Generalized Proximal Point Algorithms for Quasiconvex Programming

Arnaldo Silva Brito* P. Roberto Oliveira† Jurandir O. Lopes ‡ J.X. da Cruz Neto §

Abstract

In this paper, we proposed algorithms interior proximal methods based on entropy-like distance for the minimization of the quasiconvex function subjected to nonnegativity constraints. Under the assumptions that the objective function is bounded below and continuously differentiable, we established the well definedness of the sequence generated by the algorithms and obtained two important convergence results, the principal one is a sufficient condition for the convergence point of the sequence generated by the algorithms is a point of solution of the problem.

Keywords: Interior proximal methods, entropy-like distance, quasiconvex programming.

1 Introduction

Consider the quasiconvex minimization problem

$$\begin{array}{lll} \text{(P)} & \min & f(x) \\ & \text{s.t.} & x \geq 0. \end{array} \tag{1.1}$$

where $f : I\!R^n \to I\!R \cup \{\infty\}$ is a closed proper quasiconvex function.
The quasiconvex minimization problem has many applications in Economics (e.g. see [11]), microeconomy (utility function, e.g. see [20]), location theory (e.g. see [13]), approximation theories (fractional programming, e.g. see [4]) etc.
The classical proximal point algorithm to minimize a convex function f on $I\!R^n$ generates a sequence $\{x^k\}$ through the iterative scheme: Start with an initial point $x^0 \in I\!R^n$ and find

$$x^{k+1} = \arg \min_{x \in I\!R^n} \{f(x) + \lambda_k \|x - x^k\|^2\}, \tag{1.2}$$

*Partially supported by CAPES. PESC-COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. E-mail: arnaldo@cos.urfj.br

†Corresponding author. PESC/COPPE - Programa de Engenharia de Sistemas e Computação Cidade Universitária, Centro de Tecnologia, Bloco H, Sala 319 C. P. 68511, 21941-972 Tel: +55 21 2562-8672 / +55 21 2562-8673 Fax: +55 21 2562-8676 Rio de Janeiro - RJ - Brazil E-mail: poliveir@cos.ufrj.br

‡Partially supported by CNPq, Federal University of Piauí, Teresina, Brazil. E-mail: jurandir@ufpi.br

§Partially supported by CNPq GRANT 301625-2008, Federal University of Piauí, Teresina, Brazil. E-mail: jxavier@ufpi.br

where $\{\lambda_k\}$ is a sequence of positive real numbers and $\|\cdot\|$ denote the Euclidean norm $I\!R^n$. This method, first proposed by Martinet [19] and subsequently studied by Rockafellar [24], Guller [14] and many others.

Many researchers have attempted to replace the quadratic term in (1.2) by distance-like function [1, 3, 8, 17, 26, 27]. Usually, Bregman function and φ-divergence distance are considered to obtain iterates lying within the constraint set, see for example [8], [27] and their references; another interesting interior proximal family is the so-called logarithmic-quadratic method studied in [3]. However, as few researches exist regarding the quasiconvex case, we highlight the recent study in [25] where the distance of Bregman, [9] and [10] works with the φ-divergence distance. Both work with the following entropy-like distance, also called φ-divergence.

$$d_\varphi(x,y) := \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right) \tag{1.3}$$

where $\varphi : I\!R_+ \to I\!R$ is the closed proper strictly convex function $\varphi(t) = t - \ln t - 1$, for all $t > 0$. This paper deals with two classes of functions Φ_1 and Φ_2, defined in the next section. We propose the algorithms $A1$, $A2$ and $A3$ based on the interior proximal algorithm, where for all $\varphi \in \Phi_1$ we replace the quadratic term in (1.2), by d_φ is defined in (1.3), and since we do not require the convexity of f, the algorithm is given by:

$$\begin{array}{l} x^0 > 0 \\ x^{k+1} \in \arg\min\{f(x) + \lambda_k d_\varphi(x, x^k)\}, \end{array} \tag{1.4}$$

where $\{\lambda_k\}$ is sequence of positive numbers. Now, for $\varphi \in \Phi_2$ use (1.4) with d_φ defined by

$$d_\varphi(x,y) := \sum_{i=1}^{n} y_i^2 \varphi\left(\frac{x_i}{y_i}\right). \tag{1.5}$$

The algorithm $A3$ works with D_h, the Bregman distance induced by h, defined by

$$D_h(x,y) := h(x) - h(y) - \langle \nabla h(y), x - y\rangle. \tag{1.6}$$

where h defined in the next section. Use (1.4) replaced d_φ by D_h defined by (1.6).

Regarding the algorithm $A1$, retrieve and expand the works of [9] and [10], because we work not only with a function $\varphi(t) = t - \ln t - 1$, but with a class Φ_1 which contains this function, and also established, under certain assumption, a sufficient condition for the convergence of our algorithm to an optimal solution to the problem (P). A condition which does not appear in [9], [10] and will be applied in [25] with the algorithm $A3$.

In relation to algorithm $A2$, we obtain similar results to those obtained by the algorithm A1 and we have as a particular case the logarithmic-quadratic proximal distance [2].

This paper is organized as it follows: after this introduction, we present some definitions and results about the quasiconvex functions, the definition of φ-divergence and the definition of Bregman distance, and some of their properties. In section 3, we define the algorithms and show its well-definedness. In section 4, convergence analysis . Finally, in section 5, we set the conclusions.

We will use the following notation throughout this paper. $I\!R^n$ denotes the n-dimensional Euclidean space. The Euclidean inner product is denoted by $\langle ., . \rangle$ and $\| \ . \ \|$ indicates the Euclidean norm. The i th component of vector $x \in I\!R^n$ is denoted by $x_i \ \forall \ i = 1, ..., n$. $I\!R^n_+$ and $I\!R^n_{++}$ represent, respectively, the non-negative orthant of $I\!R^n$ and its interior, i.e., $I\!R^n_+ = \{(x_1, ..., x_n) \in I\!R^n : x_i \geq 0, \ i = 1, ..., n\}$ and $I\!R^n_{++} = \{(x_1, ..., x_n) \in I\!R^n : x_i > 0, \ i = 1, ..., n\}$. The gradient of f in x is denoted by $\nabla f(x)$ and the ith partial derivative of f in relation to x is represented by $(\nabla f(x))_i$. The notation $\nabla_x f(x, y)$ indicates the partial derivative of f in relation to its first component.

2 Preliminary

In this section, we recall some preliminary results that will be used in the next section. We start with the definitions of quasiconvex functions, properties and some of theirs characterizations. This theory can be found in [18] and their references.

Definition 2.1 *A function $f : I\!R^n \to I\!R$ is said quasiconvex if for every $x, y \in R^n$ and for every $\alpha \in [0, 1]$ the following inequality holds:*

$$f((1 - \alpha)x + \alpha y) \leq \max\{f(x), f(y)\}. \tag{2.7}$$

When the strict inequality occurs in (2.7) the function f is said to be strictly quasiconvex. It is immediate that a convex function is quasiconvex and the domain of a quasiconvex function is convex.
For a given $\lambda \in R^n$ the level (respectively, strict level) set of f, corresponding to λ, is the set:

$$S_f(\lambda) := \{x \in R^n : f(x) \leq \lambda\},$$

respectively,

$$S^<_f(\lambda) := \{x \in R^n : f(x) < \lambda\}.$$

The next proposition characterizes the quasiconvex functions.

Proposition 2.1 *A function $f : I\!R^n \to I\!R$ is quasiconvex if and only if $S_f(\lambda)$ is a convex set for all $\lambda \in R$.*

Proof: See [18]. □

Proposition 2.2 *Assume that $f : I\!R^n \to I\!R$ is differentiable. Then f is quasiconvex if and only if for all $x, y \in R^n$ such that*

$$f(x) < f(y) \Rightarrow \langle \nabla f(y), x - y \rangle \leq 0. \tag{2.8}$$

Proof: See [18]. □

The following definition of pseudoconvex function can be found in [18].

Definition 2.2 *A function $f : I\!R^n \to I\!R$ is said to be pseudoconvex if for any $x, x_0 \in I\!R^n$, one has*

$$f(x) < f(x_0) \Rightarrow \langle \nabla f(x_0), x - x_0 \rangle < 0.$$

We relate now pseudoconvex functions to and quasiconvex functions.

Lemma 2.1 *Assume that $f : I\!R^n \to I\!R$ is quasiconvex and differentiable. If $\nabla f(x) \neq 0$ and $f(y) < f(x)$, then $\langle \nabla f(x), y - x \rangle < 0$.*

Proof: Since f is continuous, there exists $\delta > 0$ such that $z \in B(y, \delta)$ implies $f(z) < f(x)$, where $B(y, \delta) := \{w \in I\!R^n \ : \| y - w \| < \delta\}$.
Take

$$\hat{z} = y + \frac{1}{2\delta} \frac{\nabla f(x)}{\| \nabla f(x) \|} \in B(y, \delta) \quad \text{so} \quad f(\hat{z}) < f(x)$$

by the Proposition 2.2 , it follows that $\langle \nabla f(x), \hat{z} - x \rangle \leq 0$, which means that

$$\langle \nabla f(x), y - x \rangle + \frac{\| \nabla f(x) \|}{2\delta} \leq 0$$

consequently,

$$\langle \nabla f(x), y - x \rangle \leq -\frac{\| \nabla f(x) \|}{2\delta} < 0.$$

□

An interesting observation is that if f is a quasiconvex and differentiable function and $\nabla f(x) = 0$, x is not necessarily a minimum of f as it can be seen for instance in $f : I\!R \to I\!R$ given by $f(x) = x^3$ at point $x = 0$.
We now present the definition of φ-divergence function and some of their properties as used in the context of optimization.
Let $\varphi : I\!R \to I\!R \cup \{\infty\}$ be a closed proper convex function. We denote its domain by dom$\varphi := \{t : \varphi(t) < \infty\} \neq \emptyset$ with dom$\varphi \subset [0, \infty)$. We assume that φ satisfies the following.

(i) φ is twice continuously differentiable on int(dom φ)$= (0, \infty)$,

(ii) φ is strictly convex on its domain,

(iii) $\lim_{t \to 0^+} \varphi'(t) = -\infty$,

(iv) $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$.

We denote by Φ the class of functions satisfying (i) - (iv). For any $\varphi \in \Phi$, since $argmin\{\varphi(t) : t \in I\!R\} = \{1\}$ φ is coercive. Consider two subclasses of Φ , defined by.

$$\Phi_1 := \{\varphi \in \Phi; \varphi''(1) \left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1) \log t \quad \forall t > 0\} \tag{2.9}$$

$$\Psi := \{\varphi \in \Phi; \varphi''(1) \left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1)(t-1) \quad \forall t > 0\} \tag{2.10}$$

Examples of function in $\Phi_1 \cap \Psi$ are:

$$\varphi_1(t) = t\log t - t + 1, \qquad dom\varphi_1 = [0, +\infty)$$

$$\varphi_2(t) = -\log t + t - 1, \qquad dom\varphi_2 = (0, +\infty)$$

$$\varphi_3(t) = 2(\sqrt{t} - 1)^2, \qquad dom\varphi_3 = [0, +\infty)$$

This first time, we fixed our attention to the subclass Φ_1.

Definition 2.3 *If $\varphi \in \Phi_1$, then $d_\varphi : I\!R^n_{++} \times I\!R^n_{++} \to I\!R$ defined by*

$$d_\varphi(x, y) := \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right) \tag{2.11}$$

is said to be a $\varphi - divergence$.

Denoting by $\bigtriangledown_1$ the gradient concerning the first variable, it holds that $[\bigtriangledown_1 d_\varphi(x, y)]_i = \varphi'\left(\frac{x_i}{y_i}\right)$ for all $i = 1, ..., n$. From the strict convexity of φ and (iv) we obtain,

$$\varphi(t) \geq 0 \qquad \forall t \geq 0 \qquad \varphi(t) = 0 \Leftrightarrow t = 1.$$

Hence, d_φ satisfies

$$d_\varphi(x, y) \geq 0 \quad (x, y) \in I\!R^n_{++} \times I\!R^n_{++}, \quad and$$
$$d_\varphi(x, y) = 0 \Leftrightarrow x = y.$$

Consider the function $\varphi(t) = \varphi_1(t) = t\log t - t + 1$, we have

$$d_\varphi(x, y) := H(x, y) = \sum_{i=1}^{n} (x_i \log \frac{x_i}{y_i} + y_i - x_i) \tag{2.12}$$

which is the so-called Kullback-Leibler relative entropy distance functional. $H(x, y)$ can be continuously extended to $I\!R^n_{+} \times I\!R^n_{++}$, adopting the convention the $0\log 0 = 0$. In order words, H admits a point with zero component in its first argument. The next result shows some properties of H.

Lemma 2.2 *Let H defined in (2.12), we have:*

(i) *The level sets of $H(x, \cdot)$ are bounded for all $x \in I\!R^n_{+}$,*

(ii) *If $\{y^k\} \subset I\!R^n_{++}$ converge to $y \in I\!R^n_{+}$, then $\lim_{k\to\infty} H(y, y^k) = 0$,*

(iii) *If $\{z^k\} \subset I\!R^n_{+}, \{y^k\} \subset I\!R^n_{++}$ are such that $\{z^k\}$ is bounded, $\lim_{k\to\infty} y^k = y \in I\!R^n_{+}$ and $\lim_{k\to\infty} H(z^k, y^k) = 0$, them $\lim_{k\to\infty} z^k = y$,*

(iv) *For any $w, z \in I\!R^n_{++}$ and $v \in I\!R^n_{+}$. If $\varphi \in \Phi_1$, then $\langle \bigtriangledown_1 d_\varphi(z, w), v - z\rangle \leq \varphi''(1)[H(v, w) - H(v, z)]$.*

Proof: The proof (i) - (iii) is elementary using (2.12) and (iv) see [27] lemma 4.1 (ii). $\square$

Now, consider the class Ψ. Let Φ_2 the class of closed proper convex functions $\varphi : R_+ \to R \cup \{+\infty\}$ is given by

$$\varphi(t) := \mu h(t) + \left(\frac{\nu}{2}\right)(t-1)^2. \tag{2.13}$$

with $\nu > \mu h''(1) > 0$, $h \in \Psi$ and let the associated proximal distance be defined by

Definition 2.4 *For a given $\varphi \in \Phi_2$, the distance-like function $d_\varphi : R^n \times R^n \to R \cup \{+\infty\}$ is defined by*

$$d_\varphi(x,y) := \begin{cases} \sum_{i=1}^n y_i^2 \varphi(x_i/y_i), & x, y \in R^n_{++} \\ +\infty, & \textit{otherwise} \end{cases} \tag{2.14}$$

In particular, $h(t) = \varphi_2(t) = -\log t + t - 1$ gives the so-called logarithmic - quadratic proximal distance [2].
Considering d_φ given in (2.14) we have that, $[\nabla_1 d_\varphi(x,y)]_i = y_i \varphi'(\frac{x_i}{y_i})$ for all $i = 1, ..., n$.

The following lemma will be used in convergence analysis. Defined by

$$\theta := \nu + \rho.\mu, \quad \tau := \nu - \rho.\mu \quad \rho := h''(1)$$

Lemma 2.3 *For all $w, z \in I\!R^n_{++}$ and $v \in I\!R^n_+$, then $\langle \nabla_1 d_\varphi(w,z), w - v\rangle \geq \frac{\theta}{2}(\| w - v \|^2 - \| w - z \|^2) + \frac{\tau}{2} \| w - z \|^2$.*

Proof: See [3]. $\square$

We focus our attention on the properties of a given function $\varphi \in \Phi_2$ and the induced function $d_\varphi(\cdot,\cdot)$, which will be used in the subsequent analysis. Initially, we summarize some special properties of φ. Since their verifications are directed by computations, we omit the details.

Proposition 2.3 *Let $\varphi \in \Phi_2$. Then, the following results hold.*

(i) $\varphi(t) \geq 0$ *e* $\varphi(t) = 0$ *if and only if* $t = 1$,

(ii) $\varphi(t)$ *is decreasing in* $(0,1)$ *with* $\lim_{t\to 0^+} \varphi(t) = \infty$, *and increasing in* $(1,\infty)$ *with* $\lim_{t\to\infty} \varphi(t) = \infty$,

(iii) $\varphi(1) = \varphi'(1) = 0$, *and* $\varphi''(1) > 0$,

(iv) $\varphi'(t)$ *is nondecreasing on* $(0,\infty)$ *and* $\lim_{t\to\infty} \varphi'(t) = 1, \lim_{t\to 0^+} \varphi'(t) = -\infty$.

From the strict convexity of $\varphi \in \Phi_2$ and property (iii) by the last proposition, φ satisfies

$$\varphi(t) \geq 0 \quad t > 0, \quad \textit{and} \quad \varphi(t) = 0 \Leftrightarrow t = 1.$$

Hence, d_φ satisfies

$$d_\varphi(x,y) \geq 0 \quad (x,y) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}, \quad \textit{and}$$
$$d_\varphi(x,y) = 0 \Leftrightarrow x = y$$

Lemma 2.4 *Let $\varphi \in \Phi_r$, fixed $y \in \mathbb{R}^n_{++}$. Then, $L_r(y,\gamma) := \{x \in \mathbb{R}^n_{++} : d_\varphi(x,y) \leq \gamma\}$ are bounded for all $\gamma \geq 0$, with $r \in \{1,2\}$.*

Proof: To $r = 1$ is direct by definition $\varphi \in \Phi_1$ and to $r = 2$ is direct by Proposition 2.3. □

Now, we present the definition of Bregman distance. Let S be an open and convex subset of $\mathbb{R}^n$ and $\overline{S}$ its closure. Consider a convex real function h defined on $\overline{S}$ and let $D_h : \overline{S} \times S \to \mathbb{R}$ be

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle. \tag{2.15}$$

Definition 2.5 *The function h is said to be a Bregman function (and D_h the Bregman distance induced by h) if the following conditions hold:*

(B1) *h is continuously differentiable S;*

(B2) *h is strictly convex and continuous in $\overline{S}$;*

(B3) *For all $\delta \in \mathbb{R}$ the partial level sets $\Gamma_1(y,\delta) = \{x \in \overline{S} : D_h(x,y) \leq \delta\}$, $\Gamma_2(x,\delta) = \{y \in \overline{S} : D_h(x,y) \leq \delta\}$ are bounded for all $y \in S$, all $x \in \overline{S}$ respectively;*

(B4) *If $\{y^k\} \subset S$ converges to y then $D_h(y, y^k)$ converges to 0.*

(B5) *If $\{x^k\} \subset \overline{S}$ and If $\{y^k\} \subset S$ are sequences such that $\{x^k\}$ is bounded, $\lim_{k\to\infty} y^k = y$ and $\lim_{k\to\infty} D_h(x^k, y^k) = 0$ then $\lim_{k\to\infty} x^k = y$.*

S is called the zone of h. It is easy to check that $D_h(x,y) \geq 0$ for all $x \in \overline{S}, y \in S$ and $D_h(x,y) = 0$ if and only if $x = y$.

The following definition presents a subclass of Bregman functions.

Definition 2.6 *Let h be a Bregman function with zone S.*

(i) *h is called zone coercive when for all $y \in \mathbb{R}^n$ there exists $x \in S$ such that $\nabla h(x) = y$, that is, ∇h is onto.*

(ii) *h is called separable when h can be written in the form*

$$h(x) = \sum_{i=1}^n h_i(x_i),$$

with h_i scalar Bregman functions. In this case, distance D_h associated to h is also called separable.

Examples of Bregman functions can be found in [8] , [11]. In particular, let $S = I\!R^n_{++}, h(x) = \sum_{i=1}^n x_i \log x_i$, extended with continuity to $\partial I\!R^n_+$ with the convention that $0 \log 0 = 0$, in this case $D_h(x,y) = H(x,y) = \sum_{i=1}^n (x_i \log \frac{x_i}{y_i} + y_i - x_i)$, where H is given in (2.12).

We next recall the definition of Fejér convergent sequence regarding the Euclidean distance.

Definition 2.7 *A sequence $\{x^k\}$ in $I\!R^n$ is Fejér convergent to a set $U \subset R^n$ with respect to the Euclidean distance if*

$$\| x^{k+1} - u \| \leq \| x^k - u \| \quad \textit{for all} \quad k \geq 0, \quad \textit{for all} \quad u \in U.$$

We have the following result

Proposition 2.4 *If $\{x^k\}$ is Fejér convergent to $U \neq \emptyset$ then $\{x^k\}$ is bounded. If a cluster point x of $\{x^k\}$ belongs to U then $x = lim_{k \to \infty} x^k$.*

Proof: See [16]. □

3 Algorithm

In this section, we propose proximal point algorithms to solve the problem (P). Denoted by Algorithm A1, A2 and A3, respectively. We show the well definedness of the generated sequence. All have the following common structure:

Initialization: $x^0 > 0$
Iterative step: Given x^k, calculate the next iterative x^{k+1} from

$$x^{k+1} \in argmin\{f_k(x)\}$$

where $f_k(x) := f(x) + \lambda_k d_\varphi(x,y)$ and $\{\lambda_k\}$ a sequence of positive real number satisfying $0 < \lambda_k \leq \bar{\lambda}$ for some $\bar{\lambda} > 0$.
According to the choice of φ and $d_\varphi(\cdot,\cdot)$ we have the following two algorithms:

Algorithm A1 Consider the above structure with $\varphi \in \Phi_1$ and $d_\varphi(\cdot,\cdot)$ defined in (2.11).

Algorithm A2 Consider the above structure with $\varphi \in \Phi_2$ and $d_\varphi(\cdot,\cdot)$ defined in (2.14).

The following algorithm we are interested in Bregman functions with the following requirement:
(R) A Bregman distance $D_h(\cdot,\cdot)$, defined in (2.15), induced by a Bregman function h, with zone $S = I\!R^n_{++}$ and such that h is separable and coercive zone.

Algorithm A3 Consider the above structure where $d_\varphi(\cdot,\cdot)$ is replaced by a Bregman distance $D_h(\cdot,\cdot)$.

We show that the algorithm **A1** and **A2** are well-defined under the following assumption.

(H1) $f : I\!R^n \to I\!R$ is bounded below, that is, there exists $\beta \in I\!R$ such that $\beta \leq f(x)$ for all $x \in I\!R^n_+$.

(H2) f is continuously differentiable and quasiconvex function.

We denote by U the following subset of $I\!R^n_+$, which is associated to the sequence $\{x^k\}$.

$$U := \{x \in I\!R^n_+ : f(x) < f(x^k), \quad k = 0, 1, ...\} \tag{3.16}$$

In the following result we establish the well-definedness of the **A1** and **A2** algorithms, and the existence of a limit point to sequence $\{x^k\}$ generated by **A1** or **A2** algorithms.

Proposition 3.1 *Assuming (H1) and (H2), we have:*

(i) *The sequences $\{x^k\}$ generated by* **A1** *and* **A2** *are well-defined;*

(ii) *$\{f(x^k)\}$ is a decreasing and convergent sequence, where $\{x^k\}$ is generated by* **A1** *or* **A2**;

(iii) *If $U \neq \emptyset$ then, the sequence $\{x^k\}$ generated by* **A1** *is bounded;*

(iv) *If $U \neq \emptyset$ then, the sequence $\{x^k\}$ generated by* **A2** *is Fejér convergent to U.*

Proof: Considering that the proof of (i) and (ii), for both the sequence generated by **A1** and by **A2**, follows the same structure, we prove only the case where $\{x^k\}$ is generated by **A2**.
(i) By induction, $x^0 > 0$ there exists by the initialization of the algorithm **A2**. Let $x^k > 0$, by **H1** it follows that

$$f_k(x) = f(x) + \lambda_k d_\varphi(x, x^k) \geq \beta + \lambda_k d_\varphi(x, x^k)$$

by Definition 2.4 and $(x^k)_i > 0$, we have

$$\lambda_k d_\varphi(x, x^k) \to \infty, \quad \text{as} \quad x \to \partial I\!R^n_{++} \tag{3.17}$$

Since f is bounded below and (3.17) holds, then

$$f_k(x) = f(x) + \lambda_k d_\varphi(x, x^k) \to \infty \quad \text{as} \quad x \to \partial I\!R^n_{++} \tag{3.18}$$

With f_k is continuous, bounded below and (3.18) holds, f_k reaches its minimum in a point $w > 0$. Therefore, there exists $x^{k+1} \in argmin\{f_k(x^k)\}$, with $x^{k+1} = w > 0$, which can not unique due to the non convexity of f.

(ii) By definition of x^k given by algorithm **A2**, we have

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

Setting $x = x^{k-1}$ in the last inequality, we have that

$$f(x^k) + \lambda_k d_\varphi(x^k, x^{k-1}) \leq f(x^{k-1})$$

which, by the nonnegativity of d_φ and λ_k, implies that

$$0 \leq \lambda_k d_\varphi(x^k, x^{k-1}) \leq f(x^{k-1}) - f(x^k).$$

This shows that $\{f(x^k)\}$ is a decreasing sequence. As f is bounded below $\{f(x^k)\}$ is convergent.

(iii) Given $x \in U$ then $f(x) < f(x^{k+1}) \quad \forall k \in I\!N$, where $\{x^k\}$ is generated by **A1**. By the assumption (H2) and Proposition (2.2), we obtain that

$$\langle \nabla f(x^{k+1}), x - x^{k+1} \rangle \leq 0. \tag{3.19}$$

Since $z \in argmin\{f_k(x)\}$ is such that $z > 0$. Hence, from the optimality conditions of this problem, it results $\nabla f_k(z) = 0$, and so $\nabla f(z) = -\lambda_k \bigtriangledown_1 d_\varphi(z, x^k) \quad \forall k \in I\!N$. Particularly,

$$\nabla f(x^{k+1}) = -\lambda_k \bigtriangledown_1 d_\varphi(x^{k+1}, x^k) \quad \forall k \in I\!N. \tag{3.20}$$

By (3.19), (3.20) and Lemma 2.2 (iii) with $w = x^{k+1}, z = x^k$ and $v = x$, we have that

$$\begin{aligned} 0 \geq \langle \nabla f(x^{k+1}, x - x^{k+1}) \rangle &= \langle -\lambda_k \bigtriangledown_1 d_\varphi(x^{k+1}, x^k), x - x^{k+1} \rangle \\ &= \lambda_k \langle \bigtriangledown_1 d_\varphi(x^{k+1}, x^k), x^{k+1} - x \rangle \\ &\geq \lambda_k \varphi''(1)[H(x, x^{k+1}) - H(x, x^k)] \end{aligned} \tag{3.21}$$

with $\lambda_k > 0$ and $\varphi''(1) > 0$, we obtain that

$$H(x, x^{k+1}) \leq H(x, x^k) \tag{3.22}$$

Now, from Lemma 2.2 (i) the level sets of $H(x, \cdot)$ are bounded, thus implying the boundedness of $\{x^k\}$.

(iv) In a way similar to the previous item, given $x \in U$ where $\{x^k\}$ is generated by **A2** then, we obtain

$$\langle \nabla f(x^{k+1}), x - x^{k+1} \rangle \leq 0. \tag{3.23}$$

By (3.20), (3.23) and Lemma 2.3, with $v = x, w = x^{k+1}$ and $z = x^k$, we have

$$\begin{aligned} 0 \geq \langle \nabla f(x^{k+1}, x - x^{k+1}) \rangle &= \langle -\lambda_k \bigtriangledown_1 d_\varphi(x^{k+1}, x^k), x - x^{k+1} \rangle \\ &= \lambda_k \langle \bigtriangledown_1 d_\varphi(x^{k+1}, x^k), x^{k+1} - x \rangle \\ &= \lambda_k \left[\frac{\theta}{2} \left(\| x^{k+1} - x \|^2 - \| x^k - x \|^2 \right) + \frac{\tau}{2} \| x^{k+1} - x^k \|^2 \right] \\ &\geq \lambda_k \frac{\theta}{2} \left(\| x^{k+1} - x \|^2 - \| x^k - x \|^2 \right) \end{aligned} \tag{3.24}$$

with $\lambda_k > 0$ and $\theta > 0$, we obtain that

$$\| x^{k+1} - x \|^2 \leq \| x^k - x \|^2$$

then $\{x^k\}$ is Fejér convergent to U. $\square$

4 Convergence Analysis

In this section we prove important convergence results. In the first, we establish convergence to a stationary point of the problem (P), when $U \neq \emptyset$.

Theorem 4.1 *If assumptions (H1), (H2) hold and $U \neq \emptyset$, then the sequence $\{x^k\}$ generated by the algorithm* **A1** *or* **A2** *converge to a stationary point of (P).*

Proof: First, prove that the sequence $\{x^k\}$, generated by **A1** or **A2**, is convergent. Let $\overline{x}$ be an accumulation point of $\{x^k\}$ and $\{x^{k_j}\}$ be a subsequence converging to $\overline{x}$. From the continuity of f, it we have that

$$\lim_{j\to\infty} f(x^{k_j}) = f(\overline{x})$$

by the definition of U, it implies that $\overline{x} \in U$.
To $\{x^k\}$ generated by **A1**, by (3.22) $\{H(\overline{x}, x^k)\}$ is decreasing, and the nonnegativity of H, we have $\{H(\overline{x}, x^k)\}$ is convergent. With $\{x^{k_j}\}$ converge to $\overline{x}$, by Lemma 2.2 (ii), $\lim_{j\to\infty} H(\overline{x}, x^{k_j}) = 0$, then it follows that

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

Consider, now, another accumulation point $\widehat{x}$ of $\{x^k\}$ and $\{x^{q_j}\}$ is a subsequence of $\{x^k\}$ such that, $\lim_{j\to\infty} x^{q_j} = \widehat{x}$. By the same argument

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

Using, again, by Lemma 2.2 (iii) with $y^k = x^{q_j}, y = \widehat{x}$ and $z^k = \overline{x}$, we have that $\overline{x} = \widehat{x}$. Soon $\{x^k\}$ is convergent.
To $\{x^k\}$ generated by **A2**, using Proposition 2.4 and Proposition 3.1 (iv), we have that $\{x^k\}$ is convergent.
We will show, now, that $\overline{x}$ is a stationary point of the problem (P), that is,

$$\overline{x} \geq 0, \quad \nabla f(\overline{x}) \geq 0 \quad \text{and} \quad \overline{x}_i(\nabla f(\overline{x}))_i = 0 \quad \forall i = 1, ..., n. \tag{4.25}$$

Henceforth the proof basically follows the same structure, for both the sequence generated by **A1** and **A2**, we will do only the case where $\{x^k\}$ is generated by **A2**.
The first condition in (4.25) is immediate, since $\lim_{k\to\infty} x^k = \overline{x}$ and $x^k > 0$. To prove the other two conditions in (4.25) we consider the sets:

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

Clearly, the two disjoint sets from a division of $\{1, ..., n\}$, and we analyze the cases when $i \in I(\overline{x})$ or $i \in J(\overline{x})$ for any $i \in \{1, ..., n\}$.

Case 1: If $i \in I(\overline{x})$, it supposes by contradiction that $(\nabla f(\overline{x}))_i < 0$. By the continuous differentiability of f, we have $\nabla f(x^{k+1})_i \to \nabla f(\overline{x})_i < 0$. Therefore $\nabla f(x^{k+1})_i < 0$ for k sufficiently large. By (3.20) and Definition 2.4, we have

$$(\nabla f(x^{k+1}))_i = -\lambda_k(\nabla_1 d_\varphi(x^{k+1}, x^k))_i = -\lambda_k \varphi' \left(\frac{x_i^{k+1}}{x_i^k} \right) \tag{4.26}$$

Since $\lambda_k > 0, \varphi'\left(\frac{x_i^{k+1}}{x_i^k}\right) > 0$ for k sufficiently large. By Proposition 2.3, we have $x_i^{k+1} > x_i^k > 0$ for k sufficiently large, which contradicts the fact that $\{x_i^k\}$ converges to $\overline{x}_i$ and $\overline{x}_i = 0$. Therefore $(\nabla f(\overline{x}))_i \geq 0$ for each $i \in I(\overline{x})$.

Case 2: If $i \in J(\overline{x})$, then $\lim_{k\to\infty} \frac{x_i^{k+1}}{x_i} = 1$. Using, again, Proposition 2.3, we obtain,

$$\varphi'\left(\frac{x_i^{k+1}}{x_i}\right) \to \varphi'(1) = 0.$$

This, together with (4.26) and the boundedness of $\{\lambda_k\}$, we have

$$(\nabla f(x^{k+1}))_i \to (\nabla f(\overline{x}))_i = 0 \quad \forall i \in J(\overline{x}).$$

From cases (1) and (2), we conclude that

$$\nabla f(\overline{x}) \geq 0 \quad \text{and} \quad \overline{x}_i(\nabla f(\overline{x}))_i = 0 \quad \forall i = 1, ..., n.$$

$\square$

With the hypotheses of the previous theorem, the next result establishes a sufficient condition for $\overline{x}$ is a solution of the problem (P). Condition which is not included in [9], [10] and [25]. Henceforward $\overline{x} = \lim_{k\to\infty} x^k$, where $\{x^k\}$ is generated by **A1** or **A2**.

Theorem 4.2 *Suppose (H1), (H2) hold and $U \neq \emptyset$, we have*

(i) *If $\nabla f(\overline{x}) \neq 0$ then $\overline{x}$ is solution of the problem (P);*

(ii) *If $\lambda_k \to 0$, then $\overline{x}$ is solution of the problem (P).*

Proof: **(i)** Suppose by contradiction that $\overline{x}$ is not one of minimizing f, then there exists $\widehat{x} \geq 0$ such that $f(\widehat{x}) < f(\overline{x})$. With $\nabla f(\overline{x}) \neq 0$ and $f(\widehat{x}) < f(\overline{x})$ by Lemma 2.1, we have

$$\langle \nabla f(\overline{x}), \widehat{x} - \overline{x} \rangle < 0$$

Consequently,

$$0 > \langle \nabla f(\overline{x}), \widehat{x} - \overline{x} \rangle = \langle \nabla f(\overline{x}), \widehat{x} \rangle - \langle \nabla f(\overline{x}), \overline{x} \rangle = \langle \nabla f(\overline{x}), \widehat{x} \rangle.$$

Which is a contradiction, because $\nabla f(\overline{x}) \geq 0$ and $\widehat{x} \geq 0$. Therefore $\overline{x}$ is a solution of the problem (P).

(ii) Since x^{k+1} is the minimizer of $f(x) + \lambda_k d_\varphi(x, x^k)$, we have

$$f(x^{k+1}) + \lambda_k d_\varphi(x^{k+1}, x^k) \leq f(x) + \lambda_k d_\varphi(x, x^k) \quad \forall x \in I\!R^n_{++}$$

Taking the limit $k \to \infty$ in the last inequality, and using the continuity of f, $\lim_{k\to\infty} \lambda_k = 0$ and Lemma 2.4, we have that

$$f(\overline{x}) \leq f(x) \quad \forall x \in I\!R^n_{++} \tag{4.27}$$

Consider $\{y^k\}$ sequence in $I\!R^n_{++}$ and $y \in I\!R^n$, with $\lim_{k\to\infty} y^k = y$.
It follows directly from (4.27) that,

$$f(\overline{x}) \leq f(y^k) \quad \forall k \in I\!N \tag{4.28}$$

taking the limit in (4.28) with $k \to \infty$ and using, again, the continuity of f, we have

$$f(\overline{x}) \leq f(y) \quad \forall y \in I\!R^n_+$$

Therefore, $\overline{x}$ is a solution of the problem (P). $\square$

Considering now the sequence $\{x^k\}$ generated by algorithm **A3**. Let S^* the solution set of problem (P). The following proposition shows the results of sequence $\{x^k\}$ under (H2) and the following additional assumption.

(H3) S^*, the solution set of (P), is nonempty.

Proposition 4.1 *Let $\{x^k\}$ the sequence generated by algorithm* **A3***. Suppose (H2) and (H3) hold and, we have:*

(i) *The sequence $\{x^k\}$ is well-defined;*

(ii) *The sequence $\{x^k\}$ converges to a stationary point $\overline{x}$ of (P);*

(iii) *If $\lambda_k \to 0$, then $\overline{x}$ is the solution of the problem (P);*

(iv) *If $\nabla f(\overline{x}) \neq 0$, then $\overline{x}$ is the solution of the problem (P).*

Proof: For part (i) - (iii) see [25]. For part (iv) the proof as exactly as it is in the proof of Theorem 4.2 (i). $\square$

Finally, consider now the case $U = \emptyset$.

Proposition 4.2 *Assume(H1), (H2) hold. If $U = \emptyset$ then $\lim_{k\to\infty} f(x^k) = \inf_{x\geq 0} f(x)$ and $\{x^k\}$ is unbounded, where $\{x^k\}$ generated by* **A1** *or* **A2**.

Proof: Considering that the proof follows the same structure, so that $\{x^k\}$ can be generated by **A1** or **A2**, prove only the case where $\{x^k\}$ is generated by **A2**. Therefore, by Proposition 3.1 $\{f(x^k)\}$ is decreasing and a convergent sequence. Let $\alpha := \lim_{k\to\infty} f(x^k)$, and suppose by contradiction that $\alpha \neq \inf_{x\geq 0} f(x)$. Then there exists $x^* \in I\!R^n_+$ such that $f(x^*) < \alpha$. On the other hand, $f(x^*) < \lim_{k\to\infty} f(x^k)$, implies that $x^* \in U$, which is a contradiction with $U = \emptyset$, consequently,

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

Suppose, again, by contradiction that $\{x^k\}$ is bounded. Then there exists the subsequence $\{x^{k_j}\}$ of $\{x^k\}$ such that,

$$\lim_{j\to\infty} f(x^{k_j}) = w \geq 0$$

From the continuity of f,

$$\lim_{j\to\infty} f(x^{k_j}) = f(w)$$

on the other hand,

$$\lim_{j\to\infty} f(x^{k_j}) = \inf_{x\geq 0} f(x)$$

Soon $w = \inf_{x\geq 0} f(x)$, implies that $w \in U$, which is a contradiction. Therefore, $\{x^k\}$ is unlimited. $\square$

5 Conclusions and future work

We analyze the proximal point algorithms defined by (1.4) associated to distance-like functions (1.3), (1.5) and (1.6) for minimizing continuously differentiable quasiconvex functions in the nonnegative orthant. We have shown that, the sequences generated by algorithms $A1$, $A2$ and $A3$ converge to a stationary point $\overline{x}$ of problem (P). It shows that, if $\nabla f(\overline{x}) \neq 0$ or if the parameters λ_k satisfy the condition $\lambda_k \to 0$ then $\overline{x}$ is a solution of (P). As a future research, we are interested in extending this work, considering the objective function only lower semicontinuous quasiconvex.

References

[1] Auslender, A. and Teboulle, M. Interior gradient and proximal methods for convex and conic optimization. SIAM Journal Optimization, 16 (2006); 693-725.

[2] Auslender, A., Teboulle, M. and Ben-Tiba, S. Interior proximal and multiplier methods based on second order homogeneous kernels. Mathematics of Operations Research, 24 (1999), 645-668.

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

[4] Bajona-Xandrei C., Martinez-Legaz, J. E., Lower subdifferentiability in minimax fractional programming, Optimization, 45 (1999), 1-22.

[5] Burachik, R.S., Lopes, J. O. and Silva, G.J.P. An inexact interior proximal method for the variational inequality problem. Computational Applied Mathematics, 28 (2009), 1-22.

[6] Byrne,C. and Censor, Y. Proximity functions minimization using multiple Bregman projections, with applications to split feasibility an Kullback-Leibler distance minimization. Annals of Operations Research, 105 (2001), 77-98.

[7] Censor, Y., Iusem, A.N. and Zenios, S. An interior point method Bregman functions for the variational inequality problem with paramonotone operators. Mathematical Programming. 81 (1998), 373-400.

[8] Censor, Y. and Zenios, S.A. The proximal minimization algorithm with D-functions. Journal of Optimization Theory and Applications, 73 (1992), 451-464.

[9] Chen, J-S. and Pan, S. A proximal-like algorithm for a class of nonconvex programming. Pacific Journal of Optimization, 4(2008), 319-333.

[10] Cunha, F.G.M., Cruz Neto, J.X. da, and Oliveira, P.R. A proximal point algorithm with φ- divergence to quasiconvex programming, Optimization: A journal of Mathematical Programming and Operations Research, (2010), 1029-1045.

[11] Crouzeix, J.-P., Martinez-Legaz J.-E. and Volle, M. Generalized convexity, generalized monotonicity. Non Convex optimization and its applications. Kluwer Academic Publishers; 1998.

[12] Eckstein, J. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Mathematics of Operations Research 1993; 18, 1, 202-226.

[13] Gromicho, J. Quasiconvex optimization and location theory. Kluwer Academic Publishers: Dordrecht, The Netherlands; 1998.

[14] Güler, O. On the convergence of the proximal point proximal algorithm for convex minimization. SIAM Journal Control and Optimization, 29 (1991), 403-419.

[15] Güler, O. New proximal point proximal algorithms for convex minimization. SIAM Journal on Optimization 2 (1992), 649-664.

[16] Iusem, A. Métodos de ponto proximal em otimização, 20º Colóquio Brasileiro de Matemática, IMPA - Rio de Janeiro, 1995.

[17] Iusem, A.N., Svaiter, B. and Teboulle, M. Entropy-like proximal methods in convex programming. Mathematics of Operation Research, 17 (1992), 670-690.

[18] Mangasarian, O.L. Nonlinear Programming (Classics in Applied Mathematics), Society for Industrial Mathematics, Philadelphia, 1987.

[19] Martinet, B. Regularization d'inequations variationelles par approximation successives. Révue Française d'Informatique et Recherche Opérationelle 4 (1970), 154-159.

[20] Mas-Colell, A., Whinston, M.D. and Green, J. R., Microeconomic Theory, Oxford University Press, New York, USA, 1995.

[21] Linh, N.T.H. and Penot, J-P. Optimality conditions for quasiconvex programs. SIAM Journal on Optimization. 17 (2006), 500-510.

[22] Penot, J-P. Characterization of solution sets of quasiconvex progrms, Journal of Optimization Theory and Applications, 3 (2003), 627-636.

[23] Rockafellar, R.T. Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

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

[25] Souza, S.S., Oliveira. P.R., Cruz Neto, J. X. da. and Soubeyaran. A. A proximal method with separable distance for quasiconvex minimization over the nonnegative orthant. 2(2010), 365-376.

[26] Teboulle, M. Entropic proximal mapping with applications to nonlinear programming, Mathematics of Operation Research, 17 (1992), 670-690.

[27] Teboulle, M. Convergence of proximal-like algorithms. SIAM Journal on Optimization, 7 (1997), 1069-1083.