An Extension of the Proximal Point Method for Quasiconvex Minimization

E. A. Papa Quiroz *and P. Roberto Oliveira

Department of Mathematics and Natural Sciences,
Universidad Nacional del Callao, Callao-Perú.
erikpapa@gmail.com

Department of Systems Engineering and Computer Sciences,
PESC-COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
poliveir@cos.ufrj.br

December 3, 2008

Abstract

In this paper we propose an extension of the proximal point method to solve minimization problems with quasiconvex objective functions on the Euclidean space and the nonnegative orthant. For the unconstrained minimization problem, assumming that the function is bounded from below and lower semicontinuous we prove that iterations $\{x^k\}$ given by $0 \in \widehat{\partial}(f(.)+(\lambda_k/2)||.-x^{k-1}||^2)(x^k)$ are well defined and if, in addition, f is quasiconvex then $\{f(x^k)\}$ is decreasing and $\{x^k\}$ converges to a point of $U := \{x \in I\!R^n : f(x) \leq \inf_{j\geq 0} f(x^j)\}$ assumed nonempty. Under the assumption that the sequence of parameters is bounded and f is continuous it is proved that $\{x^k\}$ converges to a generalized critical point of f. Furthermore, if $\{\lambda_k\}$ converge to zero and the iterations $\{x^k\}$ are global minimizers of the regularized subproblems $f(.) + (\lambda_k/2)||. - x^{k-1}||^2$, the sequence converges to an optimal solution. For the quasiconvex minimization on the nonnegative orthant, using the same premises of the unconstrained case and using a general proximal distance (which includes as particular cases Bregman distances, $\phi-$ divergence distances and second order homogeneous distances), we find that the iterations $\{x^k\}$ given by $0 \in \widehat{\partial}(f(.) + \lambda_k d(., x^{k-1}))(x^k)$, are well defined and rest in the positive orthant and $\{f(x^k)\}$ is decreasing and convergent. If $U_+ = \{x \in I\!R^n_+ \cap \texttt{dom}f : f(x) \leq \inf_{j\geq 0} f(x^j)\}$ is nonempty, $\{\lambda_k\}$ is bounded and the distance is separable, we obtain the convergence to a generalized KKT point of the problem. Furthermore, in the smooth case we introduce a sufficient condition on the proximal distance such that the sequence converges to an optimal solution of the problem. When the parameters converge to zero we obtain a similar result, of the unconstrained case for self-proximal distances.

Keywords: Proximal point method, quasiconvex functions, nonnegative orthant, proximal distances, generalized critical point.

*Corresponding author: E.A. Papa Quiroz

1 Introduction

This paper studies the global convergence of an extension of the proximal point method for solving minimization problems with quasiconvex objective functions.

Let us consider the unconstrained optimization problem

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

where $f : I\!R^n \to I\!R \cup \{+\infty\}$ is a proper lower semicontinuous function and $I\!R^n$ is the Euclidean space with norm denoted by $||.||$. The classical proximal point method generates a sequence $\{x^k\}$ given by $x^0 \in I\!R^n$ (an arbitrary point) and

$$x^k \in \arg\min\{f(x) + (\lambda_k/2)||x - x^{k-1}||^2 : x \in I\!R^n\},$$

where λ_k is a certain positive parameter. This method was introduced by Martinet [23] for convex optimization problems and further developed by Rockafellar in [31] and [32] for the problem of finding zeros of maximal monotone operators. It is well known, see Guler [16], that if f is convex and $\{\lambda_k\}$ satisfies

$$\sum_{k=1}^{+\infty}(1/\lambda_k) = +\infty, \tag{1.2}$$

then $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in I\!R^n\}$. Furthermore, if the optimal set is nonempty we obtain that $\{x^k\}$ converges to an optimal solution of the problem (1.1).

On the other hand, consider the problem

$$\min\{f(x) : x \geq 0\} \tag{1.3}$$

where $f : I\!R^n \to I\!R \cup \{+\infty\}$ is a proper lower semicontinuous function and $x \geq 0$ means that $x_i \geq 0, \forall i = 1, ..., n$. The class of interior proximal point methods to solve (1.3) generates a sequence $\{x^k\}$ given by $x^0 \in I\!R^n_{++}$ (arbitrary given point) and

$$x^k \in \arg\min_{x\geq 0}\{f(y) + \lambda_k d(x, x^{k-1})\}, \tag{1.4}$$

where λ_k is a positive parameter and d is a function satisfying some desirable properties. For example we can define a like-distance function d such that the iterations x^k rest on the positive orthant. Some examples of d based on the literature are

- Bregman distances

  $$d(x, y) = D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y\rangle,$$

  where $h : I\!R^n \to I\!R$ is a Bregman function, see [12, 15, 18, 22];

- $\varphi-$divergence distances

  $$d_\varphi(x, y) = \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right),$$

  where $\varphi : I\!R_+ \to (-\infty, +\infty)$ is a function satisfying some desirable properties, see [19, 20, 29, 30]. This class was introduced by Teboulle [29] in order to define entropy-like proximal maps;

- Second order homogeneus distances

$$d_\phi(x,y)=\sum_{i=1}^{n} y_i^2\phi\left(\frac{x_i}{y_i}\right),$$

where $\phi(t)=\mu\varphi(t)+\frac{\nu}{2}(t-1)^2$ with $\mu,\nu>0$ and φ is as above. This class was introduced by Auslender et.al [7] to solve convex minimization problems and extended by the same authors in [8] to solve a class of variational inequalities problems.

Assuming that f is a proper, lower semicontinuous and convex function, $\texttt{dom}f \cap I\!R^n_{++}\neq\phi$, and λ_k satisfies (1.2) and using a class of proximal distances which contain the three classes mentioned above, Auslender and Teboulle [6] have proved that $\lim_{k\to\infty} f(x^k)=\inf\{f(x):x\geq 0\}$. Furthermore, if the optimal set is nonempty, then sequence $\{x^k\}$ converges to an optimal solution of the problem (1.3).

On the other hand, several applications in diverse Science and Engineering areas are sufficient motivation to work with non convex objective functions and proximal point methods, see for example [2, 5, 27]. In particular the class of quasiconvex minimization problems has been receiving attention from many researches due to the broad range of applications, for example, in location theory [17], control theory [9] and specially in economic theory [28].

The classical proximal point method to solve the problem (1.1) for non convex objective functions was studied by some researchers. Tseng [34] proved a weak convergence result, that is, assumming that f is lower semicontinous and bounded from below function and $\lambda_k=\lambda>0$ then every cluster point z is a stationary point of f, i.e.

$$f'(z,d):=\liminf_{\lambda\downarrow 0}\left(\frac{f(z+\lambda d)-f(z)}{\lambda}\right)\geq 0.$$

Kaplan and Tichatschke [21] studied the method for a class of nonconvex functions when the auxiliary function $f(x)+(\lambda_k/2)||x-x^{k-1}||^2$ becomes strongly convex on some certain convex sets under a suitable choice of λ_k. The authors proved that the sequence stop in a finite number of iterations in a stacionary point or any accumulation point of $\{x^k\}$ is a stationary point of f. Unfortunately, it has been known, see Absil et. al [1] and Palis-De Melo [24], that the full convergence of the sequence generated by the proximal method is not true even when $\{x^k\}$ is bounded and f is sufficiently smooth. So a natural question arose: under what minimal condition may the full convergence of the proximal point method be proved?

An advance in this direction, for unconstrained minimization, was given by Attouch and Bolte [2] where, under the assumptions that f satisfies a Lojasiewicz property and $\{x^k\}$ is bounded, they have proved the convergence of the method to some generalized critical point of the problem. For smooth quasiconvex minimization on the non negative orthant there are some recent works in the literature. Attouch and Teboulle [4], with a regularized Lotka-Volterra dynamical system, have proved the convergence of the continuous method to a point which belongs to certain set which contains the set of optimal points. Cunha et al.[13] and Chen and Pan [14], with a particular $\phi-$divergence distance, have proved the full convergence of the proximal method to the KKT-point of the problem when parameter λ_k is bounded and convergence to an optimal solution when $\lambda_k\to 0$. Pan and Chen [25], with the second-order homogeneus distance, which includes the Logarithmic-Quadratic proximal point method, studied in [26], and Souza et al. [27] with a class of separated Bregman distances, have proved the same convergence result of [13, 14].

In this paper we are interested in extending the global convergence of a large class of proximal point methods to minimize quasiconvex functions without unconstrained and constrained on the nonnegative orthant.

The mean difficulty we observed in extending the proximal method for nonconvex function is that due to the nonconvexity of f the subproblems may not be convex and thus, from a practical point of view, we may obtain that minimization subproblems may be as hard to solve globally as the original one due to the existence of multiple isolated local minimizers. Consider for example function $f(x_1, x_2) = -\exp(-x_1^2 - x_2^2)$ which is nonconvex as also the regularized function $f(x_1, x_2) = -\exp(-x_1^2 - x_2^2) + (1/40)((x_1-6)^2 + (x_2-6)^2)$. Another observation is that in a possible extension of the proximal method the property that $\{f(x^k)\}$ let be nonincreasing would be satisfied. Observe that this condition is essencial to obtain the convergence of the proximal method in the convex case.

To solve this disadvantage we propose the following extensions for the proximal method:

- For the unconstrained case, we introduce the following iteration: given x^{k-1} find x^k such that
$$0 \in \widehat{\partial}(f(.) + (\lambda_k/2)||. - x^{k-1}||^2)(x^k),$$
where $\widehat{\partial}$ is the regular subdifferential (see Subsection 2.1). This iteration resorts to the method, to be more practical and to embrace more applications in comparison with early research works where to find a global minimum in each iteration is required, see [2, 21, 34]. Under the assumption that f is proper lower semicontinuous and bounded from below we will prove that $\{x^k\}$ is well defined and if, in addition, f is quasiconvex it will be proved that $\{f(x^k)\}$ is decreasing and $\{x^k\}$ converges to some point of $U := \{x \in I\!R^n : f(x) \leq \inf_{j\geq 0} f(x^j)\}$, assumed nonempty. If $\{\lambda_k\}$ is bounded from above and f is continuous we prove that $\{x^k\}$ converges to a generalized stationary point. Furthermore, under the condition that $\{x^k\}$ are global minimizers of $f(.) + (\lambda_k/2)||. - x^{k-1}||^2$ and $\lambda_k \to 0$ we obtain the convergence to a minimum point of the problem. Observe that the last result is usefull for a subclass of problems where the objective function is weak convex functions or weak semistrict quasi-convex functions. Observe that for those classes of functions all local minimums are global minimums.

- For the minimization on the nonnegative orthant, we introduce the following iteration: given $x^{k-1} \in I\!R^n_{++}$ find x^k such that
$$0 \in \widehat{\partial}(f(.) + \lambda_k d(., x^{k-1}))(x^k),$$
where d is a proximal distance, see Definition 4.1. Observe that this condition is more practical than (1.4) where a global minimum point is required and therefore more practical than the works of [13, 14, 25, 26, 27]. Assuming the same assumption on f as in the unconstrained case we prove that $\{x^k\}$ is well defined and rests in the positive orthant $I\!R^n_{++}$, $\{f(x^k)\}$ is decreasing and $\{x^k\}$ converges in $U_+ := \{x \in I\!R^n_+ : f(x) \leq \inf_{j\geq 0} f(x^j)\}$, assumed nonempty. If $\{\lambda_k\}$ is bounded from above and d is separable we prove that $\{x^k\}$ converges to a generalized KKT point of the problem and if $\{\lambda_k\}$ converges to zero and assuming that $\{x^k\}$ are global minmizers of the subproblems we prove the convergence to a minimum point of (1.3). Furthermore, for the smooth case, we introduce a condition on the proximal distance to obtain the convergence to an optimal point when $\{\lambda_k\}$ is bounded.

Due to the general approach of this work we recover as particular cases recently works on the subject and also extend some other research works. Indeed, for unconstrained quasiconvex minimization we extend our result obtained in [26]. Also, the particular proximal method studied by Cunha et al.[13], and Chen and Pan [14] is extended to a large class of $\varphi-$ divergences; another extension is the paper of Souza et.al [27] on the class of boundary coercive and non smooth quasiconvex minimization.

The paper is organized as follows: In Section 2 we give some results on regular and general subgradients, quasiconvex theory, Fejér convergence in $I\!R^n$ and the sufficient condition of Arrown and Enthoven for quasiconvex minimization. In Section 3, we analyze an extension of the proximal point method for solving unconstrained minimization problems with quasiconvex functions. In Section 4 we introduce a proximal point method with proximal distances for solving minimization problems with quasiconvex functions on the nonnegative orthant and analyse its convergence properties. Finally, in Section 5 we give our conclusions.

2 Basic Results

In this section we first recall some definitions and results on regular subgradients (called also Fréchet subdifferential) and general subgradients (also called limiting subdifferentials), see Chapter 8 of Rockafellar and Wets [33]. Then, we give some results on quasiconvex functions, Fejér convergence and a theorem by Arrown and Enthoven which gives a sufficient condition to minimize quasiconvex minimization problems.

2.1 Subdifferential of Nonconvex Functions

Throughout this paper $I\!R^n$ is the Euclidean space endowed with the canonical inner product $\langle , \rangle$ and the norm of x given by $||x|| := \langle x, x\rangle^{1/2}$. Given an extended real valued function $f : I\!R^n \to I\!R \cup \{\pm\infty\}$ we denote its domain by

$$\texttt{domf} := \{x \in I\!R^n : f(x) < +\infty\}$$

and its epigraph by

$$\texttt{epi}f := \{(x, \beta) \in I\!R^n \times I\!R : f(x) \leq \beta\}.$$

f is said to be proper if $\texttt{domf} \neq \phi$ and $\forall x \in \texttt{domf}$ we have $f(x) > -\infty$.
Finally, f is a lower semicontinuous in $\bar{x}$ if for each $\{x^l\}$ such that $\lim_{l\to+\infty} x^l = \bar{x}$ we have

$$f(\bar{x}) \leq \liminf_{l\to+\infty} f(x^l).$$

We said that f is a lower semicontinuous if it is a lower semicontinuous at each point of $\texttt{dom}f$. It can be easily proven that the lower semicontinuity of f is equivalent to the closedness of the level set $L_f(\alpha) = \{x \in I\!R^n : f(x) \leq \alpha\}$, for each $\alpha \in I\!R$.

Definition 2.1 *Let $f : I\!R^n \to I\!R \cup \{+\infty\}$ be a proper function. For each $x \in \texttt{domf}$, the set of* regular subgradients *(also called* Fréchet subdifferential*) of f at x, denoted by $\widehat{\partial} f(x)$, is the set of vectors $s \in I\!R^n$ such that*

$$\liminf_{y\neq x, y\to x} \frac{1}{||x-y||}[f(y) - f(x) - \langle s, y - x\rangle] \geq 0.$$

If $x \notin \texttt{domf}$ then $\widehat{\partial} f(x) = \phi$.

Observe that the above definition is equivalent to

$$f(y) \geq f(x) + \langle s, y - x \rangle + o(||x - y||),$$

where

$$\lim_{y \to x} \frac{o(||x - y||)}{||x - y||} = 0.$$

It can be proved that $\widehat{\partial} f(x)$ is a closed and convex set. However, regular subgradients are inadequate for the calculus covering some of the properties we need, so we recall the following

Definition 2.2 *Let $f : I\!R^n \to I\!R \cup \{+\infty\}$ be a proper function. The set of* general subgradients *(also called* limiting subdifferential*) of f at $x \in I\!R^n$, denoted by $\partial f(x)$, is defined as follows:*

$$\partial f(x) := \{s \in I\!R^n : \exists x^n \to x, f(x^n) \to f(x), s^n \in \widehat{\partial} f(x^n) \to s\}.$$

It can be proved, see Chapter 8 of [33], that for an arbitrary lower semicontinuous function f, both the domain of $\widehat{\partial} f$ and the domain of ∂f are dense in the domain of f.

Proposition 2.1 *The following properties are true*

a. $\widehat{\partial} f(x) \subset \partial f(x)$*, for all* $x \in I\!R^n$;

b. *If f is differentiable at $\bar{x}$ then $\widehat{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$, so $\nabla f(\bar{x}) \in \partial f(\bar{x})$;*

c. *If f is continuously differentiable in a neighborhood of x, then $\widehat{\partial} f(x) = \partial f(x) = \{\nabla f(x)\}$;*

d. *If $g = f + h$ with f finite at $\bar{x}$ and h is continuously differentiable on a neighbourhood of $\bar{x}$ then*

$$\widehat{\partial} g(\bar{x}) = \widehat{\partial} f(\bar{x}) + \nabla h(\bar{x}),$$
$$\partial g(\bar{x}) = \partial f(\bar{x}) + \nabla h(\bar{x}).$$

Proof. See Rockafellar and Wets [33], pp. 304, exercise 8.8. ■

In order to work with minimization problems we need the following generalized definition.

Definition 2.3 *Let $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ be a proper function. A point $x \in$ `domf` is said to be a generalized critical point of f if $0 \in \partial f(x)$.*

Theorem 2.1 *If a proper function $g : I\!R^n \to I\!R \cup \{+\infty\}$ has a local minimum at $\bar{x}$ then $0 \in \widehat{\partial} g(\bar{x})$ and therefore,*

$$0 \in \partial g(\bar{x}).$$

Proof. Defining $o(||y - \bar{x}||) = \min\{0, g(y) - g(\bar{x})\}$ and using Proposition 2.1, **a**, we obtain the intended result. ■

2.2 Quasiconvex Functions

Definition 2.4 *Let $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ be a proper function. f is called quasiconvex if for all $x, y \in I\!R^n$, and for all $t \in [0, 1]$, it holds that*

$$f(tx + (1 - t)y) \leq \max\{f(x), f(y)\}.$$

Observe that if f is quasiconvex then `domf` is a convex set. On the other hand, while a convex function can be characterized by the convexity of its epigraph, a quasiconvex function can be characterized by the convexity of the level sets, as shown in the following result.

Theorem 2.2 *Let $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ be a proper function. Then f is quasiconvex if and only if the set $\{x \in I\!R^n : f(x) \leq c\}$ is convex for each $c \in I\!R$.*

Proof. See [10], Theorem 3.5.2, pp. 108. ■

For differentiable functions we have the following characterization of quasiconvexity

Theorem 2.3 *Let $f : I\!R^n \to I\!R$ be a differentiable function on $I\!R^n$. Then, f is quasiconvex if and only if, the following statement holds:*
if $f(x) \leq f(y)$, then

$$\langle \nabla f(y), x - y \rangle \leq 0.$$

Proof. See [10], Theorem 3.5.4, pp. 109. ■

For the nondifferentiable case we will use the following result:

Theorem 2.4 *Let $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ be a proper lower semicontinuous and quasiconvex function. If $g \in \widehat{\partial} f(x)$ and $f(y) \leq f(x)$ then*

$$\langle g, y - x \rangle \leq 0.$$

Proof. Let $t \in (0, 1)$, then from de quasiconvexity of f and the assumption that $f(y) \leq f(x)$ we have

$$f(x + t(y - x)) \leq \max\{f(x), f(y)\} = f(x).$$

As $g \in \widehat{\partial} f(x)$ we obtain

$$f(x + t(y - x)) \geq f(x) + t\langle g, y - x \rangle + o(t),$$

where $\lim_{t \to 0} \frac{o(t)}{t} = 0$. From both inequalities we conclude that

$$t\langle g, y - x \rangle + o(t) \leq 0.$$

Dividing by t and taking $t \to 0$ we obtain that

$$\langle g, y - x \rangle \leq 0,$$

the desired result. ■

Definition 2.5 *A function $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ is said to be pseudoconvex if for any $x, y \in$ `domf` one has that there exists $s \in \partial f(y)$ such that if*

$$\langle s, x - y \rangle \geq 0, \text{ we have } f(y) \leq f(x).$$

2.3 Fejér Convergence Results

Definition 2.6 *A sequence* $\{y^k\} \subset I\!R^n$, $k \geq 0$, *is* Fejér convergent *to a nonempty set* $U \subset I\!R^n$, *if for every* $u \in U$ *we have*

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

Theorem 2.5 *If* $\{y^k\}$ *is Fejér convergent to a nonempty set* $U \subseteq I\!R^n$, *then* $\{y^k\}$ *is bounded. If, furthermore, a cluster point* $\bar{y}$ *of* $\{y^k\}$ *belongs to* U, *then* $\{y^k\}$ *converges and* $\lim_{k\to+\infty} y^k = \bar{y}$.

Proof. See Burachik et al. [11]. ■

2.4 Arrown-Enthoven Sufficient Condition for Quasiconvex Minimization

Consider the problem

$$\min\{f(x) : g(x) \leq 0, x \geq 0\}, \tag{2.5}$$

where $f : I\!R^n \to I\!R$ and $g : I\!R^n \to I\!R^m$ such that $g(x) = (g_1(x), ..., g_m(x))$ with $g_i : I\!R^n \to I\!R$ and $x \geq 0$ mean that $x_i \geq 0$, for each $i = 1, ..., n$.
A variable x_{i_0} is a relevant variable if there exists some point in the constraint set, say $\bar{x}$, at which $\bar{x}_{i_0} > 0$.

Theorem 2.6 *Let* f *be a differentiable quasiconvex function and* g_i *are differentiable quasiconvex functions. If* $\bar{x}$ *satisfies the KKT necessary condition and let one of the following conditions be satisfied*

a) $\frac{\partial f}{\partial x_i}(\bar{x}) > 0$ *for at least one variable* x_i;

b) $\frac{\partial f}{\partial x_i}(\bar{x}) < 0$ *for some relevant variable* x_i;

c) $\nabla f(\bar{x}) \neq 0$ *and* f *is twice differentiable in a neighbourhood of* $\bar{x}$;

d) f *is convex.*

then $\bar{x}$ *is a global minimum of the problem (2.5).*

Proof. See Theorem 1 of Arrown and Enthoven, [3]. ■

3 Proximal Point Method for Unconstrained Minimization

In this section we are interested in solving the unconstrained optimization problem:

$$(p)\ \ \min\{f(x) : x \in I\!R^n\}$$

where $f : I\!R^n \to I\!R \cup \{+\infty\}$ is a proper function. To solve (p) we introduce the following method:

Given a sequence of positive parameters $\{\lambda_k\}$ and an initial point

$$x^0 \in I\!R^n. \tag{3.6}$$

For each $k = 1, 2...., $ if $0 \in \widehat{\partial} f(x^{k-1})$, then stop. Otherwise, find $x^k \in I\!R^n$ such that

$$0 \in \widehat{\partial}(f(.) + (\lambda_k/2)||. - x^{k-1}||^2)(x^k) \tag{3.7}$$

Remark 3.1 *Observe that the proposed method is an extension (for nonconvex functions) of the proximal point method. In fact, if f is convex then (3.7) becomes*

$$x^k = \arg\min\{f(x) + (\lambda_k/2)||x - x^{k-1}||^2 : x \in I\!R^n\}.$$

Remark 3.2 *As we are interested in solving (p) when f is nonconvex, it is important to observe that the method (3.6)-(3.7) only needs, in each iteration, to find an stationary point (and not necessarily a global minimum) of the regultarized function $f(.)+(\lambda_k/2)||.-x^{k-1}||^2$, so we believe that local algorithms can be used satisfactorily in each iteration.*

We will prove, under the only assumptions that f is proper, bounded from below and lower semicontinuous, that the iterations $\{x^k\}$ are well defined and if f is quasiconvex $\{f(x^k)\}$ is non increasing and converges. Then defining a set which contains the optimal set of the problem, we will obtain that $\lim_{k\to+\infty} f(x^k) = \inf_{x\in I\!R^n} f(x)$ if that set is empty and the convergence of $\{x^k\}$ to a generalized critical point if $\{\lambda_k\}$ is bounded and f is continuous in its domain. Furthermore, if each x^k is a global minimum of the regularized function $f(.)+(\lambda_k/2)||.-x^{k-1}||^2$ and $\lambda_k \to 0$ we prove that $\{x^k\}$ converges to a minimum point of the problem (p). As particular cases we obtain the convergence of the method to a global minimum point for pseudoconvex and convex functions.

3.1 Convergence Results

Theorem 3.1 *If $f : I\!R^n \to I\!R \cup \{+\infty\}$ is a proper lower bounded and lower semicontinuous on* `domf`*, then sequence $\{x^k\}$, given by (3.6)-(3.7), exists.*

Proof. We proceed by induction. It holds for $k = 0$, due to (3.6). Assume x^k exists. As f is lower semicontinuous and bounded below and $||\,.-x^k||^2$ is coercive then $f(.)+(\lambda_k/2)||\,.-x^k||^2$ is lower semicontinuous and coercive, thus this function has a global minimum (in particular a local minimum) x^{k+1} and thus $0 \in \widehat{\partial}(f(.) + (\lambda_{k+1}/2)||.-x^k||^2)(x^{k+1})$. ■

As we are interested in solvING (p) when f is quasiconvex we have given the following assumptions

Assumption A. $f : I\!R^n \to I\!R \cup \{+\infty\}$ is a proper function bounded from below.

Assumption B. $f : I\!R^n \to I\!R \cup \{+\infty\}$ is lower semicontinuous and quasiconvex.

As we are interested in the asymptotic convergence of the method we also assume that in each iteration $0 \notin \widehat{\partial} f(x^k)$ which implies that $x^k \neq x^{k-1}$, for all k.

Remark 3.3 *From (3.7), Proposition 2.1,* **d**, *and the smoothness of $(\lambda_k/2)||.-x^{k-1}||^2$, this implies that*

$$0 \in \widehat{\partial} f(x^k) + \lambda_k(x^k - x^{k-1}).$$

Thus, there exists $g^k \in \widehat{\partial} f(x^k)$ such that

$$g^k = \lambda_k(x^{k-1} - x^k).$$

Proposition 3.1 *Under assumptions A and B we have that $\{f(x^k)\}$ is decreasing and converges.*

Proof. As $x^k \neq x^{k-1}$, then

$$\langle x^{k-1} - x^k, x^{k-1} - x^k \rangle > 0$$

From Remark 3.3 we have

$$\langle g^k, x^{k-1} - x^k \rangle > 0.$$

Using the quasiconvexity of f and Theorem 2.4, this implies that

$$f(x^k) < f(x^{k-1}).$$

The convergence of $\{f(x^k)\}$ is inmediate from the lower boundedness of f. ■

Proposition 3.2 *The iterates x^k do not cycle.*

Proof. Suposse that there exists $l > j+1$ such that $x^l = x^j$. From Proposition 4.1 we have that

$$f(x^j) = f(x^l) < ... < f(x^{j+1}) < f(x^j),$$

which is a contradiction, so $x^l \neq x^j$. ■

Now, we define the following set

$$U := \{x \in I\!R^n : f(x) \leq \inf_{j \geq 0} f(x^j)\}.$$

Observe that this set depends on the choice of the initial iterates x^0 and sequence $\{\lambda_k\}$. If $U = \emptyset$ then it can be easily proven that

i) $\lim_{k \to +\infty} f(x^k) = \inf_{x \in I\!R^n} f(x)$,

ii) $\{x^k\}$ is unbounded.

From now on we assume that $U \neq \emptyset$ and define the following set

$$V_k := \{x \in I\!R^n : f(x) \leq f(x^k)\}.$$

We observe that $U \subset V_k$. Furthermore, from assumptions A and B, sets U and V_k are nonempty closed and convex (see Theorem 2.2 for the convexity property).

Theorem 3.2 *Under assumptions A and B, sequence $\{x^k\}$, generated by the Proximal Method is Fejér convergent to U.*

Proof. Let $x \in U$, then $f(x) \leq f(x^k)$. As $g^k = \lambda_k(x^{k-1} - x^k) \in \widehat{\partial} f(x^k)$ (see Remark 3.3) and f is quasiconvex, using Theorem 2.4 we have

$$\langle x - x^k, x^{k-1} - x^k \rangle \leq 0. \tag{3.8}$$

On the other hand, for all $x \in I\!R^n$, we have

$$\begin{aligned} ||x - x^{k-1}||^2 &= \langle (x - x^k) + (x^k - x^{k-1}), (x - x^k) + (x^k - x^{k-1}) \rangle \\ &= ||x - x^k||^2 + ||x^k - x^{k-1}||^2 - 2\langle x - x^k, x^{k-1} - x^k \rangle \end{aligned}$$

Now, the last identity and (3.8) imply, in particular, for $x \in U$,

$$0 \leq ||x^k - x^{k-1}||^2 \leq ||x - x^{k-1}||^2 - ||x - x^k||^2. \tag{3.9}$$

Thus

$$||x - x^k||^2 \leq ||x - x^{k-1}||^2. \tag{3.10}$$

Therefore, $\{x^k\}$ is Fejér convergent to U. ■

Proposition 3.3 *Under assumptions A and B, the following facts are true*

a. *For all $x \in U$, sequence $\{||x - x^k||\}$ is convergent;*

b. $\lim_{k\to+\infty} ||x^k - x^{k-1}|| = 0$*;*

Proof.
a. From (3.10), $\{||x - x^k||\}$ is a bounded below nonincreasing sequence and hence convergent.
b. Taking limit when k goes to infinity in (3.9) and using the previous result, we obtain $\lim_{k\to+\infty} ||x^k - x^{k-1}|| = 0$, as desired. ∎

Theorem 3.3 *Suppose that Assumptions A and B are satisfied. Then the sequence $\{x^k\}$ converges to a point of U.*

Proof. From Theorem 3.2, $\{x^k\}$ is Fejér convergent to U, thus $\{x^k\}$ is bounded (see Theorem 2.5). Then, there exist $\bar{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to $\bar{x}$. From the lower semicontinuity of f we obtain

$$\liminf_{j\to+\infty} f(x^{k_j}) \geq f(\bar{x}).$$

As $\{f(x^k)\}$ is decreasing and converges then

$$f(\bar{x}) \leq \lim_{k\to+\infty} f(x^k) \leq f(x^k), \forall k.$$

This implies that $\bar{x} \in U$. Now, from Theorem 2.5 we conclude that $\{x^k\}$ converges to $\bar{x}$. ∎

To obtain strong results for the proximal method we impose some conditions to parameters $\{\lambda^k\}$ and substitute assumption A by the following

Theorem 3.4 *Suppose that assumptions A and B are satisfied. If $0 < \lambda_k < \bar{\lambda}$, where $\bar{\lambda}$ is a positive real number, then sequence $\{x^k\}$ converges to a point of U and*

$$\lim_{k\to\infty} g^k = 0,$$

for some $g^k \in \widehat{\partial} f(x^k)$. Furthermore, if f is continuous on `domf`*, then it converges to a generalized critical point of f.*

Proof. The convergence was proven in Theorem 3.3. From Remark 3.3 we have that

$$g^k = \lambda_k(x^{k-1} - x^k) \in \widehat{\partial} f(x^k).$$

Taking $||.||$ in the above equation and using the condition in parameter λ_k we have that

$$||g^k|| \leq \bar{\lambda}||x^k - x^{k-1}||.$$

Now, using Proposition 3.3, **b**, we get

$$\lim_{k\to\infty} g^k = 0.$$

Finally, we will prove that $0 \in \partial f(\bar{x})$ when f is continuous. In fact, $\{x^k\}$, $\{f(x^k)\}$ and $\{g^k\}$ with $g^k \in \widehat{\partial} f(x^k)$ such that $\lim_{k\to+\infty} x^k = \bar{x}$, $\lim_{k\to\infty} f(x^k) = f(\bar{x})$ (from the continuity of f) and $\lim_{k\to\infty} g^k = 0$. From Definition 2.2 it follows that $0 \in \partial f(\bar{x})$. ∎

As immediate particular cases of the above theorem we obtain the following

Corollary 3.1 *Let $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ be a convex and continuous function on* **domf** *and assume that assumption A' is satisfied, then sequence $\{x^k\}$ converges to an optimal solution of the problem (p).*

Corollary 3.2 *Let $f : I\!R^n \to I\!R \bigcup \{+\infty\}$ be a pseudoconvex and continuous function on* **domf** *and assume that the assumption A' is satisfied, then sequence $\{x^k\}$ converges to an optimal solution of the problem (p).*

Proof. From the previous theorem we have $0 \in \partial f(\bar{x})$. Now, due to the pseudoconvexity of f we obtain that $f(\bar{x}) \leq f(x)$, for all $x \in I\!R^n$. ■

Assumption A'. The optimal set of the problem (p), denoted by X^*, is nonempty. Next, we give a condition to obtain the global convergence to a global minimum of (p).

Theorem 3.5 *Suppose that assumptions A' and B are satisfied. If x^k are global minimums of $f(.) + (\lambda_k/2)||. - x^{k-1}||^2$ and*

$$\lim_{k \to +\infty} \lambda_k = 0, \tag{3.11}$$

then the sequence $\{x^k\}$ converges to a global minimum of the problem (p).

Proof. As x^k is a minimizer point of $f + (\lambda_k/2)||. - x^{k-1}||^2$, then

$$f(x^k) + (\lambda_k/2)||x^k - x^{k-1}||^2 \leq f(x) + (\lambda_k/2)||x - x^{k-1}||^2, \ \forall x \in I\!R^n.$$

As $\{x^k\}$ converges to a point of U, let $\bar{x} \in U$ such that $\lim_{k \to +\infty} x^k = \bar{x}$. From above inequality, we have in particular that

$$f(x^k) + (\lambda_k/2)||x^k - x^{k-1}||^2 \leq f(x) + (\lambda_k/2)||x - x^{k-1}||^2, \ \forall x \in U.$$

Taking $k \to +\infty$ and using (3.11), the lower semicontinuity of f, Proposition 3.3, **a** and **b**, we obtain that

$$f(\bar{x}) \leq f(x), \forall x \in U.$$

This implies that $\bar{x}$ is an optimal solution of the problem (p). ■

3.2 A Theoretical Localization of the Limit Point

Let $\pi_U : I\!R^n \to I\!R^n$ be such that

$$\pi_U(x^0) = \arg\min\{||x - x^0|| : x \in U\},$$

where U is defined as in the previous subsection. Note that $\pi_U(x^0)$ exists because U is a nonempty closed convex set. We denote

$$\rho(x^0, U) = \inf\{||x - x^0|| : x \in U\}.$$

Proposition 3.4 *Let $\{x^k\}$ be the sequence generated by the proximal point method and $\bar{x}$ the limit point, then*

$$||x^0 - \bar{x}|| \leq 2\rho(x^0, U).$$

Proof. Setting $x = \pi_U(x^0)$ in (3.10) we obtain $||\pi_U(x^0) - x^k|| \leq ||\pi_U(x^0) - x^0||$. Taking $k \to \infty$ gives

$$||\pi_U(x^0) - \bar{x}|| \leq ||\pi_U(x^0) - x^0||, \tag{3.12}$$

Now,

$$\begin{aligned} ||x^0 - \bar{x}||^2 &\leq (||x^0 - \pi_U(x^0)|| + ||\pi_U(x^0) - \bar{x}||)^2 \\ &\leq 2\left(||x^0 - \pi_U(x^0)||^2 + ||\pi_U(x^0) - \bar{x}||^2\right) \\ &\leq 2\left(||x^0 - \pi_U(x^0)||^2 + ||x^0 - \pi_U(x^0)||^2\right) \\ &= 4||x^0 - \pi_U(x^0)||^2 \end{aligned}$$

where the third inequality is due to (3.12). Thus, the proof is completed. ■

Remark 3.4 *If $\bar{x}$ is a global minimum of (p), for example when f is convex, or pseudoconvex, the above result remain true if we substitute U by X^*.*

4 Proximal Point Method on the Nonnegative Orthant

Now, we are interested in solving the problem

$$\min\{f(x) : x \geq 0\} \tag{4.13}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper function such that $\mathtt{dom} f \cap \mathbb{R}^n_+ \neq \emptyset$ and $x \geq 0$ means that each component of x, x_i, is non negative.

In this section we define and study the full convergence of a proximal point method with a generalized like-distance so called *proximal distance* that includes as particular cases known distances as Bregman, $\varphi-$divergences and second-order homogeneus distances.

4.1 Proximal Distance

In this subsection we present the definition of the proximal distance and induced proximal distance, introduced by Auslender and Teboulle in [6], but adapted for the set $\mathbb{R}^n_+$ which is the contraint of the problem (4.13).

Definition 4.1 *A function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is called proximal distance as regards the open nonempty convex set $\mathbb{R}^n_{++}$ if for each $y \in \mathbb{R}^n_{++}$ it satisfies the following properties:*

- **i.** *$d(., y)$ is proper, lower semicontinuous, strictly convex and continuously differentiable on $\mathbb{R}^n_{++}$;*
- **ii.** *$\mathtt{dom}\, d(., y) \subset \mathbb{R}^n_+$ and $\mathtt{dom}\partial_1 d(., y) = \mathbb{R}^n_{++}$, where $\partial_1 d(., y)$ denotes the classical subgradient map of the function $d(., y)$ with respect to the first variable;*
- **iii.** *$d(., y)$ is coercive on $\mathbb{R}^n$, i.e., $\lim_{||u|| \to \infty} d(u, y) = +\infty$;*
- **iv.** *$d(y, y) = 0$.*

We denote by $D(\mathbb{R}^n_{++})$ the family of functions satisfying this definition.

Property **i**. is needed to preserve convexity of $d(.,y)$, property **ii** will force the iteration of the proximal method to stay in $I\!R^n_{++}$, and **iii** will be used to guarantee the existence of the proximal iterations. For each $y \in I\!R^n_{++}$, let $\nabla_1 d(.,y)$ denote the gradient map of the function $d(.,y)$ with respect to the first variable. Note that by definition $d(.,.) \geq 0$ and from **iv** the global minimum of $d(.,y)$ is obtained at y, which shows that $\nabla_1 d(y,y) = 0$.

Associated to a proximal distance to be an induced proximal distance which we define as follows:

Definition 4.2 *Given $d \in D(I\!R^n_{++})$, a function $H : I\!R^n \times I\!R^n \to I\!R_+ \cup \{+\infty\}$ is called the induced proximal distance to d if H is a finite valued on $I\!R^n_{++} \times I\!R^n_{++}$ and for each $a, b \in I\!R^n_{++}$ satisfies*

(Ii) $H(a,a) = 0$.

(Iii) $\langle c - b, \nabla_1 d(b,a) \rangle \leq H(c,a) - H(c,b), \quad \forall\, c \in I\!R^n_{++}$.

We write $(d,H) \in \mathcal{F}(I\!R^n_{++})$ to the proximal and induced proximal distance that satisfies the premises of Definition 4.2.
We also denote $(d,H) \in \mathcal{F}(I\!R^n_{+})$ if there exists H such that:

(Iiii) H is finite valued on $I\!R^n_{+} \times I\!R^n_{++}$ satisfying (**Ii**) and (**Iii**), for each $c \in I\!R^n_{+}$.

(Iiv) For each $c \in I\!R^n_{+}$, $H(c,\cdot)$ has level bounded sets on $I\!R^n_{++}$.

Finally, we write $(d,H) \in \mathcal{F}_+(I\!R^n_{+})$ if

(Iv) $(d,H) \in \mathcal{F}(I\!R^n_{+})$.

(Ivi) $\forall\, y \in I\!R^n_{+}$ and $\forall\, \{y^k\} \subset I\!R^n_{++}$ bounded with $\lim_{k\to+\infty} H(y,y^k) = 0$, then $\lim_{k\to+\infty} y^k = y$.

(Ivii) $\forall\, y \in I\!R^n_{+}$, and $\forall\, \{y^k\} \subset I\!R^n_{++}$ such that $\lim_{k\to+\infty} y^k = y$, then $\lim_{k\to+\infty} H(y,y^k) = 0$.

The mean result on proximal point method will be when $(d,H) \in \mathcal{F}_+(I\!R^n_{+})$. Some examples of proximal distances which satisfy the above definitions, for example Bregman distances, proximal distances based on $\varphi-$divergences, self-proximal distances, and distances based on second order homogeneus proximal distances, were given in [6].

4.2 Proximal Method

In this section we propose a extension of the proximal point method with a proximal distance to solve the problem (4.13). It is as follows:
Given a sequence of positive parameters $\{\lambda_k\}$ and a initial point

$$x^0 \in I\!R^n_{++}. \tag{4.14}$$

For each $k = 1, 2....$, if $0 \in \hat{\partial} f(x^{k-1})$, then stop. Otherwise, find $x^k \in I\!R^n$ such that

$$0 \in \hat{\partial}(f(.) + \lambda_k d(., x^{k-1}))(x^k), \tag{4.15}$$

where d is a proximal distance such that $(d,H) \in \mathcal{F}_+(I\!R^n_{+})$.

Remark 4.1 *Observe that the proposed algorithm is an extension (for nonconvex functions) of the proximal point method. In fact, if f is convex then the iterations become*

$$x^k = \mathit{arg}\min_{x \geq 0}\{f(x) + \lambda_k d(x, x^{k-1})\},$$

Particular cases of this proximal method for minimizing smooth quasiconvex functions were studied by Souza et al. [27] with a class of separated Bregman distances, Cunha et al. [13] and Chen and Pan [14] with a particular $\phi-$divergence distance and Pan and Chen [25] with a second-order homogeneous distance for solving non-smooth quasiconvex functions, which includes the Logarithmic-Quadratic proximal point method introduced by Auslender Teboulle and Ben-Tiba in [7] and [8].

Remark 4.2 *As we are interested in solving (4.13) when f is nonconvex it is important to observe (differently from previous research works) that the algorithm only needs, in each iteration, to find a stationary point (and not a global minimum) of the regularized function $f(.)+\lambda_k d(., x^{k-1})$ so we believe that local algorithms can be used satisfactorily in each iteration.*

Theorem 4.1 *If f is bounded from below and lower semicontinuous, then the sequence $\{x^k\}$, generated by proximal method (4.14)-(4.15) is well defined and furthermore $x^k \in I\!R^n_{++}$.*

Proof. The existence of $\{x^k\}$ is immediate from the lower boundedness and lower semicontinuity of f and coercivity of $d(., y)$. The positiveness of x^k is immediate from the definition of d. ■

We are interested in analyzing the convergence of the proximal method when f is quasiconvex, so we list our assumptions:

Assumption A**.** $f : I\!R^n \to I\!R \cup \{+\infty\}$ is a proper function bounded from below.

Assumption B**.** $f : I\!R^n \to I\!R \cup \{+\infty\}$ is lower semicontinuous and quasiconvex.

As we are interest in the asymptotic convergence of the method, we also assume that in each iteration $0 \notin \widehat{\partial} f(x^k)$ which implies that $x^k \neq x^{k-1}$, for all k.

Remark 4.3 *From (4.15), Proposition 2.1,* **d***, and the smoothness of $\lambda_k d(., x^{k-1})$, this implies that*

$$0 \in \widehat{\partial} f(x^k) + \lambda_k \nabla_1 d(x^k, x^{k-1}).$$

Thus, there exists $g^k \in \widehat{\partial} f(x^k)$ such that

$$g^k = -\lambda_k \nabla_1 d(x^k, x^{k-1}).$$

Proposition 4.1 *Under assumptions A and B we have that $\{f(x^k)\}$ is decreasing and converges.*

Proof. As $x^k \neq x^{k-1}$ and $d(., x^{k-1})$ is strictly convex then

$$\langle \nabla_1 d(x^k, x^{k-1}) - \nabla_1 d(x^{k-1}, x^{k-1}), x^k - x^{k-1} \rangle > 0 \tag{4.16}$$

As $\nabla_1 d(x^{k-1}, x^{k-1}) = 0$, this implies that

$$\langle -\lambda_k \nabla_1 d(x^k, x^{k-1}), x^{k-1} - x^k \rangle > 0.$$

From Remark 4.3, we obtain

$$\langle g^k, x^{k-1} - x^k \rangle > 0$$

Using the quasiconvexity of f and Theorem 2.4, this implies that

$$f(x^k) < f(x^{k-1}).$$

The convergence of $\{f(x^k)\}$ is immediate from the lower boundedness of f. ■

Proposition 4.2 *The iterates $\{x^k\}$ do not cycle.*

Proof. Suppose that there exists $l > j+1$ such that $x^l = x^j$. From Proposition 4.1 we have that

$$f(x^j) = f(x^l) < ... < f(x^{j+1}) < f(x^j),$$

which is a contradiction, so $x^l \neq x^j$. ■

Now, we define the following set

$$U_+ := \{x \in \texttt{dom}\, f \cap I\!R^n_+ : f(x) \leq \inf_{j \in I\!N} f(x^j)\}.$$

Observe that this set depends on the choice of the initial iterates x^0 and the sequence $\{\lambda_k\}$. If $U_+ = \emptyset$ then it can be easily proven that

i) $\lim_{k \to +\infty} f(x^k) = \inf_{x \geq 0} f(x)$,

ii) $\{x^k\}$ is unbounded.

From now on we assume that $U_+ \neq \emptyset$. From the assumptions we obtain that U_+ is a nonempty, closed and convex set (see Theorem 2.2 for the convexity property).

Theorem 4.2 *Under assumptions A and B, sequence $\{x^k\}$, generated by the proximal method, (4.14)-(4.15), is $H-$ Fejér convergent to U_+, that is,*

$$H(x, x^k) \leq H(x, x^{k-1}), \forall x \in U_+.$$

Proof. Let $x \in U_+$, then

$$f(x) \leq f(x^k). \tag{4.17}$$

From Remark 4.3, (4.17) and using Theorem 2.4 we obtain

$$\langle -\lambda_k \nabla_1 d(x^k, x^{k-1}), x - x^k \rangle \leq 0,$$

that is,

$$\langle \nabla_1 d(x^k, x^{k-1}), x - x^k \rangle \geq 0,$$

Then, using property **Iii** of Definition 4.2 we have

$$H(x, x^k) \leq H(x, x^{k-1}). \quad ■$$

Corollary 4.1 *Under assumptions A and B, sequence $\{x^k\}$ is bounded.*

Proof. From the previous theorem we deduce

$$H(x, x^k) \leq H(x, x^0) = \alpha, \forall x \in U_+.$$

This implies that

$$x^k \in L_H(x, \alpha) = \{y \in I\!R^n_+ : H(x, y) \leq \alpha\},$$

and from Definition 4.2, **Iiv**, $L_H(x, \alpha)$ is bounded. Thus $\{x^k\}$ is bounded. ■

Theorem 4.3 *Under assumptions A and B, sequence $\{x^k\}$ converges to some point of U_+.*

Proof. From previous Corollary $\{x^k\}$ is bounded, then there exists a subsequence $\{x^{k_j}\}$ which converges to $\bar{x}$, that is

$$\lim_{j\to+\infty} x^{k_j} = \bar{x}.$$

As f is decreasing and converges we have $\bar{x} \in U$. Suppose that there exists another sequence $\{x^{k_l}\}$ such that

$$\lim_{l\to+\infty} x^{k_l} = x^* \in U_+.$$

Using property **Ivii**, from Definition 4.2, we obtain

$$\lim_{l\to+\infty} H(x^*, x^{k_l}) = 0.$$

Due to $\{H(x^*, x^k)\}$ is non-increasing and bounded from below then it converges and therefore

$$\lim_{k\to+\infty} H(x^*, x^k) = 0.$$

Thus

$$\lim_{j\to+\infty} H(x^*, x^{k_j}) = 0.$$

Using property **Ivi**, from Definition 4.2, we obtain that

$$\lim_{j\to+\infty} x^{k_j} = x^*,$$

that is, $\bar{x} = x^*$. Then sequence $\{x^k\}$ converges to some point of U_+. ■

To obtain strong results for the proximal method with proximal distances we impose some conditions to parameter $\{\lambda^k\}$ and substitute assumption A for the following

Assumption A'. The optimal set of the problem (p), denoted by X^*, is nonempty.

Theorem 4.4 *Suppose that assumptions A' and B are satisfied. If d is a self-proximal distance, $\{x^k\}$ are global minimums of function $f(.) + \lambda_k d(., x^{k-1})$ and*

$$\lim_{k\to+\infty} \lambda_k = 0, \tag{4.18}$$

then the sequence $\{x^k\}$ converges to a global minimum of the problem (4.13).

Proof. As x^k is a global minimizer of $f + \lambda_k d(., x^{k-1})$, then

$$f(x^k) + \lambda_k d(x^k, x^{k-1}) \leq f(x) + \lambda_k d(x, x^{k-1}), \forall x \in I\!R^n_+.$$

Given that d is a self-proximal distance then

$$f(x^k) + \lambda_k d(x^k, x^{k-1}) \leq f(x) + \lambda_k H(x, x^{k-1}), \forall x \in I\!R^n_+. \tag{4.19}$$

As $\{x^k\}$ converges to a point of U_+, let $\bar{x} \in U_+$ such that $\lim_{k\to+\infty} x^k = \bar{x}$. From the above inequality, we have in particular that

$$f(x^k) + \lambda_k d(x^k, x^{k-1}) \leq f(\bar{x}) + \lambda_k H(\bar{x}, x^{k-1}).$$

Thus

$$0 \leq \lambda_k d(x^k, x^{k-1}) \leq f(\bar{x}) - f(x^k) + \lambda_k H(\bar{x}, x^{k-1})$$

Taking lim inf when $k \to +\infty$ and using (4.18), the lower semicontinuity of f, and Theorem 4.2 (the sequence $\{H(x, x^k)\}$ is bounded for each $x \in U_+$) we obtain that $\liminf_{k\to+\infty} \lambda_k d(x^k, x^{k-1}) = 0$. Finally, taking lim inf when $k \to +\infty$ in (4.19) we obtain

$$f(\bar{x}) \leq f(x), \forall x \in U.$$

This implies that $\bar{x}$ is an optimal solution of the problem (4.13). ■

To obtain further information on the limit point we restrict the class of proximal distances given in Definition 4.1 to a special class of separable proximal distances, that is,

$$d(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i),$$

where $d_i : I\!R \times I\!R \to I\!R \cup \{+\infty\}$ such that if there exist i with

$$\frac{\partial d_i}{\partial x_i}(x_i, y_i) > 0 \Rightarrow x_i > y_i. \tag{4.20}$$

Remark 4.4 *Observe that the well known Bregman separable, $\varphi-$divergence and second order homogeneous distances satisfied the above property.*

Theorem 4.5 *Suppose that assumptions A and B are satisfied, d is separable and holds (4.20). If $0 < \lambda_k < \bar{\lambda}$, for all k and some $\bar{\lambda} > 0$, then the sequence $\{x^k\}$ converges to a generalized KKT point of the problem (4.13), that is, the limit point $\bar{x}$ satisfies*

i). $\bar{x}_i \geq 0, \forall i = 1, ..., n;$

ii). $\liminf_{k\to+\infty} g_i^k \geq 0, \forall i = 1, ..., n;$

iii). $\liminf_{k\to+\infty} x_i^k g_i^k \geq 0, \forall i = 1, ..., n.$

Proof. From the previous theorem, there exists $\bar{x} \in U_+$ such that

$$\lim_{k\to+\infty} x^k = \bar{x}.$$

Due to Theorem 4.1 the condition **i**) is satisfied. To prove the other conditions consider the following sets

$$I(\bar{x}) = \{i \in \{1, ..., n\} : \bar{x}_i = 0\},$$
$$J(\bar{x}) = \{i \in \{1, ..., n\} : \bar{x}_i > 0\}.$$

We will prove **ii**). Let $i \in I(\bar{x})$ and suppose that there exists $\beta > 0$ such that

$$\liminf_{k\to\infty} g_i^k := \sup_{n\in I\!N} \inf_{n\leq k} g_i^k < -\beta < 0.$$

That is,

$$\inf_{n\leq k} g_i^k < -\beta, \ \forall n \in I\!N.$$

Then, there exists a subsequence $\{g_i^{k_n}\}$ such that

$$g_i^{k_n} < -\beta.$$

From Remark 4.3 that implies that

$$\frac{\partial d_i}{\partial x_i}(x_i^{k_n}, x_i^{k_n-1}) > 0,$$

from assumption (4.20) we have

$$x_i^{k_n} > x_i^{k_n-1} > ... > x_i^{k_1} > 0.$$

Taking $n \to +\infty$ we obtain that $0 > 0$, which is a contradiction. Therefore $\liminf_{k\to+\infty} g_i^k \geq 0$, for $i \in I(\bar{x})$.
Now let $i \in J(\bar{x})$. From Remark 4.3 we have

$$\lim_{k\to\infty} g_i^k = \lim_{k\to\infty} -\lambda_k \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1}). \tag{4.21}$$

As $x_i^{k-1} \to \bar{x}_i > 0$, d is continuously differentiable on $I\!R^n_{++}$ and $\nabla_1 d(x,x) = 0$ we have

$$\lim_{k\to\infty} \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1}) = \frac{\partial d_i}{\partial x_i}(\bar{x}_i, \bar{x}_i) = 0.$$

Now, using in (4.21) the boundedness of $\{\lambda_k\}$ and the above result we have $\lim_{k\to\infty} g_i^k = 0$, for $i \in J(\bar{x})$.
Finally, we prove **iii**). Let $i \in I(\bar{x})$, and, by contradiction, suppose that $\liminf_{k\to+\infty} x_i^k g_i^k < 0$. Then for each $n \in I\!N$ we have

$$\inf_{k\geq n} x_i^k g_i^k < 0.$$

Thus, there exists a subsequence $\{x_i^{k_n} g_i^{k_n}\}$ such that

$$x_i^{k_n} g_i^{k_n} < 0.$$

As $x_i^k > 0$ then $g_i^{k_n} < 0$. Then from Remark 4.3 we obtain that

$$\frac{\partial d_i}{\partial x_i}(x_i^{k_n}, x_i^{k_n-1}) > 0,$$

from assumption we have

$$x_i^{k_n} > x_i^{k_n-1} > ... > x_i^{k_1} > 0.$$

Taking $n \to +\infty$ we obtain that $0 > 0$, which is a contradiction. Therefore $\liminf_{k\to+\infty} g_i^k x_i^k \geq 0$, for $i \in I(\bar{x})$.
Finally, let $i \in J(\bar{x})$, then from the proof of **ii** we have that $\lim_{k\to+\infty} g_i^k = 0$, so we conclude that $\lim_{k\to+\infty} g_i^k x_i^k = 0$.
■

Remark 4.5 *From* **ii** *of the previous theorem and Remark 4.3 we obtain that*

$$\limsup_{k\to+\infty} \left(\lambda_k \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1}) \right) \leq 0$$

hols for each $i = 1, 2, ..., n$.

Corollary 4.2 *Under the assumptions of the previous theorem and the continuous differentiability of f we have that sequence $\{x^k\}$ converges to a KKT point of the problem (4.13). Furthermore, if assumption A' is satisfied and there exists $i \in I(\bar{x})$ such that*

$$\limsup_{k\to+\infty}\left(\lambda_k \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1})\right) < 0,$$

then $\{x^k\}$ converges to an optimal solution of the problem (4.13).

Proof. Let $\{x^k\}$ such that

$$\lim_{k\to\infty} x^k = \bar{x}.$$

From the previous theorem and the continuously differentiability of f we obtain that $\bar{x}$ satisfies

$$\bar{x} \geq 0, \quad \nabla f(\bar{x}) \geq 0, \text{and } (\nabla f(\bar{x}))_i \bar{x}_i = 0,$$

Thus, $\bar{x}$ is a KKT point of the problem (4.13).
Now, we suppose that

$$\limsup_{k\to+\infty}\left(\lambda_k \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1})\right) < 0.$$

This implies that

$$\liminf_{k\to+\infty}\left(-\lambda_k \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1})\right) > 0.$$

From Remark 4.3 we obtain

$$\liminf_{k\to+\infty} g_i^k > 0.$$

As f is continuously differentiable and from Proposition 2.1, **b**, we have

$$\liminf_{k\to+\infty} \frac{\partial f}{\partial x_i}(x^k) > 0.$$

As $x^k \to \bar{x}$ and $\frac{\partial f}{\partial x_i}$ is continuous we obtain

$$\frac{\partial f}{\partial x_i}(\bar{x}) > 0.$$

Now using item **a** of Theorem 2.6, for the case when $g(x) = -x$, we obtain that $\bar{x}$ is an optimal solution of the problem (4.13). ■

Remark 4.6 *The same ideas used in the proof of the previous theorem can be used to extend the convergence result of the continuous method studied by Attouch and Teboulle [4]. Indeed, to solve the problem (4.13) in [4] an itroduction was made of the dynamic system*

$$x_j'(t) + \left[\frac{x_j(t)}{\mu + \nu x_j(t)}\right] \frac{\partial f(x(t))}{\partial x_j} = 0, \quad j = 1, ..., p \tag{4.22}$$

$$x_j(0) = x_{0_j} \in I\!R^p_{++}, \quad j = 1, .., p. \tag{4.23}$$

Under the following assumption:

(H_1) *$f : I\!R^n \to I\!R$ is a quasiconvex and C^2 function;*

(H_2) *For any* $x_0 \in I\!R^n_{++}$*, there exists a global solution* $x : [0, +\infty) \to I\!R^n_{++}$ *of the dynamic system (4.22)-(4.23);*

it has been proven that there exists $x_\infty \in Z := \{x \in I\!R^n_+ : f(x) \le \inf_{t\ge 0} f(x(t))\}$ *(assumed nonnempty), such that*

$$\lim_{t\to\infty} x(t) = x_\infty.$$

Now, using the same arguments given in Theorem 4.5 we will prove that x_∞ *is a KKT-point of the problem, that is,*

$$(x_\infty)_i \ge 0,\ (\nabla f(x_\infty))_i \ge 0 \ \text{ and } \ (x_\infty)_i(\nabla f(x_\infty))_i = 0 \tag{4.24}$$

In fact, due to Theorem 4.1 the first condition in (4.24) is satisfied. To prove the other conditions, consider the following sets:

$$I(x_\infty) = \{i \in \{1, ..., n\} : (x_\infty)_i = 0\},$$

$$J(x_\infty) = \{i \in \{1, ..., n\} : (x_\infty)_i > 0\}.$$

Let us spearately consider the cases when $i \in I(x_\infty)$ *and* $i \in J(x_\infty)$.
First Case: $i \in I(x_\infty)$**.** *We will show that* $(\nabla f(x_\infty))_i \ge 0$*. By contradiction, suppose that* $(\nabla f(x_\infty))_i < 0$*. Because* f *is continuously differentiable, we have*

$$\lim_{t\to\infty} (\nabla f(x(t))_i = (\nabla f(x_\infty))_i < 0.$$

Hence,$\nabla f(x(t))_i < 0$ *for* t *sufficiently large. Thus, from (4.22), we obtain*

$$\lim_{t\to\infty} \nabla f(x(t))_i = \lim_{t\to\infty} -x_i'(t)\left(\frac{\mu + \nu x_i(t)}{x_i(t)}\right) < 0.$$

Thus we have

$$x_i'(t)\left(\frac{\mu + \nu x_i(t)}{x_i(t)}\right) > 0,$$

for t *sufficiently large. As* $x_i(t) > 0$*, the last inequality implies that*

$$x_i'(t) > 0,$$

for t *sufficiently large. Thus* $x_i(t)$ *is increasing for* t *sufficiently large, that is, there exists* $M > 0$*, such that* $\forall t > t_1 > M$ *we have*

$$0 < x_i(t_1) < x_i(t).$$

Taking $t \to +\infty$ *we obtain that* $0 < x_i(t_1) \le \lim_{t\to+\infty} x_i(t) = (x_\infty)_i = 0$*, which is a contradiction. Therefore,* $(\nabla f(x_\infty))_i \ge 0$*.*

Second Case: $i \in J(x_\infty)$**.** *We will show that* $(\nabla f(x_\infty))_i = 0$*. In fact, since from the dynamical system (4.22) we have*

$$(\nabla f(x_\infty))_i = \lim_{t\to+\infty} \nabla f(x(t))_i = \lim_{t\to+\infty} -x_i'(t)\left(\frac{\mu + \nu x_i(t)}{x_i(t)}\right).$$

Given that $x_i(t)$ *converges to* $(x_\infty)_i > 0$ *and* $\lim_{t\to+\infty} -x_i'(t) = 0$ *(see Theorem 4.1, (c), of Attouch and Bolte [2]) we obtain that* $(\nabla f(x_\infty))_i = 0$*.*

5 Conclusion and Some Discusion

Our approach contributes to the progress in the eficient solution of minimization problems with quasiconvex objective functions using an extension of the proximal point method. Previous and recent results turn out to be particular cases of our general approach as also the full convergence for a large class of proximal distances, which was not studied previously, is obtained. Furthermore, for the proximal point method on the nonnegative orthant we give a sufficient condition on the proximal distance to obtain the convergence of the method to an optimal solution of the problem. In future research we hope to improve our obtained results, with a better condition on $\{\lambda_k\}$ to attain the convergence of the proximal point method for an optimal solution of the problem.

References

[1] P.A. Absil, R. Mahony and B. Andrews, Convergence of the Iterates of Descent Methods for Analytic Cost Function, SIAM J. Optim., 16, 531-547, 2005.

[2] H. Attouch, and J. Bolte, On the Convergence of the Proximal Algorithm for Nonsmooth Functions Involving Analytic Features, Mathematical Programming, Vol 116, 1-2, Ser. B, 5-16, 2009.

[3] K.J., Arrow and A., Enthoven, Quasi-Concave Programming, Econometrica, Vol 29, N 4, October, 1961.

[4] H. Attouch, and M. Teboulle, Regularized Lotka-Volterra Dynamical System as Continuous Proximal-Like Method in Optimization, Journal Optimiation Theory and Applications, 121, pp. 541-580, 2004.

[5] H. Attouch, A. Soubeyran, Worthwhile to Move Behaviors and Inertia: Temporary Satisficing Process with not too Much Sacrificing, in revision, Journal of Mathematical Psychology, 2007.

[6] A. Auslender, M. Teboulle, Interior Gradient and Proximal Methods for Convex and Conic Optimization, SIAM J. Optim., Vol 16, n. 3, pp. 697-725, 2006.

[7] A. Auslender, M. Teboulle and S. Ben-Tiba, Interior Proximal and Multiplier Methods Based on Second Order Homogeneous Functionals, Math Oper. Research 24, 3, pp.645-668, 1999.

[8] A. Auslender, M. Teboulle and S. Ben-Tiba, A Logarithmic-Quadratic Proximal Method for Variational Inequalities, Computational Optimization and Applications 12, 31-40, 1999.

[9] N. Barron and W. Liu, Calculus of Variation l^{∞}, Applied Math. Optim. 35, 237-263, 1997.

[10] M.S. Bazaara, H.D. Sherali, and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 2nd Edition, John Wiley and Sons, New York, NY, 1993.

[11] R.S. Burachik, L.M. Graña Drummond, A.N. Iusem and B.F. Svaiter, Full Convergence of the Steepest Descent Method with Inexact Line Searches. Optimization, **32**, pp. 137-146, 1995.

[12] Y. Censor, A. Iusem, and S. Zenios, An Interior Point Method Bregman Functions for the Variational Inequality Problem with Paramonotone Operators. Mathematical Programming, 81, 373-400, 1998.

[13] F.G.M. Cunha, J.X. Cruz Neto and P.R. Oliveira, A Proximal Point Algorithm with $\phi-$divergence to Quasiconvex Programming, Acepted to Publication in Optimization.

[14] J.S. Chen and S. Pan, A Proximal-Like Algorithm for a Class of Nonconvex Programming, Pacific Journal of Optimization, Vol 4, pp. 319-333, 2008.

[15] G. Chen and M. Teboulle, Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions, SIAM Journal on Optimization, Vol 3, pp. 538-543, 1993.

[16] O. Guler, On the Convergence of the Proximal Point Algorithm for Convex Minimization, SIAM J. Control and Optimization, 29(2), 403-419, 1991.

[17] J. Gromicho, Quasiconvex Optimization and Location Theory, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998.

[18] J. Eckstein, Nonlinear Proximal Point Algorithms Using Bregman Functions with Applications to Convex Programming, Mathematics of Operation Research, Vol 18, pp. 206-226, 1993.

[19] A. Iusem and M. Teboulle, Convergence Rate Analysis of Nonquadratic Proximal Method for Convex and Linear Programming, Mathematics of Operation Research, vol 20, pp. 657-677, 1995.

[20] A. Iusem, B. Svaiter, and M. Teboulle, Entropy-Like Proximal Methods in Convex Programming, Mathematics of Operation Research, vol 19, pp. 790-814, 1994.

[21] A. Kaplan and R. Tichatschke, Proximal Point Methods and Nonconvex Optimization, Journal of global Optimization, 13, pp. 389-406, 1998.

[22] K.C. Kiwiel, Proximal Minimization Methods with Generalized Bregman Functions, SIAM Journal of Control and Optimization, vol 35, pp. 1142-11268, 1997.

[23] B. Martinet, Régularisation D'inéquations Variationnelles par Approximations Sucessives. (French) Rev. Française Informat. Recherche Opérationnelle 4, Ser. R-3, 154-158, 1970.

[24] J. Palis and W. De Melo, Geometric Theory of Dynamical Systems. An introduction (Translated from the Portuguese by A.K. Manning). Springer-Verlag, New York-Berlin, 1982.

[25] S. Pan and J.S. Chen, Entropy-like Proximal Algorithms Based on a Second-Order Homogeneus distances function for Quasiconvex Programming, Journal of Global Optimization, Vol 39, pp. 555-575, 2007.

[26] E.A. Papa Quiroz and P. Roberto Oliveira, Classical and Logarithmic-Quadratic Proximal Methods for Quasiconvex Functions, Federal University of Rio de Janeiro, Brazil, 2006.

[27] S.S. Souza, P.R. Oliveira, J.X. Cruz Neto and A. Soubeyran, A proximal Point Algorithm with Separable Bregman Distances for Quasiconvex Optimization over the Nonnegative Orthant, Acepted to Publication in European Journal of Operation Research.

[28] A. Takayama, Mathematical Economics, $2nd$ Edition, Cambrigde Univ. Press, 1995.

[29] M. Teboulle, Entropic Proximal Mappings with Application to Nonlinear Programming, Mathematics of Operation Research, vol 17, pp. 670-690, 1992.

[30] M. Teboulle, Convergence of Proximal-Like Algorithms, SIAM Journal on Optimization, vol 7, pp. 1069-1083, 1997.

[31] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Res., 1(2), 97-116, 1976.

[32] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control and Optimization, 14 (5), 877-898, 1976.

[33] R.T. Rockafellar and R. Wets, Variational Analysis. Grundlehren der Mathematischen, Wissenschaften, 317, Springer, 1998.

[34] Tseng, P., Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, Journal of Optimization Theory and Applications (JOTA), 109 (3), 475-494, 2001.