpret our results in terms of behavioural science. Algorithms for multiobjective DC problems has not been considered yet. An extension of our algorithm to this context will be studied as well as modified versions of the algorithms where linear or quadratic approximations of g will be considered. We expect that the results of this paper become a further step towards solving DC optimization problems. We foresee further progress in this topics in the near future.

Acknowledgments The authors wish to express their gratitude to the anonymous referees for them helpful comments.

References

1. Martinet, B.: Regularisation d'inéquations variationelles par approximations succesives. Rev. Française d'Inform. Recherche Oper. **4**, 154–159 (1970)
2. Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
3. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. control. optim. **14**, 877–898 (1976)
4. Kaplan, A., Tichatschke, R.: Proximal point methods and nonconvex optimization. J. Glob. Optim. **13**, 389–406 (1998)
5. Hare, W., Sagastizábal, C.: Computing proximal points of nonconvex functions. Math. Program. **116**(1), 221–258 (2009)
6. Otero, R.G., Iusem, A.N.: Proximal methods in reflexive Banach spaces without monotonicity. J. Math. Anal. Appl. **330**(1), 433–450 (2007)
7. Iusem, A.N., Pennanen, T., Svaiter, B.F.: Inexact variants of the proximal point algorithm without monotonicity. SIAM J. Optim. **13**(4), 1080–1097 (2003)
8. 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)
9. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. Math. Program **137**(1–2), 91–129 (2013)
10. Burachik, R.S., Svaiter, B.F.: A relative error tolerance for a family of generalized proximal point methods. Math. Oper. Res. **26**(4), 816–831 (2001)
11. Solodov, M.V., Svaiter, B.F.: Error bounds for proximal point subproblems and associated inexact proximal point algorithms. Math. Program **88**(2), 371–389 (2000)
12. Solodov, M.V., Svaiter, B.F.: A unified framework for some inexact proximal point algorithms. Numer. Funct. Anal. Optim. **22**(7–8), 1013–1035 (2001)
13. Zaslavski, A.: Convergence of a proximal point method in the presence of computational errors in Hilbert spaces. SIAM J. Optim. **20**(5), 2413–2421 (2010)

14. Bento, G.C., Soubeyran, A.: Generalized inexact proximal algorithms: Routine's formation with resistance to change, following worthwhile changes. J. Optim. Theory Appl. **172**(1), 1–16 (2015)
15. Sun, W., Sampaio, R.J.B., Candido, M.A.B.: Proximal point algorithm for minimization of DC Functions. J. Comput. Math. **21**, 451–462 (2003)
16. Moudafi, A., Maingé, P.-E.: On the convergence of an approximate proximal method for d.c. functions. J. Comput. Math. **24**, 475–480 (2006)
17. Souza, J.C.O., Oliveira, P.R.: A proximal point algorithm for DC functions on Hadamard manifolds. J. Glob. Optim. (2015). doi:10.1007/s10898-015-0282-7
18. Hartman, P.: On functions representable as a difference of convex functions. Pac. J. Math. **9**, 707–713 (1959)
19. Bomze, I., Lemaréchal, C.: Necessary conditions for local optimality in difference-of-convex programming. J. Convex Anal. **17**, 673–680 (2010)
20. Horst, R., Thoai, N.V.: DC programming: overview. J. Optim. Theory Appl. **103**(1), 1–43 (1999)
21. Hiriart-Urruty, J.B.: Generalized differentiabity, duality and optimization for problems dealing with difference of convex functions, Convexity and Duality in Optimization. Lectur. Notes Econ. Math. Syst **256**, 37–70 (1986)
22. Pham, D.T., Souad, E.B.: Algorithms for solving a class of nonconvex optimization problems: methods of subgradient. Fermat Days 85: Math. Optim. **129**, 249–271 (1986)
23. Ferrer, A., Bagirov, A., Beliakov, G.: Solving DC programs using the cutting angle method. J. Glob. Optim. **61**(1), 71–89 (2015)
24. Pham, D.T., An, L.T.H., Akoa, F.: The DC (Difference of Convex Functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**, 23–46 (2005)
25. Holmberg, K., Tuy, H.: A production-transportation problem with stochastic demand and concave production costs. Math. Program. **85**, 157–179 (1999)
26. Chen, P.C., Hansen, P., Jaumard, B., Tuy, H.: Solution of the multisource weber and conditional weber problems by d.c. programming. Oper. Res. **46**(4), 548–562 (1998)
27. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex analysis and minimization algorithms. Springer, Berlin (1993)
28. Rockafellar, R.T.: Convex analysis. Princeton University Press, Princeton, New Jersey (1970)
29. Ginchev, I., Gintcheva, D.: Characterization and recognition of dc functions. J. Glob. Optim. **57**, 633–647 (2013)
30. Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F.: Full convergence of the steepest descent method with inexact line searches. Optimization **32**(2), 137–146 (1995)
31. Soubeyran, A: Variational rationality. Human behaviors as worthwhile stay and change transitions, possibly ending in traps, before reaching desires. Preprint at GREQAM-AMSE (2015)
32. Polyak, B.T.: Sharp Minima Institute of Control Sciences Lecture Notes, Moscow, USSR, 1979. Presented at the IIASA workshop on generalized Lagrangians and their applications, IIASA, Laxenburg, Austria (1979)
33. Ferris, M.C.: Weak sharp minima and penalty functions in mathematical programming. Ph.D. Thesis. University of Cambridge, UK (1988)
34. Li, G., Mordukhovich, B.S.: Holder metric subregularity with applications to proximal point method. SIAM J. Optim. **22**, 1655–1684 (2012)