An Inexact Proximal Method for Quasiconvex Minimization

E.A. Papa Quiroz [*], L. Mallma Ramirez and P.R. Oliveira

Federal University of Rio de Janeiro, Brazil
Mayor de San Marcos National University, Perú.
erikpapa@gmail.com

Callao National University, Perú
lenninmr@gmail.com

Federal University of Rio de Janeiro, Brazil
poliveir@cos.ufrj.br

January 10, 2015

Abstract

In this paper we propose an inexact proximal point method to solve constrained minimization problems with locally Lipschitz quasiconvex objective functions. Assuming that the function is also bounded from below, lower semicontinuous and using proximal distances, we show that the sequence generated for the method converges to a stationary point of the problem.

Keywords: Computing science, global optimization, nonlinear programming, Clarke subdifferential, proximal point methods, quasiconvex minimization.

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 to solve convex optimization problems, see Auslender and Teboulle (2006), Burachik and Iusem (1998), Burachik and Scheimberg (2000), Chen and Teboulle (1993), Kiwiel (1997), Rockafellar (1976).

In the last decades a great interest has emerged to extend the PPA for non monotone variational inequalities and non convex minimization problems not only for extending the convergence theory but by several applications in diverse science and engineering areas, see for example the works of Attouch and Bolte (2009); Attouch et al. (2010); Kaplan and Tichatschke (1998); Pennanen (2002); Chen and Pan (2008).

In particular the class of quasiconvex minimization problems has been receiving attention from many researches due to the broad range of applications in location theory, Gromicho (1998), Fractional Programming and specially in economic theory, see for example Takayama

---

[*]Corresponding author: E.A. Papa Quiroz

(1995) and Mas-Colell et al. (1995). Some related papers are the following: Goudou and Munier (2009), Souza et al. (2010), Chen and Pan (2008), Cunha et al. (2010), Pan and Chen (2007), Papa Quiroz and Oliveira (2012), Langenberg and Tichatschke (2012), Brito et. al (2012).

In this paper we are interested in extending the global convergence of an inexact proximal point method to minimize a quasiconvex function constrained on a nonempty closed convex set, that is,

$$\min\{f(x) : x \in \bar{C}\}, \tag{1.1}$$

where $f : I\!\!R^n \to I\!\!R \cup \{+\infty\}$ is a quasiconvex function, C is open convex set on the euclidean space $I\!\!R^n$ and $\bar{C}$ is the closure of C. Obviously, if $C = I\!\!R^n$ we obtain the unconstrained minimization problem. Some convergence results have been recently obtained for some researches:

Attouch and Teboulle (2004), 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; see also Alvarez et al. (2004), that treats a general class of dynamical systems that includes the one of Attouch and Teboulle.

Souza et al. (2010), Cunha et al. (2010), Chen and Pan (2008) and Pan and Chen (2007) studied the iteration:

$$x^k \in \arg\min_{x \in \bar{C}}\{f(x) + \lambda_k d(x, x^{k-1})\}, \tag{1.2}$$

where $\bar{C} = I\!\!R_+^n$, d is a certain distance to force the iterates x^k to stay in C. Some examples of d based on the literature are the class of Bregman, $\varphi-$divergence and second order homogeneous distances. Under the assumption that f is bounded from below, $\text{dom}(f) \cap I\!\!R_+^n \neq \emptyset$, the proximal parameter is bounded and assuming that f is differentiable, these researches obtained the global convergence of the method to a KKT point of (1.1). Furthermore, the sequence generated converges to a solution of the problem, should it exist, if the proximal parameters approach to zero.

Brito et. al (2012), proposed an interior proximal algorithm inspired by the logarithmic-quadratic proximal method for linearly constrained quasiconvex minimization problems. For that method, they proved the global convergence when the proximal parameters go to zero. The latter assumption could be dropped when the function is assumed to be pseudoconvex.

Langenberg and Tichatschke (2012), motivated from the work of Kaplan and Tichatschke (1998), studied the iteration (1.2) when C is an arbitrary open convex set and d is a Bregman distance. Assuming that f is locally Lipschitz and using the Clarke subdifferential, the authors proved the global convergence of the method to a critical point of (1.1).

The above works although important have some disadvantages:

- The main difficulty in extending the proximal method for nonconvex function, which was observed by Kaplan and Tichatschke (1998) and Langenber and Tichatschke (2012), is that due to the nonconvexity of f the subproblems of (1.2) 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. These authors have been proved, under some appropriate conditions and chose a sufficiently large regularization parameters, the strong convexity of the proximal subproblems (and thus efficiently solvable subproblems) for a class of non convex functions, see Theorem 2 and Theorem 5 of Kaplan and Tichatschke (1998) and Langenber and Tichatschke (2012) respectively. However, the above property is not true in general for arbitrary quasiconvex functions. So, we believe that a basic idea is to weaken the condition of minimizing a strongly convex regularization function by another one such that in each iteration we may use local information of the subproblems. If the regularized

function is, for example, locally strongly convex we will have, in certain sense, efficiency in solving the subproblems. This motivate the following question: Is it possible to introduce a local stationary iteration that makes much more sense that the previously considered (1.2) for dealing with nonconvex problems?

- In Rockafellar (1976) it is shown that in some cases let the proximal parameter converges to zero, although the regularizing effect vanishes, provides superlinear convergence of the algorithm in the convex case. Motivated by this fact, Brito et. al (2012), Souza et.al (2010), Cunha et. al (2010), Chen and Pan (2008) and Pan and Chen (2007) have been proved the convergence of the proximal method to an optimal point when the parameters converges to zero. On the other hand, when the proximal parameters are sufficiently greater than zero but bounded from above, Langenber and Tichatschke (2012), see Theorem 9 of that work, proved the convergence to a stationary point which may be in the worst case a saddle point (observe that stationary point or critical point point does not necessarily a global nor local minimum point). This motive the following question: is it possible to obtain a convergence theory to an optimal point when the proximal parameters are bounded from above?

- Despite the fact that the proximal point method is not practical in its exact version, several works, for the convex case, have been shown that it is possible to obtain implementable algorithms with good convergence properties, see for example Alvarez et al. (2010), Liu et al. (2012), Santos and Silva (2014). In our case, for a computational implementation of the proximal point algorithm for the quasiconvex case it is needed to solve the iteration (1.2) using a local optimization algorithm, which only provides an approximate solution. Thus it is important to consider inexact methods. Therefore, from the computational point of view, is it possible to introduce a inexact proximal method to solve (1.2) and prove the convergence of the iteration?

In this paper, motivated by a recently work of Papa Quiroz and Oliveira (2012), we answer the questions of the first and third bullet points and partially we answer the question of the second bullet one, see Subsection 4.3, proposing the following proximal method: given $x^{k-1} \in C$, find $x^k \in C$ and $g^k \in \partial^\circ f(x^k)$ such that

$$\left\| x^{k-1} - x^k - e^k \right\| \leq max \left\{ \left\| e^k \right\|, \left\| x^k - x^{k-1} \right\| \right\} \tag{1.3}$$

where

$$e^k = g^k + \lambda_k \nabla_1 d(x^k, x^{k-1}) \tag{1.4}$$

with ∂° is the Clarke subdifferential and d is a proximal distance, see sections 2 and 3 respectively.

The condition (1.3) is a variant of the inexact proximal method introduced by Solodov and Svaiter (1999) in the general context of proximal algorithm for variational inequalities and studied by Humes and Silva (2005) for the classical proximal point method to solve convex minimization problems. When $e^k = 0$ in (1.4) and $C = \mathbb{R}^n_{++}$ we obtain the exact version studied by Papa Quiroz and Oliveira (2012) for the nonnegative orthant.

Observe also that the conditions (1.3)-(1.4) are more practical than (1.2), where a global minimum point is required in each iteration, and thus more practical than the works of Souza et.al (2010), Cunha et. al (2010), Chen and Pan (2008), Pan and Chen (2007) and Langenberg and Tichatschke (2012). Therefore in our opinion the local stationary iterations (1.3)-(1.4)

3

makes much more sense than the previously considered (1.2) for dealing with nonconvex problems.

Under the assumption that f is proper, lower semicontinuous, locally Lipschitz and bounded from below on $\bar{C}$ and using a class of proximal distance 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 $U_+ := \{x \in \bar{C} : f(x) \leq \inf_{j \geq 0} f(x^j)\}$, assumed nonempty. Then, under the additional conditions that the proximal parameter $\{\lambda_k\}$ is bounded from above and

$$\sum_{k=1}^{+\infty} \frac{\|e^k\|}{\lambda_k} < +\infty \tag{1.5}$$

$$\sum_{k=1}^{+\infty} \frac{|\langle e^k, x^k \rangle|}{\lambda_k} < +\infty, \tag{1.6}$$

we obtain that the sequence $\{x^k\}$ converges to a stationary point of the problem. Observe that the above conditions (1.5)-(1.6) have been used in convex proximal methods, see for example Auslender et. al (1999), Kaplan and Tichatschke (2004), Eckstein (1998), Xu et. al (2006), Solodov and Svaiter (2000). We also get to rid the assumption (1.6) for a class of induced proximal distances which includes Bregman distances given by the standard entropy kernel and all strongly convex Bregman functions.

The paper is organized as follows: In Section 2 we give some basic results on quasiconvex theory and Clarke subdifferential of locally Lipschitz functions. In Section 3 we introduce the class of proximal distances that we will use along the paper. In Section 4 we present the inexact algorithm for solving minimization problems with quasiconvex functions and analyze its convergence properties. Finally, in Section 5 we give our conclusions.

2   Basic Results

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 $X \subset I\!\!R^n$ we denote $bd(X)$ and $\bar{X}$ the boundary and closure of X respectively. Let $B \in I\!\!R^{n \times n}$ be a symmetric and positive definite matrix, we define $\|x\|_B = \langle Bx, x \rangle^{1/2}$.

Given an extended real valued function $f : I\!\!R^n \to I\!\!R \cup \{\pm\infty\}$ we denote its domain by $\text{dom}(f) := \{x \in I\!\!R^n : f(x) < +\infty\}$. f is said to be proper if $\text{dom}(f) \neq \emptyset$ and $\forall x \in \text{dom}(f)$ we have $f(x) > -\infty$.

Finally, f is a lower semicontinuous function if for each $x \in I\!\!R^n$ we have that all $\{x^l\}$ such that $\lim_{l \to +\infty} x^l = x$ implies that $f(x) \leq \liminf_{l \to +\infty} f(x^l)$. It is easy to prove 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 $\alpha \in I\!\!R$.

Lemma 2.1 *Let $\{v_k\}, \{\gamma_k\}$, and $\{\beta_k\}$ be nonnegative sequences of real numbers satisfying $v_{k+1} \leq (1 + \gamma_k) v_k + \beta_k$ and such that $\sum_{k=1}^{\infty} \beta_k < \infty$, $\sum_{k=1}^{\infty} \gamma_k < \infty$. Then, the sequence $\{v_k\}$ converges.*

Proof. See Lemma 2, pp. 44, of Polyak (1987). ∎

Observe that there is a sharper version of the above result, presented as Lemma 7 in Langenberg (2010).

2.1   Fréchet and Limiting Subdifferentials

Definition 2.1 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function.*

(a) *For each $x \in \mathrm{dom}(f)$, the set of regular subgradients (also called Fréchet subdifferential) of f at x, denoted by $\hat{\partial}f(x)$, is the set of vectors $v \in \mathbb{R}^n$ such that*

$$f(y) \geq f(x) + \langle v, y - x \rangle + o(\|y - x\|), \text{ where } \lim_{y \to x} \frac{o(\|y-x\|)}{\|y-x\|} = 0.$$

Or equivalently, $\hat{\partial}f(x) := \left\{ v \in \mathbb{R}^n : \liminf_{y \neq x, \ y \to x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y-x\|} \geq 0 \right\}$. If $x \notin \mathrm{dom}(f)$ then $\hat{\partial}f(x) = \emptyset$.

(b) *The set of general subgradients (also called limiting subdifferential) of f at $x \in \mathbb{R}^n$, denoted by $\partial f(x)$, is defined as follows:*

$$\partial f(x) := \left\{ v \in \mathbb{R}^n : \exists \ x^l \to x, \quad f(x^l) \to f(x), \quad v^l \in \hat{\partial}f(x^l) \text{ and } v^l \to v \right\}.$$

Proposition 2.1 *For a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point $\bar{x} \in \mathrm{dom}(f)$, the subgradient sets $\partial f(\bar{x})$ and $\hat{\partial}f(\bar{x})$ are closed, with $\hat{\partial}f(\bar{x})$ convex and $\hat{\partial}f(\bar{x}) \subset \partial f(\bar{x})$.*

Proof. See Rockafellar and Wets (1998), Theorem 8.6. ∎

Proposition 2.2 *If a proper function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ has a local minimum at $\bar{x} \in \mathrm{dom}(f)$, then $0 \in \hat{\partial}f(\bar{x})$.*

Proof. See Rockafellar and Wets (1998), Theorem 10.1. ∎

Proposition 2.3 *Let $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ proper functions such that f is locally Lipschitz at $\bar{x} \in \mathrm{dom}(f) \cap \mathrm{dom}(g)$ and g is lower semicontinuous function at this point. Then,*

$$\partial(f + g)(\bar{x}) \subset \partial f(\bar{x}) + \partial g(\bar{x})$$

Proof. See Mordukhovich (2006), Theorem 2.33. ∎

2.2   Clarke Subdifferential

Definition 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper function. We say that f is locally Lipschitz if for each $x \in \mathrm{dom}(f)$ there exists $\epsilon_x > 0$ such that*

$$|f(z) - f(y)| \leq L_x \|z - y\|, \forall z, y \in B(x, \epsilon_x),$$

where L_x is some positive number (called the Lipschitz constant of f in a neighborhood of x) and $B(x, \epsilon_x) := \{y \in M : \|x - y\| < \epsilon_x\}$.

Remark 2.1 *From the above definition we obtain that if $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper locally Lipschitz function, then $\mathrm{dom}(f)$ is open in $\mathbb{R}^n$.*

Definition 2.3 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper locally Lipschitz function. Given $x \in \operatorname{dom}(f)$ the generalized directional derivative of f at x in the direction $v \in \mathbb{R}^n$ denoted by $f^\circ(x, v)$, is defined as*

$$f^\circ(x, v) = \limsup_{\substack{t \downarrow 0 \\ y \to x}} \frac{f(y + tv) - f(y)}{t},$$

which is also called the Clarke's generalized directional derivative, see Clarke (1975) or Clarke (1990).

Definition 2.4 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper locally Lipschitz function. The generalized subdifferential of f at $x \in \operatorname{dom}(f)$, denoted by $\partial^\circ f(x)$, is defined by*

$$\partial^\circ f(x) := \{w \in: \mathbb{R}^n : f^\circ(x, v) \geq \langle w, v \rangle, \forall \, v \in \mathbb{R}^n\}.$$

Remark 2.2 *If the proper locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, then $f^\circ(x, v) = f'(x, v) := \lim_{t \downarrow 0} \frac{f(x+tv)-f(x)}{t}$ (respectively, $\partial^\circ f(x) = \partial^F f(x)$, where $\partial^F f$ is the Fenchel Subdifferential of f) for all $x \in \operatorname{dom}(f)$, see Clarke (1990), pp. 40, Proposition 2.3.6.*

Remark 2.3 *From the above definitions it follows directly that for all $x \in \mathbb{R}^n$, one has $\hat{\partial} f(x) \subset \partial f(x) \subset \partial^\circ f(x)$ (see Bolte et al. (2007), Inclusion (7)).*

Definition 2.5 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper locally Lipschitz functions. A point $x \in M$ is said to be a stationary point of f if $0 \in \partial^\circ f(x)$.*

Definition 2.6 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper locally Lipschitz function. The operator $\partial^\circ f$ is called locally weakly monotone iff for each $x \in \operatorname{dom}(\partial^\circ f)$ there exists ϵ_x and $\rho_x > 0$ such that for all $z, y \in B(x, \epsilon_x)$ we have*

$$\langle u - v, z - y \rangle \geq -\rho_x \|z - y\|^2. \tag{2.7}$$

for all $u \in \partial^\circ f(z)$, and for all $v \in \partial^\circ f(y)$

Definition 2.7 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function. f is called quasiconvex if for all $x, y \in \mathbb{R}^n$, 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 a quasiconvex function then $\operatorname{dom}(f)$ 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:

Theorem 2.1 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function. Then, f is quasiconvex if and only if the set $\{x \in \mathbb{R}^n : f(x) \leq c\}$ is convex for each $c \in \mathbb{R}$.*

Proof. See Theorem 3.5.2 of Bazaara et. al (1993). ∎

Lemma 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper locally Lipschitz and quasiconvex function. If $g \in \partial^\circ f(x)$ and $\langle g, z - x \rangle > 0$, then $f(x) \leq f(z)$.*

Proof. See Theorem 2.1 of Aussel (1998) . ∎

2.3 Sufficient Conditions for Quasiconvex Minimization

Consider the problem

$$\min\{f(x) : g(x) \leq 0, x \geq 0\}, \qquad (2.8)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ such that $g(x) = (g_1(x), ..., g_m(x))$ with $g_j : \mathbb{R}^n \to \mathbb{R}$, for each $j = 1, ..., m$ and $x \geq 0$ mean that $x_i \geq 0$, for each $i = 1, ..., n$.

Theorem 2.2 *Let $f, g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, ..., m$, be differentiable quasiconvex functions. If $\bar{x}$ satisfies the KKT necessary condition and one of the following conditions is satisfied:*

a) *$\frac{\partial f}{\partial x_i}(\bar{x}) > 0$, for at least one variable x_i, $i \in \{1, 2, ..., n\}$,*

b) *$\frac{\partial f}{\partial x_i}(\bar{x}) < 0$, for some i such that $\bar{x}_i > 0$,*

c) *$\frac{\partial f}{\partial x_i}(\bar{x}) \neq 0$, and f is twice differentiable in a neighborhood of $\bar{x}$,*

then $\bar{x}$ is a global minimum of the problem (2.8).

Proof. See Theorem 1 of Arrown and Enthoven (1961). ∎

3 Proximal Distance

In this section we present the definitions of proximal distance and induced proximal distance, introduced by Auslender and Teboulle (2006).

Definition 3.1 *A function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is called a proximal distance with respect to an open nonempty convex set $C \subset \mathbb{R}^n$ if for each $y \in C$ it satisfies the following properties:*

  i. *$d(., y)$ is proper, lower semicontinuous, strictly convex and continuously differentiable on C;*

  ii. *$\mathrm{dom}(d(., y)) \subset \bar{C}$ and $\mathrm{dom}(\partial_1^F d(., y)) = C$, where $\partial_1^F d(., y)$ denotes the classical subdifferential 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 $D(C)$ the family of functions satisfying the above definition.

Property **i.** is needed to preserve convexity of $d(., y)$, property **ii** will force the iteration of the proximal method to stay in C, and **iii** will be used to guarantee the existence of the proximal iterations. For each $y \in C$, let $\nabla_1 d(., y)$ denotes 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 $\nabla_1 d(y, y) = 0$.

Definition 3.2 *Given $C \subset \mathbb{R}^n$, open and convex, and $d \in D(C)$, a function $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{+\infty\}$ is called the induced proximal distance to d if H is a finite-valued function on $C \times C$ and for each $a, b \in C$ satisfies*

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

(Iii) $\langle c - b, \nabla_1 d(b,a) \rangle \le H(c,a) - H(c,b), \quad \forall\, c \in C.$

Let us denoted by $(d, H) \in \mathcal{F}(C)$ to the proximal and induced proximal distance that satisfies the conditions of Definition 3.2.
We also denote $(d, H) \in \mathcal{F}(\bar{C})$ if there exists H such that:

(Iiii) H is finite valued on $\bar{C} \times C$ satisfying **(Ii)** and **(Iii)**, for each $c \in \bar{C}$.

(Iiv) For each $c \in \bar{C}$, $H(c, \cdot)$ has level bounded sets on C.

Finally, to establish the global convergence of the proposed method we need to make further assumptions on the induced proximal distance H, mimicking the behavior of the euclidean norm, so we write $(d, H) \in \mathcal{F}_+(\bar{C})$ if

(Iv) $(d, H) \in \mathcal{F}(\bar{C})$.

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

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

The main result on proximal point method will be when $(d, H) \in \mathcal{F}_+(\bar{C})$.
Several 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 homogeneous proximal distances, were given by Auslender and Teboulle (2006), section 3.

Remark 3.1 *The conditions* **(Ivi)** *and* **(Ivii)** *will ensure the global convergence of the sequence generates by the algorithm introduced in subsection 4.1. As we will see in Theorem 4.4, the condition* **(Ivii)** *may be substitute by the following:*

(Iviii) *$H(.,.)$ is continuous in $C \times C$ and if $\{y^k\} \subset C$ such that $\lim_{k \to +\infty} y^k = y \in bd(C)$ and $\bar{y} \ne y$ is another point in $\bar{C}$ then $\lim_{k \to +\infty} H(\bar{y}, y^k) = +\infty$.*

According Kaplan and Tichatschke (2010),pag. 643, the above condition for induced Bregman distances holds when nonlinear constraints are active at y while the condition **(Ivii)** *holds when only affine constraints are active at y.*

Definition 3.3 *Given a symmetric and positive definite matrix $B \in \mathbb{R}^{n \times n}$ and $d \in \mathcal{D}(C)$. We say that d is strongly convex with respect to the first variable and with respect to the norm $\|.\|_B$, if for each $y \in C$ exists $\alpha > 0$ such that*

$$\langle \nabla_1 d(x_1, y) - \nabla_1 d(x_2, y), x_1 - x_2 \rangle \ge \alpha \, \|x_1 - x_2\|_B^2, \forall x_1, x_2 \in C.$$

Examples of proximal distances satisfying the above definition are the second order homogeneous distances, see Brito et al., Lemma 3.1. In particular, for $C = \{x \in \mathbb{R}^n : Ax < b\}$ with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and $rank(A) = n$, the following logarithmic-quadratic proximal distance satisfies the above property

$$d(x, z) = \sum_{i=1}^m \mu \left(y_i(z) y_i(x) - y_i(z)^2 \ln \left(\frac{y_i(x)}{y_i(z)} \right) - y_i(z)^2 \right) + \frac{\nu}{2} \left(y_i(x) - y_i(z) \right)^2$$

where $y_i(x) = b_i - a_i^T x$, $y_i(z) = b_i - a_i^T z$, a_i denotes the row i of the matrix A, b_i the components of b and $\nu > \mu > 0$ with $B = A^T A$ and $\alpha = \nu$. Another class that satisfies the above property is the class of Bregman distances with strongly convex Bregman functions.

Definition 3.4 *Given a symmetric and positive definite matrix $B \in \mathbb{R}^{n \times n}$ and $d \in \mathcal{D}(C)$. We say that d is locally strongly convex with respect to the first variable and with respect to the norm $||.||_B$, if for each $y \in C$ and each $x \in C$ there exists $\epsilon_x > 0$ and $\alpha_x > 0$ such that*

$$\langle \nabla_1 d(x_1, y) - \nabla_1 d(x_2, y), x_1 - x_2 \rangle \geq \alpha_x \|x_1 - x_2\|_B^2, \forall x_1, x_2 \in B(x, \epsilon_x).$$

Lemma 3.1 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper locally Lipschitz function, $d \in D(C)$, $\text{dom}(\partial^\circ f) \cap C \neq \emptyset$ and $B \in \mathbb{R}^{n \times n}$ a symmetric and positive definite matrix. Given an arbitrary point $y \in C$, if $\partial^\circ f$ is a locally weakly monotone operator with constant $\rho > 0$, $d(., y)$ is locally strongly convex with respect to the norm $||.||_B$, with constant α and $\{\beta_k\}$ is a sequence of positive numbers satisfying*

$$\beta_k \geq \beta > \frac{\rho}{\alpha \lambda_{min}(B)},$$

where $\lambda_{min}(B)$ denotes the smallest eigenvalue of B, then $F(.) := \partial^\circ f(.) + \beta_k \nabla_1 d(., y)$ is locally strongly monotone with constant $\beta \alpha \lambda_{min}(B) - \rho$.

proof. Similar to Brito, et al. (2012), Lemma 5.1 ∎

Lemma 3.2 *Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper locally Lipschitz function, if $\partial^\circ f$ is locally Lipschitz then it is locally weakly monotone.*

Proof. It is inmediate. ∎

4 Proximal Interior Method

We are interested in solving the problem defined by:

$$\min\{f(x) : x \in \bar{C}\} \tag{4.9}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous quasiconvex function, $C \subset \mathbb{R}^n$ is a nonempty open convex set satisfying $\text{dom}(f) \cap \bar{C} \neq \emptyset$.

Example 4.1 *(Demand Theory).*
In the classical approach to consumer demand, the analysis of consumer behavior begins specifying the consumer's preference $\succeq$ (an " at-least-as-good-as" binary relation) over the commodity bundles in the consumption set $X \subset \mathbb{R}_+^n$. This preference $\succeq$ is assumed to be rational, that is, $\succeq$ is complete and transitive, see Definition 3.B.1 of Mas-Colell et al. (1995).
A function $\mu : X \rightarrow \mathbb{R}$ is said to be an utility function representing a preference relation $\succeq$ on X if the following condition is satisfied

$$x \succeq y, \text{ if and only if } \mu(x) \geq \mu(y),$$

for all $x, y \in X$. We should observe that the utility function not always exist. In fact, define in $X = \mathbb{R}^2$ a lexicographic relation:

$$\text{For } x, y \in \mathbb{R}^2, x \succeq y \text{ if and only if } x_1 > y_1 \text{ or } (x_1 = y_1 \text{ and } x_2 \geq y_2).$$

Fortunately, a very general class of preference relations can be represented by utility functions, see for example Proposition 3.C.1 of Mas Colell et al. (1995).

If a preference relation $\succeq$ is represented by an utility function μ, then the problem of maximizer the preference of the consumer on X is equivalent to solve the optimization problem

$$(P) \max\{\mu(x) : x \in X\}.$$

On the other hand, a natural psychological assumption in economy is that the consumer tends to diversify his consumption among all goods, that is, the preference $\succeq$ satisfies the following convexity property: X is convex and if $x \succeq z$ and $y \succeq z$ then $\lambda x + (1 - \lambda)y \succeq z, \forall \lambda \in [0, 1]$.
It can be proved that the if there exists an utility function representing the preference, then the convexity property of $\succeq$ is equivalent to the quasiconcavity of the utility function μ. Thus, (P) becomes in a maximization problem with quasiconcave objective function.
Taking $f = -\mu$ we obtain a minimization problem with a quasiconvex objective function on X.
Some examples of quasiconcave functions in economic theory are:

- *The Cobb-Douglas function $\mu : \mathbb{R}_+^n \to \mathbb{R}$ such that $\mu(x) = k \prod_{i=1}^{n} x_i^{\alpha_i}$, where $\alpha_i > 0$, $\forall i = 1, ..., n$ and $k > 0$.*

- *The C.E.S production function $\mu : \mathbb{R}_+^n \to \mathbb{R}$, such that $\mu(x) = k \left[\sum_{i=1}^{n} \delta_i x_i^{-\rho} \right]^{-v/\rho}$ where $\delta_i \in (0, 1), \forall i = 1, 2..., n, \sum_{i=1}^{n} \delta_i = 1, k > 0, v \in (0, 1)$ and $\rho > -1, \rho \neq 0$.*

4.1   The algorithm

To solve (4.9), we propose the following Interior Proximal Point Method (IPPM):

Initialization: Let a sequence of real positive numbers $\{\lambda_k\}$ and an initial point

$$x^0 \in C \tag{4.10}$$

Main Steps: For $k = 1, 2, ...$, and given $x^{k-1} \in C$, find $x^k \in C$ and $g^k \in \partial^\circ f(x^k)$ such that

$$\left\| x^{k-1} - x^k - e^k \right\| \leq max \left\{ \left\| e^k \right\|, \left\| x^k - x^{k-1} \right\| \right\} \tag{4.11}$$

where

$$e^k = g^k + \lambda_k \nabla_1 d(x^k, x^{k-1}) \tag{4.12}$$

and d is a proximal distance such that $(d, H) \in \mathcal{F}(\bar{C})$.
Stop Criterium: If $x^k = x^{k-1}$, or $0 \in \partial^\circ f(x^k)$ then stop. Otherwise to do $k - 1 \leftarrow k$ and return to Main Steps.

Remark 4.1 *The inexact iteration (4.11) is a variant of the one studied by of Humes and Silva (2005), from the geometric point of view this condition means that the side $\|e^k + (x^{k-1} - x^k)\|$ of the triangle is smaller (or equal) than the maximum of $\|x^k - x^{k-1}\|$ and $\|e^k\|$. We will prove that this condition implies that the angle between the vectors $x^k - x^{k-1}$ and e^k is acute and it will be used to assure that the sequence $\{f(x^k)\}$ is non increasing.*

Remark 4.2 *Observe that if $e^k = 0$, then (4.11)-(4.12) becomes: find $x^k \in C$ and $g^k \in \partial^\circ f(x^k)$ such that*

$$0 = g^k + \lambda_k \nabla_1 d(x^k, x^{k-1}),$$

which is more weak than

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

Lemma 4.1 *If the condition (4.11) is satisfied, then $\langle e^k, x^{k-1} - x^k \rangle \geq 0$.*

Proof. From the cosines law and (4.11), we have:

$$\langle e^k, x^{k-1} - x^k \rangle = \frac{1}{2}\left(||x^{k-1} - x^k||^2 + ||e^k||^2 - \left\|x^{k-1} - x^k - e^k\right\|^2\right) \geq 0.$$

Theorem 4.1 *If f is a proper lower semicontinuous function and locally Lipschitz function and lower bounded on $\mathrm{dom}(f) \cap \bar{C}$ and $d \in D(C)$, then the sequence $\{x^k\}$, generated by the proximal method is well defined, that is, x^k exists and $x^k \in C$.*

Proof. We proceed by induction. It holds for $k = 0$ due to (4.10). Assume that x^{k-1} exists and $x^{k-1} \in C$. Define $g(x) = f(x) + \lambda_k d(x, x^{k-1}) + \delta_{\bar{C}}(x)$ and consider the problem

$$\min\{g(x) : x \in \mathbb{R}^n\}.$$

Now, due to the lower boundedness and lower semicontinuity of f, as also, the lower semicontinuity and coercivity of $d(., y)$, there exists $\bar{x} \in \mathbb{R}^n$ (not necessarily unique by the nonconvexity of f) which is a global minimum of g.
From Proposition 2.2

$$0 \in \hat{\partial}\left(f(.) + \lambda_k d(., x^{k-1}) + \delta_{\bar{C}}(.)\right)(\bar{x})$$

Then from Proposition 2.1 and Proposition 2.3 we have

$$0 \in \partial f(\bar{x}) + \partial\left(\lambda_k d(., x^{k-1}) + \delta_{\bar{C}}(.)\right)(\bar{x})$$

Since that $d(., x^{k-1})$ and $\delta_{\bar{C}}(.)$ are proper, lower semicontinuous and convex functions and using Remark 2.3 we obtain that

$$0 \in \partial^\circ f(\bar{x}) + \lambda_k \partial_1^F d(., x^{k-1})(\bar{x}) + \partial^F(\delta_{\bar{C}})(\bar{x})$$

From Definition 3.1, **ii**, we conclude that $\bar{x} \in C$. Finally, taking $x^k = \bar{x}$ there exists $g^k \in \partial^\circ f(x^k)$ such that:

$$0 = g^k + \lambda_k \nabla_1 d(x^k, x^{k-1}),$$

that is, (4.11)-(4.12) is satisfied. ∎

Remark 4.3 *Under the assumptions of the above theorem and if, furthermore, $\partial^0 f$ is locally weakly monotone, $d(., y)$ is strongly convex with respect to the first variable and with respect to the norm $||.||_B$ with constant α, then, from Lemma 3.1, the function*

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

is locally strongly convex for $\lambda_k \geq \lambda > \frac{\rho}{\alpha \lambda_{\min}(B)}$, where ρ is the locally weakly monotone constant.

The above remark provides a class of non convex functions, in particular the class of functions f which $\partial^\circ f$ is locally Lipschitz, see Lemma 3.2, with regularized locally strongly convex functions. Thus, we expect local algorithms to perform efficiently enough to find good approximate solutions at reasonable time. For example, if this regularized function is locally self-concordant, see Nesterov and Nemirovskii (1994), we can use for computing the point x^k satisfying (4.12) the damped Newton method which is commonly recognized as a standard approach particularly efficient in such cases. When the objective function f is nonsmooth we can work with the so called bundle methods.

4.2 Convergence Results

In this section, under some natural conditions, we prove the convergence of the proposed method. We begin imposing the following assumptions:

Assumption A. f is a proper lower semicontinuous quasiconvex function.

Assumption B. f is locally lipschitzian and bounded from below

Assumption C. $\mathrm{dom}(f) \cap \bar{C} \neq \emptyset$.

As we are interest in the asymptotic convergence of the method, we also assume that in each iteration $x^k \neq x^{k-1}$, for all k. If $x^k = x^{k-1}$ then $\nabla_1 d(x^k, x^{k-1}) = 0$ and from (4.12) we have that $e^k \in \partial^\circ f(x^k)$, that is, x^k is an approximation of the critical point of f.

Proposition 4.1 *Under assumptions A, B and C and $d \in D(C)$, we have that $\{f(x^k)\}$ is non increasing and converges.*

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

$$\langle \nabla_1 d(x^k, x^{k-1}) - \nabla_1 d(x^{k-1}, x^{k-1}), x^k - x^{k-1} \rangle > 0 \tag{4.13}$$

As $\nabla_1 d(x^{k-1}, x^{k-1}) = 0, \lambda_k > 0$ and from (4.12) we have

$$\left\langle g^k, x^{k-1} - x^k \right\rangle > \left\langle e^k, x^{k-1} - x^k \right\rangle \geq 0 \tag{4.14}$$

where the last inequality is due to Lemma 4.1. Finally from Lemma 2.2 we obtain

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

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

Now, we define the following set

$$U_+ := \{x \in \bar{C} : f(x) \leq \inf_{j \in I\!N} f(x^j)\}.$$

Observe that this set depends on the choice of the initial iterates x^0 and the sequence $\{\lambda_k\}$. Furthermore, if $U_+ = \emptyset$ then it is easy to prove that $\lim_{k \to +\infty} f(x^k) = \inf_{x \in \bar{C}} f(x)$, and $\{x^k\}$ is unbounded.

From now on we assume that $U_+ \neq \emptyset$. So from the assumption A, we obtain that U_+ is closed and convex set (see Theorem 2.1 for the convexity property).

Proposition 4.2 *Under assumptions A, B and C and $(d, H) \in \mathcal{F}(\bar{C})$, we have*

$$H(x, x^k) \leq H(x, x^{k-1}) - \frac{1}{\lambda_k} \left\langle e^k, x - x^k \right\rangle, \forall x \in U_+$$

Proof. Let $x \in U_+$,then $f(x) \leq f(x^k)$ from Lemma 2.2 we obtain $\left\langle g^k, x - x^k \right\rangle \leq 0$, which from (4.12) implies

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

12

then, using property **Iii** of Definition 3.2 we have

$$\lambda_k^{-1} \left\langle e^k, x - x^k \right\rangle \leq H(x, x^{k-1}) - H(x, x^k), \tag{4.15}$$

that is,

$$H\left(x, x^k\right) \leq H\left(x, x^{k-1}\right) - \frac{1}{\lambda_k} \left\langle e^k, x - x^k \right\rangle \tag{4.16}$$

∎

Proposition 4.3 *Suppose the assumptions A, B, C and $(d, H) \in \mathcal{F}(\bar{C})$. If the following additional conditions hold:*

$$\sum_{k=1}^{+\infty} \frac{\|e^k\|}{\lambda_k} < +\infty \tag{4.17}$$

$$\sum_{k=1}^{+\infty} \frac{|\langle e^k, x^k \rangle|}{\lambda_k} < +\infty \tag{4.18}$$

then

a. *$\{x^k\}$ is Quasi-Féjer convergent to U_+, that is,*

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

where $\epsilon^k = \frac{1}{\lambda_k} \left(\|e^k\| \|x\| + |\langle e^k, x^k \rangle| \right)$ with $\sum_{k=1}^{+\infty} \epsilon^k < +\infty$.

b. *$\{H(x, x^k)\}$ converges $\forall x \in U_+$.*

c. *Furthermore, if $(d, H) \in \mathcal{F}(\bar{C})$, then $\{x^k\}$ is bounded.*

Proof.
a. Using the Cauchy-Schwartz inequality in the right-side of (4.16) we obtain

$$H(x, x^k) \leq H(x, x^{k-1}) + \frac{1}{\lambda_k} \left(\|e^k\| \|x\| + |\langle e^k, x^k \rangle| \right) \tag{4.19}$$

Let $\epsilon^k = \frac{1}{\lambda_k} \left(\|e^k\| \|x\| + |\langle e^k, x^k \rangle| \right)$, then $H\left(x, x^k\right) \leq H\left(x, x^{k-1}\right) + \epsilon^k$, and from the assumptions (4.17) and (4.18) we obtain that $\sum_{k=1}^{+\infty} \epsilon^k < \infty$.

b. It is immediate from **a.** and Lemma 2.2.

c. From **a.** we have

$$H\left(x, x^k\right) \leq H\left(x, x^0\right) + \sum_{k=1}^{+\infty} \epsilon^k \tag{4.20}$$

this implies that $x^k \in L_H(x, \alpha) = \left\{ y \in \bar{C} : H(x, y) \leq \alpha \right\}$, where $\alpha = H\left(x, x^0\right) + \sum_{k=1}^{+\infty} \epsilon^k$. From Definition 3.2, *Iiv*, $L_H(x, \alpha)$ is bounded and thus $\{x^k\}$ is bounded. ∎

Remark 4.4 *It is possible to get rid the assumption (4.18) in Proposition 4.3 for a class of induced proximal distances which includes Bregman distances given by the standard entropy kernel and all strongly convex Bregman functions, see Kaplan and Tichatsche (2004). Of fact, suppose that the induced proximal distance H with $(d, H) \in \mathcal{F}(\bar{C})$ satisfies the following property:*

(Iix) *For each $x \in \bar{C}$ there exist $\alpha(x) > 0$ and $c(x)$ such that:*

$$H(x, v) + c(x) \geq \alpha(x) ||x - v||, \forall v \in C.$$

Then, it is easy to imitate the proof of Proposition 1 from Kaplan and Tichatsche (2004), to obtain that $H(x, x^k)$ is convergent and $\{x^k\}$ is bounded, without the assumption (4.18).

Theorem 4.2 *Under the assumptions of the previous proposition and assuming that $(d, H) \in \mathcal{F}_+(\bar{C})$, the sequence $\{x^k\}$ converges to some point of U_+.*

Proof. From previous lemma $\{x^k\}$ 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 lower semicontinuous, nonincreasing and $\{f(x^k)\}$ converges we have $\bar{x} \in U_+$. Suppose that there exists another sequence $\{x^{k_l}\}$ such that $\lim_{l \to +\infty} x^{k_l} = z \in U_+$. Using property **Ivii**, from Definition 3.2, we obtain $\lim_{l \to +\infty} H(z, x^{k_l}) = 0$, and from the convergence of $\{H(z, x^k)\}$, $\lim_{k \to +\infty} H(z, x^k) = 0$. Thus $\lim_{j \to +\infty} H(z, x^{k_j}) = 0$. Using property **Ivi**, from Definition 3.2, we obtain that $\lim_{j \to +\infty} x^{k_j} = z$, that is, $\bar{x} = z$. ∎

Remark 4.5 *The above result also is true if the condition (**Ivii**) on (d, H) is substitute by (**Iviii**). Of fact, let $\bar{x}$ and z in U_+ as in the proof of the above theorem. Consider two cases: if $\bar{x} \in bd(C)$ then if $z \neq \bar{x}$, from assumption (**Iviii**), $H(z, x^{k_j}) \to +\infty$, but this a contradiction because $z \in U_+$ and the convergence of $H(z, x^k)$. On the other hand, if $\bar{x} \in C$, from the continuity of $H(., .)$ on $C \times C$ we have $H(\bar{x}, x^{k_j}) \to 0$ and from the convergence of $H(\bar{x}, x^k)$ we have that $H(\bar{x}, x^k) \to 0$, now using the condition (**Ivi**) we obtain that $x^k \to \bar{x}$ and thus $z = \bar{x}$.*

Finally, we give the global convergence of $\{x^k\}$ to a stationary point of the problem. We should note that result is an easy adaptation to proximal distances of Theorem 9 of Langenberg and Tichatschke (2012).

Theorem 4.3 *Under assumptions A, B, C, $(d, H) \in \mathcal{F}_+(\bar{C})$, (4.17) and (4.18), and furthermore $\{\lambda_k\}$ is bounded from above, the generated sequence $\{x^k\}$ converges to a stationary point $\bar{x} \in C$ of f, i.e. exists $\bar{g} \in \partial^\circ f(x)$ such that $\forall x \in \bar{C}$ we have*

$$\langle \bar{g}, x - \bar{x} \rangle \geq 0.$$

Proof. Since f is locally Lipschitz at $\bar{x} = lim_{k \to \infty} x^k$ there exists an open neighborhood $B(\bar{x}, \delta)$ and $L > 0$ such that f is Lipschitzian with modulus L on $B(\bar{x}, \delta)$. Thus, there is $k_0 \in I\!N$ such that $x^k \in B(\bar{x}, \delta)$ for any $k \geq k_0$. Hence, $\partial^\circ f(x^k)$ is bounded for $k \geq k_0$ in the sense that $\|g^k\| \leq L$ for all $g^k \in \partial^\circ f(x^k)$ with $k \geq k_0$. Without loss of generality we can assume the convergence $g^k \to \bar{g}$ for some $\bar{g} \in \partial^\circ f(\bar{x})$.
On the other hand,

$$\left\langle g^k, x - x^k \right\rangle = \left\langle e^k, x - x^k \right\rangle - \lambda_k \left\langle \nabla_1 d \left(x^k, x^{k-1} \right), x - x^k \right\rangle$$

In view of (4.17) and that λ_k is bounded from below we obtain that $\langle e^k, x - x^k \rangle \to 0$, so it is sufficient to analyze the convergence of $-\lambda_k \langle \nabla_1 d(x^k, x^{k-1}), x - x^k \rangle$.
From Definition 3.2,**(Iii)**, we obtain that

$$-\lambda_k \left\langle x - x^k, \nabla_1 d \left(x^k, x^{k-1} \right) \right\rangle \geq \lambda_k \left[H(x, x^k) - H(x, x^{k-1}) \right]. \tag{4.21}$$

We analyze two cases:
If $\{H(x, x^k)\}$ converges, then $\lambda_k \left[H(x, x^k) - H(x, x^{k-1}) \right] \to 0$, since $\{\lambda_k\}$ is bounded.

14

If $\{H(x, x^k)\}$ does not converge, then this sequence cannot be monotonically decreasing, since it would otherwise be convergent as a nonnegative sequence. Thus, there are infinitely many $k \in \mathbb{N}$ such that $H\left(x, x^k\right) \geq H\left(x, x^{k-1}\right)$. Let $\{k_l\} \subset \mathbb{N}$ be a subsequence with $H\left(x, x^{k_l}\right) \geq H\left(x, x^{k_l-1}\right)$ for all $l \in \mathbb{N}$, then there exists $\bar{g} \in \partial^\circ f(\bar{x})$ such that

$$\langle \bar{g}, x - \bar{x} \rangle = \lim_{l \to \infty} \left\langle g^{k_l}, x - x^{k_l} \right\rangle = \limsup_{l \to \infty} \left\langle g^{k_l}, x - x^{k_l} \right\rangle \geq \lambda_k \left[H(x, x^{k_l}) - H(x, x^{k_l-1}) \right] \geq 0. \quad \blacksquare$$

Theorem 4.4 *Under assumptions A, B, C, $(d, H) \in \mathcal{F}(\bar{C})$, satisfying the condition* (**Ivi**) *and* (**Iviii**) *instead of* (**Ivii**), *(4.17) and (4.18), and furthermore $\{\lambda_k\}$ is bounded from above, the generated sequence $\{x^k\}$ converges to a stationary point $\bar{x} \in C$ of f.*

Proof. Its is immediate from Remark 4.5 and the proof of the above theorem. $\quad \blacksquare$

4.3 Differentiable Case

In the last subsection we prove the convergence of the sequence $\{x^k\}$ to a stationary point of f on $\bar{C}$ which may be in the worst case a saddle point. In this subsection we give some conditions to guarantee the convergence of the above sequence to a global minimum point. Consider the following minimization problem:

$$\min\{f(x) : g_j(x) \leq 0; x_i \geq 0; i = 1, ..., n; j = 1, ..., m\} \tag{4.22}$$

where $f, g_j : \mathbb{R}^n \to \mathbb{R}$ are quasiconvex, f is continuously differentiable and g_j are differentiable. Suppose that f is bounded from below, $(d, H) \in \mathcal{F}_+(\bar{C})$, and $\{\lambda_k\}$ bounded from above, then from Theorem 4.3 we have that the sequence $\{x^k\}$ converges to a point $\bar{x} \in \mathbb{R}^n$ which is a KKT point of the problem (4.22), that is, there exist $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{s} \in \mathbb{R}^n$ such that:

$$\nabla f(\bar{x}) + \sum_{j=1}^{m} \bar{\lambda}_i \nabla g_j(\bar{x}) - \bar{s} = 0,$$

$$g_j(\bar{x}) \leq 0,$$
$$\bar{\lambda}_j g_j(\bar{x}) = 0,$$
$$\bar{s}_i \bar{x}_i = 0,$$
$$\bar{s}_i, \bar{x}_i \geq 0.$$

Corollary 4.1 *Consider the problem (4.22) and suppose that $f, g_j : \mathbb{R}^n \to \mathbb{R}$ are quasiconvex, f is continuously differentiable and g_j are differentiable functions. Suppose also that f is bounded from below, $(d, H) \in \mathcal{F}_+(\bar{C})$, $\{\lambda_k\}$ is bounded from above. If one of the following condition is satisfied:*

a) $\lim_{k \to \infty} \lambda_k \frac{\partial d}{\partial x_i}(x^k, x^{k-1}) < 0$, for some i

b) $\lim_{k \to \infty} \lambda_k \frac{\partial d}{\partial x_i}(x^k, x^{k-1}) > 0$, for some i such that $\bar{x}_i > 0$

c) $\lim_{k \to \infty} \lambda_k \frac{\partial d}{\partial x_i}(x^k, x^{k-1}) \neq 0$, and f is twice differentiable in a neighborhood of $\bar{x}$

then, $\bar{x}$ is a global minimum of (4.22).

Proof. From Remark 4.12

$$\lambda_k \frac{\partial d}{\partial x_i}(x^k, x^{k-1}) = e_i^k - \frac{\partial f}{\partial x_i}(x^k)$$

as $e^k \to 0$ and from Theorem 2.2 we have the aimed result. $\quad \blacksquare$

5   Conclusion and Future Works

- The main motivation of the paper was to extend the global convergence properties of the inexact proximal point method from the convex case to the quasiconvex one. However, when f is not convex and considering the errors criterion from the literature it was difficult, at least for us, to obtain the claim. So we introduce the following criterion, which is new even for the convex case,

$$\left\| x^{k-1} - x^k - e^k \right\| \leq max \left\{ \left\| e^k \right\|, \left\| x^k - x^{k-1} \right\| \right\}.$$

Then, to ensure that the sequence $\{x^k\}$ satisfy the Quasi-Féjer convergent property we impose the following conditions

$$\sum_{k=1}^{+\infty} \frac{\left\| e^k \right\|}{\lambda_k} < +\infty \tag{5.23}$$

and

$$\sum_{k=1}^{+\infty} \frac{|\langle e^k, x^k \rangle|}{\lambda_k} < +\infty. \tag{5.24}$$

We get also to eliminate the condition (5.24) for a class of induced proximal distance following the paper of Kaplan and Tichatschke (2004) who presented an analysis how to get rid the assumption (5.24) for a class of Bregman functions which includes the standard entropy kernel and all strongly convex Bregman functions. They showed also the convergence of the proximal point method with the logarithmic-quadratic distance, obtained by Auslender et al. (1999b), as also of a particular entropy-like distance without the condition (5.24) but using strongly the monotonicity (convexity for the minimization case) of the operator. For the quasiconvex case, we can see that the extension is not immediate and requires further analysis, so we may consider this research as a future work.

- If the objective function of (4.9) is proper, lower semicontinuous and quasiconvex function but it is not locally Lipschitz, we can introduce the following method: given $y^{k-1} \in C$, find y^k and $s^k \in \widehat{\partial} f(y^k)$ such that

$$\left\| y^{k-1} - y^k - e^k \right\| \leq max \left\{ \left\| e^k \right\|, \left\| y^k - y^{k-1} \right\| \right\}$$

where

$$e^k = s^k + \lambda_k \nabla_1 d(y^k, y^{k-1})$$

with $\widehat{\partial}$ is the set of general subgradients (also called Fréchet subdifferential), see Definition 2.1 of Papa Quiroz and Oliveira (2012).
In this case, all the results from Remark 4.1 to Theorem 4.2 can be easily adapted and thus to obtain the convergence of the sequence $\{y^k\}$ to a point $\bar{y} \in V_+$, where

$$V_+ := \left\{ x \in \bar{C} : f(x) \leq \inf_{j \in I\!N} f(y^j) \right\}.$$

Unfortunately, $\widehat{\partial} f$ does not locally bounded at $\bar{y}$, see Theorem 9.13 of Rockafellar and Wets (1998), and so the extension of Theorem 4.3 is an open question.

- Other future works may be the extension of this method to solve multiobjective quasi-convex minimization problems as also to solve quasiconvex vector optimization problems.

- In modern optimization, proximal algorithms are important as a basic block for structured optimization. They can handle the non-smooth part of the objective function or constraints, see for example the forward-backward algorithm, alternating proximal minimization algorithms. In this perspective it would be interesting to study these algorithms in the quasiconvex case.

References

Alvarez F, Bolte J, Brahic O (2004) Hessian riemannian gradient flows in convex programming. SIAM J. Optim., 43, 2: 477-501

Alvarez F, López J, Ramírez C (2010) Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and support vector machines, Optimization Methods Software 25, 6, 859-881

Arrow KJ, Enthoven AC (1961) Quasi-concave programming. Econometrica, 29, 4: 779-800

Attouch H, Bolte J (2009) On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Mathematical Programming, 116, 1-2, Ser. B: 5-16.

Attouch H, Bolte J, Svaiter B (2010) Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Mathematical Programming, Ser A, online first doi:10.1007/s10107-011-0484-9.

Attouch H, Teboulle M (2004) Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization. Journal Optimization Theory and Applications, 121: 541-580

Auslender A, Teboulle M (2006) Interior gradient and proximal methods for convex and conic optimization. SIAM J. Optim., 16, 3: 697-725, 2006

Auslender A, Teboulle M, Ben-Tiba S (1999) Interior proximal and multiplier methods based on second order homogeneous functionals. Math Oper. Research 24, 3: 645-668

Auslender A, Teboulle M, Ben-Tiba S (1999b) A logarithmic-quadratic proximal method for variational inequalities. Computational Optimization and Applications, 12, 31-40.

Aussel, D (1998) Subdifferential properties of quasiconvex and pseudoconvex functions: unified approach. J. Optim Theory Appl. 97, 1: 29-45.

Bazaara MS, Sherali HD, Shetty CM (1993) Nonlinear programming: theory and algorithms, 2nd Edition, John Wiley and Sons, New York, NY

Bolte J, Daniilidis A, Lewis A and Shiota, M (2007) Clarke subgradients of stratifiable functions. SIAM Journal on Optimization. 18,556-572.

Brito AS, da Cruz Neto JX, Lopez JO, Oliveira PR (2012) Interior proximal algorithm for quasiconvex programming and variational inequalities with linear constraints, J Optim Theory Appl, 154: 217-234.

Burachik RS, Iusem A (1998) Generalized proximal point algorithm for the variational inequality problem in Hilbert space. SIAM, Journal of Optimization, 8: 197-216.

Burachik RS, Scheimberg S (2000) A proximal point method for the variational inequality problem in Banach spaces. SIAM J. Control Optim.,39, 5: 1633-1649.

Cunha FGM, da Cruz Neto JX, Oliveira PR (2010) A proximal point algorithm with $\phi-$divergence to quasiconvex programming, 59: 777-792

Chen JS, Pan S (2008) A proximal-like algorithm for a class of nonconvex programming. Pacific Journal of Optimization, 4: 319-333

Chen G, Teboulle M (1993) Convergence analysis of a proximal-like minimization algorithm using Bregman functions, 3: 538-543

Clarke FH (1975) Generalized gradients and applications. Transaction of the American Mathematical Society, 205: 247-262.

Clarke FH (1990) Optimization and nonsmooth analysis. New York: Wiley.

Eckstein J (1998) Approximate iterations in Bregman-function-based proximal algorithms. Mathematical Programming 83: 113-123

Goudou X, Munier J (2009) The gradient and heavy ball with friction dynamical system: the quasiconvex case, Mathematical Programming, Serie B, 116: 173-191

Gromicho J (1998) Quasiconvex optimization and location theory. Kluwer Academic Publishers. Dordrecht, the Netherlands

Humes Jr C, Silva PJS (2005) Inexact proximal algorithms and descent methods in optimization. Optimization and Engineering, 6: 257-271.

Kaplan A, Tichatschke R (1998) Proximal point methods and nonconvex optimization. Journal of Global Optimization, 13: 389-406.

Kaplan A, Tichatschke R (2004) On inexact generalized proximal methods with a weakened error tolerance criterion, Optimization 53: 3-17

Kiwiel KC (1997) Proximal minimization methods with generalized Bregman functions. SIAM Journal of Control and Optimization, 35: 1142-11268

Langenberg N, (2010) Pseudomonotone operators and the Bregman proximal point algorithm. J Glob Optim 47: 537-555.

Langenberg N, Tichatschke R (2012) Interior proximal methods for quasiconvex optimization. J Glob Optim 52: 641-661.

Liu Yong-Jin, Sung Defeng, Toh Kim-Chuan (2012) An implementable proximal point algorithmic framework for nuclear norm minimization, Mathematical Programming, Ser A, 133, 399-436.

Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory, Oxford University Press: New York, NY, USA; 1995.

Mordukhovich, B.S. (2006) Variational analysis and generalized differentiation I: Basic theory. Grundlehren Series[Fundamental Principles of Mathematical Sciences], vol. 330, Springer-Verlag, Berlin.

Nesterov Y.E. and Nemirovskii A.N. (1994) Interior-Point Polynomial Algorithms in Convex Programming, Philadelphia, Society for Industrial and Applied Mathematics.

Pan S, Chen JS (2007) Entropy-like proximal algorithms based on a second-order homogeneous distances function for quasiconvex programming. Journal of Global Optimization, 39: 555-575

Papa Quiroz EA, Oliveira PR (2012) An extension of proximal methods for quasiconvex minimization on the nonnegative orthant. European Journal of Operational Research 216: 26-32.

Pennanen T (2002) Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. oper. Res. 27: 170-191.

Polyak BT (1997) Introduction to optimization, optimization software, New York.

Rockafellar RT (1976) Monotone operations and the proximal point method. SIAM J. Control Optimization, 14: 877-898.

Rockafellar RT, Wets R (1998) Variational analysis. Grundlehren der Mathematischen, Wissenschaften, 317, Springer.

Santos SA, Silva RC (2014), An inexact and nonmonotone proximal method for smooth unconstrained minimization, Journal of Computational and Applied Mathematics 269, 2014, 86-100.

Solodov MV, Svaiter BF (1999) A hybrid projection-proximal point algorithm. Journal of

Convex Analysis 6: 59-70.

Solodov MV, Svaiter BF (2000) An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. Mathematics of Operations Research 25: 214-230

Souza SS, Oliveira PR, da Cruz Neto JX, Soubeyran A (2010) A proximal method with separable Bregman distance for quasiconvex minimization on the nonnegative orthant. European Journal of Operational Research, 201: 365-376

Takayama A (1995) Mathematical economics, $2nd$ Edition, Cambridge Univ. Press

Xu Y, Bingsheng H, Xiaoming Y (2006) A hybrid inexact logarithmic-quadratic proximal method for nonlinear complementarity problems. J. Math. Anal. 322: 276-287