Int. J. Math. And Appl., 7(2)(2019), 115–124
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Inexact Version of the Entropic Proximal Point Algorithm

S. Kabbadj[1,*]

1 Department of Mathematic, Faculty of Sciences of Meknès, B.P. 11201, Morocco.

Abstract: In this paper, we study an inexact version of the entropic proximal point algorithm defined by

$$x^k \in \varepsilon_k - Argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\}.$$

This study recovers the most of the algotithms of the proximal point.

Keywords: Convex optimization, Bregman's function, Entropic approximation.

1. Introduction

Let's consider the problem of a convex optimization

$$(\text{P}) : \min\ \{f(x), x \in R^n\},$$

where $f \in \Gamma_0(R^n)$ set of lower semicontinuous, proper and convex functions on R^n. By using the approximation of Moreau-Yosida, Auslender [1], Martine [8], Rockafellar [9], Behraoui [2], and Lemaire [6, 7] have given the methods of resolution of (P) named proximal methods defined by:

$$x^k \in \varepsilon_k - Argmin\left\{f(.) + \frac{1}{2\lambda_k}\|. - x^{k-1}\|^2\right\}.$$

By using the entropic approximation [5], Eckstein [3], thus Chen and Teboulle [10] have given a classe of methods named entropics proximals defined by:

$$x^k := argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\}$$

where $D_h(.,.)$ is the entropic distance defined by:

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

In this labor, we propose an inexact version defined by:

$$x^k \in \varepsilon_k - Argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\},$$

* E-mail: kabbajsaid63@yahoo.com

where $h \in C(S)$ [5], we show that x^k converges to x^* solution of the problem (P). Our result of convergence obtained in section 4, recovers what has been given by Auslender (it's enough to take $h(.) = \frac{1}{2}\|.\|^2$), thus the result of convergence obtained by Eckstein (it's enough to take $\varepsilon_k = 0$).

Our notation is fairly standard; $\langle ., . \rangle$ is the scalar product on R^n; and the associated norm $\|.\|$. The closure of the set C (relative interior) is denoted by $\overline{C}$ (riC, respectivly), Adh $\{x^k\}$ is the set of adherence values of a sequence $\{x^k\}$. For any convex function f, we denote by:

(1). $domf = \{x \in R^n; f(x) < +\infty\}$ its effective domain,

(2). $\partial_\epsilon f(.) = \{v, f(y) \geq f(.) + \langle v, y - . \rangle - \epsilon, \ \forall \ y\}$ its ε-sub differential,

(3). $Arg \min f = \{x \in R^n; f(x) = \inf f\}$ its Argmin f,

(4). $\varepsilon - Arg \min f = \{x \in R^n; f(x) \leq \inf f + \varepsilon\}$ its $\varepsilon - Argminf$.

2. Preliminaries

Let S be an convex open subset of R^n and $h : \overline{S} \longrightarrow R$. We define $D_h(., .)$ by: $\forall \ x \in \overline{S}$, $\forall \ y \in S$:

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

Let us consider the following hypotheses:

H_1: h is continuously differentiable on S.

H_2: h is continuous and strictly convex on $\overline{S}$.

H_3: $\forall \ r \geq 0$, $\forall \ x \in \overline{S}$, $\forall \ y \in S$, the sets $L_1(x, r)$ and $L_2(y, r)$ are bounded where:

$$L_1(x, r) = \{y \in S / D_h(x, y) \leq r\}$$
$$L_2(y, r) = \{x \in \overline{S} / D_h(x, y) \leq r\}.$$

H_4: If $\{y^k\}_k \subset S$ is such as $y^k \longrightarrow y^* \in \overline{S}$, so $D_h(y^*, y^k) \longrightarrow 0$.

H_5: If $\{x^k\}_k \subset \overline{S}$ is such as $\{y^k\}_k \subset S$ are such as:

$$y^k \longrightarrow y^* \in \overline{S} \text{ , } \{x^k\}_k \text{ is bounded, and } D_h(x^k, y^k) \longrightarrow 0, \text{ then } x^k \longrightarrow y^*.$$

H_6: If $\{x^k\}_k$ and $\{y^k\}_k$ are two sequences of S such as:

$$D_h(x^k, y^k) \longrightarrow 0 \text{ and } x^k \longrightarrow x^* \in S, \text{ then } y^k \longrightarrow x^*.$$

Definition 2.1.

(1). $h : \overline{S} \longrightarrow R$ is a Bregman function on S or "D-function" if h verify H_1, H_2, H_3, H_4 and H_5.

(2). $D_h(., .)$, is called entropic distance if h is a Bregman function.

We put:

$$A(S) = \{h : \overline{S} \longrightarrow R \;\; verifying \;\; H_1 \; and \; H_2\}$$

$$B(S) = \{h : \overline{S} \longrightarrow R \;\; verifying \;\; H_1, H_2, H_3, H_4 \; and \; H_5\}$$

$$C(S) = \{h : \overline{S} \longrightarrow R \;\; verifying \;\; H_1, H_2, H_3, H_4 \; and \; H_6\}.$$

Lemma 2.2. $\forall\, h \in A(S)$, $\forall\, a \in \overline{S}$, $\forall\, b, c \in S$:

$$D_h(a, b) + D_h(b, c) - D_h(a, c) = \langle a - b, \nabla h(c) - \nabla h(b)\rangle.$$

Proposition 2.3. *Let $h \in A(S)$. Let $\{x^k\}$ and $\{y^k\}$ two sequences of S such that:*

(1). $x^k \to x^* \in S$ and $\{y^k\}$ bounded such as $Adh\{y^k\} \subset S$

(2). $D_h(x^k, y^k) \to 0$.

Then $y^k \to x^*$.

Proof. $\{y^k\}$ is bounded, then we can extract a sub-sequence $\{y^{k_i}\}$ of $\{y^k\}$ such that $y^{k_i} \to y \in S$. We have, $D_h(x^{k_i}, y^{k_i}) := h(x^{k_i}) - h(y^{k_i}) - \langle x^{k_i} - y^{k_i}, \nabla h(y^{k_i})\rangle$. So, when $k_i \to +\infty$, we obtain: $0 = h(x^*) - h(y) - \langle x^* - y, \nabla h(y)\rangle \Rightarrow D_h(x^*, y) = 0 \Rightarrow x^* = y \Rightarrow Adh\{y^k\} = \{x^*\} \Rightarrow y^k \to x^*$. □

Proposition 2.4. *We assume:*

(1). $h \in A(S)$.

(2). $\{x^k\} \subset S$ is bounded.

(3). $Adh\{x^k\} \subset S$.

Then $\{D_h(u, x^k)\}_k$ is bounded for all $u \in \overline{S}$.

Proof. If the sequence $\{D_h(u, x^k)\}_k$ is not bounded for $u \in \overline{S}$, then exists a sub-sequence $\{D_h(u, x^{k_i})\}_{k_i}$ such that: $D_h(u, x^{k_i}) \to +\infty$. $\{x^{k_i}\}$ is bounded, then it exists a sub-sequence $\{x^{k_j}\}$ of $\{x^{k_i}\}$ such as $x^{k_j} \to x^*$. From (3), $x^* \in S$, so $\nabla h(x^{k_j}) \to \nabla h(x^*)$.

$$D_h(u, x^{k_j}) := h(u) - h(x^{k_j}) - \langle u - x^{k_j}, \nabla h(^{k_j})\rangle.$$

when, $k_j \to +\infty$, we obtain: $+\infty = D_h(u, x^*)$, which is contradictory with $h : \overline{S} \to R$. □

Proposition 2.5. *Let $h \in A(S)$ and $f \in \Gamma_0(R^n)$ such as :*

(1). ri (dom f) $\cap S \neq \phi$,

(2). $Im\nabla h = R^n$.

Then $\forall\, \varepsilon > 0$, $\forall\, \lambda > 0$, $\forall\, x \in S$, $\exists\, y \in S : y \in \varepsilon - Argmin\{f(.) + \lambda^{-1} D_h(., x)\}$.

Proof. From the Corollary 5.8. [5], $\forall\, \varepsilon > 0$, $\forall\, \lambda > 0$, $\forall\, x \in S$, $y := h^f_\lambda(x) = argmin\{f(.) + \lambda^{-1} D_h(., x)\} \in S$. So, $h^f_\lambda(x) \in \varepsilon - Argmin\{f(.) + \lambda^{-1} D_h(., x)\}$. □

3. Proximal Point Algorithm

Let $f \in \Gamma_0(R^n)$, Auslender [1] condered the proximal following method:

$$P(1) : \begin{cases} x^0 \in R^n, \varepsilon_k > 0 \text{ and } \lambda_k > 0. \\ x^k \in \varepsilon_k - Argmin\{f(.) + \frac{1}{2\lambda_k}\|. - x^{k-1}\|^2\}. \end{cases}$$

The sequence $\{x^k\}$ generated by $P(1)$ verify the following properties:

Proposition 3.1. *It exists a sequence* $\{\overline{x}^k\}$ *such as :* $\frac{x^{k-1}-\overline{x}^k}{\lambda_k} \in \partial_{\varepsilon_k} f(x^k)$ *and* $\|x^k - \overline{x}^k\| \leq \sqrt{2\lambda_k \varepsilon_k}$.

Theorem 3.2. *We assume:*

(1). $\lambda_k \geq \underline{\lambda} > 0$,

(2). f *is inf-compact, i.e.,* $\forall\, r \in R$, $\{x \in R^n, f(x) \leq r\}$ *is compact, if it is not empty.*

(3). $\sum \varepsilon_k < +\infty$ *or* ($f(0)$ *is finite and* $\varepsilon_k \to 0$).

Then $\{x^k\}$ *defined by* $P(1)$ *is bounded and* $f(x^k) \to inf f$.

Teboule and chen [10] have considered the algorithm defined by:

$$P(2)(h) : \begin{cases} x^0 \in S, \quad \lambda_k > 0. \\ x^k = argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\}. \end{cases}$$

Theorem 3.3. *If* $f \in \Gamma_0(R^n)$ *and* $h \in B(S)$ *such that*

(1). $Im\nabla h = R^n$,

(2). $\sum \lambda_k = +\infty$,

(3). $ri(dom\ f) \subset S$.

Then:

(a). $f(x^k) \to inf f$.

(b). If $Argmin f \neq \emptyset$, *then* $x^k \to x^* \in Argmin f$.

4. Inexact Entropic Proximal Point Algorithm

In this paragraph, we assume:

$$(A) : \ h \in C(S) : Im\nabla h = R^n$$

$$(B) : \ f \in \Gamma_0(R^n) : \overline{dom f} \subset S$$

From the Proposition 2.5, we can thus construct the sequence $\{x^k\}$ defined by:

$$P(3)(h) : \begin{cases} x^0 \in S \\ x^k \in \varepsilon_k - Argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\} \\ f(x^k) \text{ is decreasing} \\ \varepsilon_k \to 0 \text{ and } \lambda_k \geq \underline{\lambda} > 0. \end{cases}$$

To choose $\{x^k\}$ such that $\{f(x^k)\}$ is decreasing is possible. Indeed:

Let $\overline{x}^k \in \varepsilon_k - Argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\}$, for all $k \geq 1$. We take

$$x^k = \begin{cases} \overline{x}^k \ if \ f(\overline{x}^k) \leq f(x^{k-1}) \\ x^{k-1} \quad if \ not \end{cases}$$

If $f(\overline{x}^k) > f(x^{k-1})$ then $x^{k-1} \in \varepsilon_k - Argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\}$, because:

$$\lambda_k^{-1} D_h(x^{k-1}, x^{k-1}) + f(x^{k-1}) < f(\overline{x}^k) \leq f(\overline{x}^k) + \lambda_k^{-1} D_h(\overline{x}^k, x^{k-1}) \leq f(x) + \lambda_k^{-1} D_h(x, x^{k-1}) + \varepsilon_k.$$

Let's consider now the function $h_{u,\lambda}$ defined by: $h_{u,\lambda} : \overline{S} \to R$, $\forall \lambda > 0$, $\forall u \in S$.

$$h_{u,\lambda}(x) = \lambda^{-1} D_h(x, u), \forall x \in \overline{S}.$$

Proposition 4.1. *$\forall \varepsilon > 0$, $\forall \lambda > 0$, $\forall x^* \in \overline{S}$, $\forall u \in S$:*

$$\partial_\varepsilon h_{u,\lambda}(x^*) = \left\{ z/z = \frac{\nabla h(\overline{x}) - \nabla h(u)}{\lambda} \ with \ \ \overline{x} \in S \ and \ D_h(x^*, \overline{x}) \leq \lambda\varepsilon \right\}.$$

Proof.

$$\begin{aligned} z \in \partial_\varepsilon h_{u,\lambda}(x^*) &\Leftrightarrow \forall x \in \overline{S}, h_{u,\lambda}(x) - h_{u,\lambda}(x^*) \geq \langle z, x - x^* \rangle - \varepsilon \\ &\Leftrightarrow \forall x \in \overline{S}, \lambda^{-1}[h(x) - h(u) - \langle x - u, \nabla h(u) \rangle - h(x^*) + h(u) + \langle x^* - u, \nabla h(u) \rangle] \geq \langle z, x - x^* \rangle - \varepsilon \\ &\Leftrightarrow \forall x \in \overline{S}, h(x^*) - h(x) - \langle x^* - x, \nabla h(u) \rangle \leq \langle \lambda z, x^* - x \rangle + \lambda\varepsilon \\ &\Leftrightarrow \forall x \in \overline{S}, h(x^*) - h(x) - \langle x^* - x, \nabla h(u) + \lambda z \rangle \leq \lambda\varepsilon. \qquad (1) \end{aligned}$$

According to (A), it exists $\overline{x} \in S$ such that $\nabla h(u) + \lambda z = \nabla h(\overline{x})$, which means:

$$\exists\, \overline{x} \in S, \ z = \frac{\nabla h(\overline{x}) - \nabla h(u)}{\lambda}.$$

Replacing in (1), $\times$ by $\overline{x}$, we get

$$h(x^*) - h(\overline{x}) - \langle x^* - \overline{x}, \nabla h(\overline{x}) \rangle \leq \lambda\varepsilon \Leftrightarrow D_h(x^*, \overline{x}) \leq \lambda\varepsilon.$$

Finally

$$\partial_\varepsilon h_{u,\lambda}(x^*) \subset \left\{ z/z = \frac{\nabla h(\overline{x}) - \nabla h(u)}{\lambda} \ with \ \ \overline{x} \in S \ and \ D_h(x^*, \overline{x}) \leq \lambda\varepsilon \right\}.$$

Conversely, let z such that $z = \frac{\nabla h(\overline{x}) - \nabla h(u)}{\lambda}$ and $D_h(x^*, \overline{x}) \leq \lambda\varepsilon$.

$$\begin{aligned} D_h(x^*, \overline{x}) \leq \lambda\varepsilon &\Rightarrow h(x^*) - h(\overline{x}) - \langle x^* - \overline{x}, \nabla h(\overline{x}) \rangle \leq \lambda\varepsilon \leq \lambda\varepsilon + D_h(x, \overline{x}) \\ &\Rightarrow h(x^*) - h(\overline{x}) - \langle x^* - \overline{x}, \nabla h(\overline{x}) \rangle - h(x) + h(\overline{x}) + \langle x - \overline{x}, \nabla h(\overline{x}) \rangle \leq \lambda\varepsilon \\ &\Rightarrow h(x^*) - h(x) - \langle x^* - x, \nabla h(\overline{x}) \rangle \leq \lambda\varepsilon. \end{aligned}$$

Replacing $\nabla h(\overline{x})$ by $\nabla h(u) + \lambda z$, we get (1). According to what precedes (1) $\Leftrightarrow z \in \partial_\varepsilon h_{u,\lambda}(x^*)$, which establishes the desired equality. □

Lemma 4.2 ([4]). *Let f_1, f_2 two functions of $\Gamma_0(R^n)$ if it exits $\overline{x} \in dom f_1$ in which f_2 is finite and continuous, then for $\varepsilon > 0$, for all $y \in dom f_1 \cap dom f_2$,*

$$\partial_\varepsilon(f_1 + f_2)(y) = \bigcup_{\varepsilon_1+\varepsilon_2=\varepsilon,\ \varepsilon_1\geq 0,\ \varepsilon_2\geq 0} \partial_{\varepsilon_1} f_1(y) + \partial_{\varepsilon_2} f_2(y).$$

Definition 4.3. *The sequence $\{(\lambda_k; a_k; b_k; c_k; d_k)\}_k \in R^{+*}XS^4$ verifies the K-property if only if the following properties are verified:*

$$K_1 : \exists\ \underline{\lambda} > 0, \forall\ k\ ,\ \lambda_k \geq \underline{\lambda}.$$
$$K_2 : \{a_k\} \textit{ is bounded and } Adh\{a_k\} \subset S.$$
$$K_3 : D_h(a_k, b_k) \to 0.$$
$$K_4 : D_h(a_k, c_k) \to 0.$$
$$K_5 : d_k = \tfrac{\nabla h(b_k) - \nabla h(c_k)}{\lambda_k}.$$

Lemma 4.4. *If the sequence $\{(\lambda_k; a_k; b_k; c_k; d_k)\}_k$ verifies the K-property then $d_k \to 0$.*

Proof. If the sequence $\{d_k\}$ does not tend to zero, then it exists $M > 0$ and the sub-sequence $\{d_{k_i}\}$ of $\{d_k\}$ such that:

$$\forall\ k_i\ ,\ \ \|d_{k_i}\|\ >\ M. \tag{2}$$

The sequence $\{a_{k_i}\}$ is bounded and $Adh\{a_{k_i}\} \subset S$, it exists then the sub-sequence $\{a_{k_j}\}$ of $\{a_{k_i}\}$ and $u^* \in S$ such as: $a_{k_j} \to u^*$. $D_h(a_{k_j}, b_{k_j}) \to 0$ and $D_h(a_{k_j}, c_{k_j}) \to 0$ allow to write, from H_6, $b_{k_j} \to u^*$ and $c_{k_j} \to u^*$. On the other hand,

$$0 \leq \|d_{k_j}\| = \|\frac{\nabla h(b_{k_j}) - \nabla h(c_{k_j})}{\lambda_{k_j}}\| \leq \underline{\lambda}^{-1}\|\nabla h(b_{k_j}) - \nabla h(c_{k_j})\|.$$

∇h is continuous on S, then $\nabla h(b_{k_j}) - \nabla h(c_{k_j}) \to 0$. It follows that $\|d_{k_j}\| \to 0$. $\{d_{k_j}\}$ being a sub-sequence of $\{d_{k_i}\}$, from (2) we have: $0 \geq M > 0$, so $d_k \to 0$. □

Theorem 4.5. *If the sequence $\{x^k\}$ generated by $P(3)(h)$ is bounded, then:*

(1). $f(x^k) \to inf f$.

(2). $Adh\{x^k\} \subset Argmin f$.

Proof. $x^k \in \varepsilon_k - Argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\} \Leftrightarrow 0 \in \partial_{\varepsilon_k}[f(.) + \lambda_k^{-1} D_h(., x^{k-1})](x^k)$. From the Lemma 4.2, it exists $\varepsilon_{k_1}, \varepsilon_{k_2} \geq 0$ such as : $\varepsilon_{k_1} + \varepsilon_{k_2} = \varepsilon_k$ and $0 \in \partial_{\varepsilon_{k_1}} f(x^k) + \partial_{\varepsilon_{k_2}}(\lambda_k^{-1} D_h(., x^{k-1}))(x^k)$. Since $\partial_\varepsilon f$ increases with ε, we have: $0 \in \partial_{\varepsilon_k} f(x^k) + \partial_{\varepsilon_k}(\lambda_k^{-1} D_h(., x^{k-1}))(x^k)$. Therefore, it exists $z_k \in \partial_{\varepsilon_k} f(x^k)$ such as $-z_k \in \partial_{\varepsilon_k}(\lambda_k^{-1} D_h(., x^{k-1}))(x^k)$. From the Proposition 4.1, it exists $\overline{x}^k \in S$ such as

$$-z_k = \frac{\nabla h(\overline{x}^k) - \nabla h(x^{k-1})}{\lambda_k} \textit{ and } D_h(x^k, \overline{x}^k) \leq \lambda_k \varepsilon_k.$$

Finally, it exists $\{\overline{x}^k\}$ such that:

$$\begin{cases} z_k = \frac{\nabla h(x^{k-1}) - \nabla h(\overline{x}^k)}{\lambda_k} \in \partial_{\varepsilon_k} f(x^k) \\ D_h(x^k, \overline{x}^k) \leq \lambda_k \varepsilon_k. \end{cases}$$

$f(x^k)$ is decreasing, so, $f(x^k) \to l \in [-\infty, +\infty[$. If $l = -\infty$, then $\lim\ f(x^k) = \inf f = -\infty$. In what follows, we assume $f(x^k) \to l > -\infty$. The sequence is such that

$$\forall\ u \in \overline{S}, \forall\ k \geq 1, f(x^k) + \lambda_k^{-1} D_h(x^k, x^{k-1}) \leq f(u) + \lambda_k^{-1} D_h(u, x^{k-1}) + \varepsilon_k.$$

Replacing u with x^{k-1} we get:

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

When $k \to +\infty$,

$$\lambda_k^{-1} D_h(x^k, x^{k-1}) \to 0. \tag{3}$$

At this level, we distinguish two cases:

Case 1: $0 < \underline{\lambda} \leq \lambda_k \leq \overline{\lambda} < +\infty$.

$$(3) \Rightarrow D_h(x^k, x^{k-1}) \to 0,$$

on the other hand, $D_h(x^k, \overline{x}^k) \leq \lambda_k \varepsilon_k \leq \overline{\lambda}\varepsilon_k$. So the sequence $\{x^k\}$ verifies then:

$$\begin{cases} z_k = \frac{\nabla h(x^{k-1}) - \nabla h(\overline{x}^k)}{\lambda_k} \in \partial_{\varepsilon_k} f(x^k) \\ D_h(x^k, \overline{x}^k) \to 0 \\ D_h(x^k, x^{k-1}) \to 0. \end{cases}$$

Since Adh $\{x^k\} \subset \overline{domf} \subset S$, the sequence $\{(\lambda_k; x^k; \overline{x}^k; x^{k-1}; z_k)\}_k$ verifies then the K-property. From the Lemma 4.4, $z_k \to 0$. On the other hand, $f(x) \geq f(x^k) + \langle z_k, x - x^k \rangle - \varepsilon_k$, which leads to $f(x) \geq \lim f(x^k)$, i.e., $\lim f(x^k) = \inf f$.

Case 2: $\{\lambda_k\}$ is not bounded. If $\{\lambda_k\}$ is not bounded, then it exists a sub-sequence $\{\lambda_{k_i}\}$ of $\{\lambda_k\}$ such that $\lambda_{k_i} \to +\infty$. We have

$$f(x^{k_i}) + \lambda_{k_i}^{-1} D_h(x^{k_i}, x^{k_i - 1}) \leq f(u) + \lambda_{k_i}^{-1} D_h(u, x^{k_i - 1}) + \varepsilon_{k_i}.$$

$\{x^k\}$ is bounded, so from the Proposition 2.4, $\{D_h(u, x^{k_i - 1})\}$ is bounded, which leads to:

$$l \leq f(u), \forall\ u \in \overline{S}.$$

Then:

$$\inf f = l = \lim f(x^k).$$

(b) Let $x^* \in Adh\{x^k\}$, it exists then the sub-sequence $\{x^{k_i}\}$ of $\{x^k\}$ such that $x^{k_i} \to x^*$. In any case, we have shown that: $\lim f(x^{k_i}) \leq \inf f$.

$$x^{k_i} \to x^* \Rightarrow f(x^*) \leq \lim f(x^{k_i}) \Rightarrow \inf f \leq f(x^*) \leq \lim f(x^{k_i}) \leq \inf f.$$

Finally we have:

$$f(x^*) = \lim f(x^k) = \inf f.$$

By additionating the hypothesis on h, we can establish the convergence of $\{x^k\}$ towards $x^* \in Argminf$. We put

$$y^k := prox_{\lambda_k f}^h(x^{k-1}) = argmin\{f(.) + \lambda_k^{-1} D_h(., x^{k-1})\}.$$

□

Proposition 4.6. $D_h(x^k, y^k) \leq \lambda_k \varepsilon_k$.

Proof.

$$y^k = prox^h_{\lambda_k f}(x^{k-1}) \Rightarrow \frac{\nabla h(x^{k-1}) - \nabla h(y^k)}{\lambda_k} \in \partial f(y^k) \Rightarrow \forall\, u, \lambda_k(f(u) - f(y^k)) \geq \langle u - y^k, \nabla h(x^{k-1}) - \nabla h(y^k)\rangle.$$

Thanks to the Lemma 2.2,

$$\langle u - y^k, \nabla h(x^{k-1}) - \nabla h(y^k)\rangle = D_h(u, y^k) + D_h(y^k, x^{k-1}) - D_h(u, x^{k-1}).$$

So:

$$\lambda_k(f(y^k) - f(u)) \leq D_h(u, x^{k-1}) - D_h(u, y^k) - D_h(y^k, x^{k-1}). \tag{4}$$

On the other hand,

$$\forall u \in \overline{S}, f(x^k) + \lambda_k^{-1} D_h(x^k, x^{k-1}) \leq f(u) + \lambda_k^{-1} D_h(u, x^{k-1}) + \varepsilon_k. \tag{5}$$

Replacing u with y^k in (5) we obtain:

$$\lambda_k(f(x^k) - f(y^k)) \leq D_h(y^k, x^{k-1}) - D_h(x^k, x^{k-1}) + \lambda_k \varepsilon_k. \tag{6}$$

Replacing u with x^k in (4) we have:

$$\lambda_k(f(y^k) - f(x^k)) \leq D_h(x^k, x^{k-1}) - D_h(x^k, y^k) - D_h(y^k, x^{k-1}). \tag{7}$$

(6) and (7) $\Rightarrow 0 \leq -D_h(x^k, y^k) + \lambda_k \varepsilon_k \Rightarrow D_h(x^k, y^k) \leq \lambda_k \varepsilon_k.$ □

Corollary 4.7. *If ∇h is a strongly convex operator, then it exists $\alpha > 0$ such that*

$$\|x^k - y^k\| \leq \sqrt{\frac{2}{\alpha}\lambda_k \varepsilon_k}.$$

Proof.
(b) From the Proposition 4.6, $D_h(x^k, y^k) \leq \lambda_k \varepsilon_k$. ∇h is a strongly convex operator, from the Proposition 3.9. [5], it exists $\alpha > 0$ such that:

$$\frac{\alpha}{2}\|x^k - y^k\|^2 \leq D_h(x^k, y^k).$$

Then $\|x^k - y^k\| \leq \sqrt{\frac{2}{\alpha}\lambda_k \varepsilon_k}$. □

Proposition 4.8. *We assume:*

(a). $\{x^k\}$ generated by $P_3(h)$ is bounded.

(b). h is twice Continuously Differentiable on S and $D_h(.,.)$ is convex jointly.

(c). ∇h is a strongly convex operator

Then

$$D_h(x^*, x^k) \to l \geq 0, \; \forall\; x^* \in Argminf.$$

Proof. From Theorem 4.5, let $Argminf$. Let $x^* \in Argminf$, from (4) we have:

$$D_h(x^*, y^k) + D_h(y^k, x^{k-1}) - D_h(x^*, x^{k-1}) \leq 0.$$

whence

$$D_h(y^k, x^{k-1}) \leq D_h(x^*, x^{k-1}) - D_h(x^*, x^k) + D_h(x^*, x^k) - D_h(x^*, y^k).$$

(2) $\Rightarrow \partial(D_h(x^*, .))(y) = H(y)(y - x^*)$, where $H(y) = \nabla^2 h(y)$. So

$$D_h(x^*, y^k) - D_h(x^*, x^k) \geq \langle y^k - x^k, H(x^k)(x^k - x^*)\rangle.$$

It follows that

$$D_h(y^k, x^{k-1}) \leq D_h(x^*, x^{k-1}) - D_h(x^*, x^k) + \|y^k - x^k\|.\|x^* - x^k\|.\|H(x^k)\|$$

h is twice Continuously Differentiable on S and $\{x^k\}$ is bounded, so $\{H(x^k)\}$ is bounded, therefore

$$\exists\ K > 0, \forall\ k, \sqrt{\frac{2}{\alpha}}\|x^* - x^k\|.\|H(x^k)\| \leq K.$$

Whence

$$D_h(y^k, x^{k-1}) \leq D_h(x^*, x^{k-1}) - D_h(x^*, x^k) + K\sqrt{\lambda_k \varepsilon_k}. \tag{8}$$

Since $D_h(y^k, x^{k-1}) \geq 0$, we have:

$$D_h(x^*, x^k) \leq D_h(x^*, x^{k-1}) + K\sqrt{\lambda_k \varepsilon_k},$$

$\sum_k \sqrt{\lambda_k \varepsilon_k} < +\infty \Rightarrow D_h(x^*, x^k) \to l \geq 0.$ □

Theorem 4.9. *We assume:*

(1). $0 < \underline{\lambda} \leq \lambda_k$*,*

(2). $\sum \sqrt{\lambda_k \varepsilon_k} < +\infty$*,*

(3). $\{x^k\}$ *generated by* $P_3(h)$ *is bounded,*

(4). h *is twice Continuously Differentiable on* S *and* $D_h(.,.)$ *is convex jointly,*

(5). ∇h *is a strongly convex operator.*

Then

(a). $f(x^k) \to \inf f$.

(b). $x^k \to x^* \in Argminf$.

Proof.

(a). is verified from the Theorem 4.5.

(b). Let $x^* \in Adh\{x^k\}$, it exists then the sub-sequence $\{x^{k_i}\}$ of $\{x^k\}$ such that $x^{k_i} \to x^*$. From H_4, $D_h(x^*, x^{k_i}) \to 0$. From the Proposition 4.8. and (B), we have: $D_h(x^*, x^{k_i}) \to l$. So

$$D_h(x^*, x^k) \to 0$$

According to the Proposition 2.3, $x^k \to x^*$. □

5. Conclusion

The entropic proximal algorithm proposed in this labor constitute a unified framework for the existing algorithms and provide others, thus:

- If $h(.) = \frac{1}{2}\|.\|^2$, then $P(1) \Leftrightarrow P_3(h)$.
- If $\varepsilon_k = 0, \forall\ k$ then $P_2(h) \Leftrightarrow P_3(h)$.

References

[1] Auslender, *Numrrical methods for non-differentiable convex optimization*, Mathematical Programming Studies, 30(1987), 102-126.

[2] M. A. Bahraoui, *Diagonal Bundle Methods for Convex Minimization, Theoritical Aspect*, Int. J. Math. And Appl., 6(4)(2018), 5157.

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

[4] J. B. Hiriart-Urruty, *$\varepsilon-$subdifferentiel calculs: Convex Analysis and Optimization*, Research Notes in Mathematics, Series 57, Pitman Publishers, (1982).

[5] S. Kabbadj, *Entropic Approximation*, Int. J. Math. And Appl., 7(2)(2019), 2338.

[6] B. Lemaire, *About the convergence of the proximal method*, proceedings 6th French German conference on optimisation 1991, Advances in Optimization Lecture Notes, Springer-Verlag, 39-51, (1992).

[7] B. Lemaire, *The proximal algorithm*, International Series Of Numerical Mathematics, 87(1989), 73-87,

[8] B. Martinet, *Perturbation des mthodes d'optimisation - Application*, R.A.I.R.O. Analyse Numrique, 12(2)(1976), 153-171.

[9] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control and Optimization, 14(5)(1975), 877-898.

[10] M. Teboulle and Gong Chen, *Convergence analysis of a proximal-like minimization algorithm using Bregman function*, SIAM Journal on Optimization, 3(3)(1993), 538-543.