Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 15, Number 6 (2019), pp. 840-857

http://www.ripublication.com/gjpam.htm

Entropic Bundle Algorithms

Said Kabbadj

Department of Mathematics , Faculty of Sciences of Meknes,

B.P. 11201, Morocco.

Abstract

Several methods have been used to calculate an approximate critical point of a nonsmooth convex function, in particular the proximal point method and the bundle method. In this paper we propose a class of methods, obtained by combining the bundle method with the entropic distances. In our analysis, the convergence properties obtained cover the convergence results of the bundle method, while giving other.

Key words : nonsmooth optimization , bundle method , proximal point , entropic distance.

2010 Mathematics Subject Classification: 90C25 · 49J52 · 65K10 · 49M05

1. INTRODUCTION

Let's consider the problem of convex optimization defined by :

$$(P): \; min \; \{ \; f(x), \; x \in R^d \; \}$$

Several numerical methods have been used to resolve (P) , especially the methods of the proximal point see $[1, 3, 5, 7, 9, 10, 11, 15, 16]$ and the methods of bundle studied by several authors $[8, 12, 13, 14]$. The pinciple of these is based of the approximaion of f by a function's sequence φ^k normally more simple than f. The principle restriction of these methods is that the function f is at finite values. By coupling the methods of bundle with the entropic distances $[2]$ definid by

$$D_h(x, y) := h(x) - h(y) - < x - y, \nabla h(y) >,$$

we propose a class of methods (EBA) called Entropic Bundle Algorithms . These, in contrary the bundle methods (BM) , allowing to solve (P) when f verifies:

(A_1) : $$f = F + \Psi_{\overline{S}},$$

where S is the function zone of Bregman h and F is a convex function at finite values . In particular , by choosing $h_1(x) = \sum_{i=1}^{i=d} x_i log x_i - x_i$, these methods allow to minimize F on the positive cone of R^d . For $h_0(x) = \frac{1}{2} \parallel x \parallel^2$, we find the bundle method (BM) described by C. Lemarechal and R.Correa [13].

Our notation is fairly standart , $< .,. >$ is the scalar product on R^d, 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) $dom f = \{x \in R^d, f(x) < +\infty\}$ its effective domain,
(2) $\partial_\epsilon f(.) = \{v, f(y) \geq f(.) + < v, y - . > -\epsilon, \forall y\}$ its $\varepsilon - subdifferential$,
(3) $arg \min f = \{x \in R^d, f(x) = \inf f\}$ its argmin f,
(4) $\varepsilon -\ arg \min f = \{x \in R^d, f(x) \leq \inf f + \varepsilon\}$ its $\varepsilon - argminf$.

2. PRELIMIMARIE

In this section, we remind some theoretical properties of entropic approximations studied by Kabbadj in [6]. These results are necessary for the analysis of the methods proposed in section 3 and 4 .

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

$$D_h(x, y) := h(x) - h(y) - < x - y, \nabla h(y) > .$$

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 $\{x^k\}_k$ and $\{y^k\}_k$ are two sequences of S such as : $D_h(x^k, y^k) \longrightarrow 0$ and $x^k \longrightarrow x^* \in S$, then $y^k \longrightarrow x^*$.

Definition 2.1.

(i) 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.
(ii) $D_h(.,.)$, 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 $H_5\}$.
$\Gamma_0(R^d) = \{f : R^d \rightarrow R \cup \{+\infty\}$ proper , closed and convex $\}$

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

$$D_h(a,b) + D_h(b,c) - D_h(a,c) =< a - b, \nabla h(c) - \nabla h(b) > .$$

$D_h(.,.)$ is not a distance because the properties of the symetry and the triangle inegality are not verfied, however positivity is ensured by the proposition below.

Proposition 2.3. If $h \in A(S)$, then :

$$D_h(x,y) = \begin{cases} 0 \ \ if \ \ x = y, \\ > 0 \ \ if \ \ x \neq y \end{cases}$$

Theorem 2.4. [6] Let $f \in \Gamma_0(R^d)$ and $h \in A(S)$ such as $dom f \cap \overline{S} \neq \phi$. If one of the two following conditions are verified :

(i) $\inf_{\overline{S}} f > -\infty$ and h verify H_3

(ii) $Im \nabla h = R^d$.

Then for all $x \in S$, for all $\lambda > 0$, the function $u \longmapsto f(u) + \lambda^{-1} D_h(u, v)$ has a unique minimum point on $\overline{S}$.

Definitions 2.5. f and h verify the hypothesis of the Theorem 2.4.
i) The entropic approximation of f compared to h, of parameter $\lambda(\lambda > 0)$ is the function defined by :

$$f_{h\lambda}(x) := \inf_{y \in \overline{S}} \{f(y) + \lambda^{-1} D_h(y, x)\}, \forall x \in S.$$

ii) The application entropic proximal of f comparing to h , of parameter λ is the operator defined by :

$$h_\lambda^f(x) := prox_{\lambda f}^h(x) := \arg\min_{y \in \overline{S}} \{f(y) + \lambda^{-1} D_h(y, x)\}, \forall x \in S.$$

Proposition 2.6. [6] Let $h \in A(S)$ and $f \in \Gamma_0(R^d)$ such as :

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

(b) $Im \nabla h = R^d$.

Then : $\forall x \in S, \; \forall \lambda > 0$

$$h_\lambda^f(x) \in S \tag{1}$$

$$\frac{\nabla h(x) - \nabla h(h_\lambda^f(x))}{\lambda} \in \partial f(h_\lambda^f(x)) \tag{2}$$

Some examples of Bregman functions are given below.

Example 2.7. If $S_0 = R^d$ and $h_0(x) = \frac{1}{2} \parallel x \parallel^2$ then

$$D_{h_0}(x, y) = \frac{1}{2} \parallel x - y \parallel^2 .$$

Example 2.8. If $S_1 = R_{++}^d := \{x \in R^d / x_i > 0, i = 1, ..., d\}$ and

$$h_1(x) = \sum_{i=1}^{i=d} x_i log x_i - x_i \; ; \forall x \in \overline{S_1},$$

with the convention : $0 \log 0 = 0$, then

$$D_{h_1}(x,y) = \sum_{i=1}^{d} x_i \log \frac{x_i}{y_i} + y_i - x_i, \forall (x,y) \in \overline{S}_1 X S_1.$$

Example 2.9. If $S_2 =]-1,1[^d$ and $h_2(x)$ = -$\sum_{i=1}^{i=d} \sqrt{1-x_i^2}$, then:

$$D_{h_2}(x,y) = h_2(x) + \sum_{i=1}^{d} \frac{1-x_i y_i}{\sqrt{1-y_i^2}},\ \forall (x,y) \in \overline{S_2} X S_2.$$

We easily verifies that

$$h_i \in B(S_i), i = 0,1,2.$$

3. MAIN RESULTS

In this section, we analyze an algorithm that plays a decisive role in establishing the convergence of the Entropic Bundle Algorithms proposed in section 4 , we suggest that h and f satisfy the following hypotheses

Assumptions A :
(i) $h \in B(S)$ such that $Im \nabla h = R^d$
(ii) $f \in \Gamma_0(R^d)$ such that $S \cap ri(domf) \neq \emptyset$

Algorithm 1: Inexact Entropic Proximal (IEP)

1: input: $x^0 \in S \cap ri(domf)$
2: for $n = 1, 2, ...$ with $\lambda_n > 0$, find $x^n \in S$ and $\varepsilon_n \geq 0$ such that $f(x^n) \leq f(x^{n-1})$ and

$$\frac{\nabla h(x^{n-1}) - \nabla h(x^n)}{\lambda_n} \in \partial_{\varepsilon_n} f(x^n)$$

Proposition 3.1. The sequence $\{x^n\}_n$ defined by (IEP) exists and verified for all $n \in N^*$:

$$x^n \in \varepsilon_n - argmin\{f(u) + \frac{1}{\lambda_n} D_h(u, x^{n-1}), u \in \overline{S}\}.$$

Proof. Existence: is deduced trivially from (2).

$$\Omega^n := \frac{\nabla h(x^{n-1}) - \nabla h(x^n)}{\lambda_n} \in \partial_{\varepsilon_n} f(x^n)$$

$\Rightarrow f(u) \geq f(x^n) + < u - x^n, \Omega^n > -\varepsilon_n$. By applying the L emma 2.2., we have :

$$f(u) \geq f(x^n) + \lambda_n^{-1}[D_{h_n}(u, x^n) + D_{h_n}(x^n, x^{n-1}) - D_{h_n}(u, x^{n-1})] - \varepsilon_n \quad (3)$$

$\Rightarrow f(x^n) + \lambda_n^{-1} D_{h_n}(x^n, x^{n-1}) \leq f(u) + \lambda_n^{-1} D_{h_n}(u, x^{n-1}) + \varepsilon_n, \forall u \in \overline{S}$

$\Rightarrow x^n \in \varepsilon_n - argmin\{f(u) + \frac{1}{\lambda_n} D_h(u, x^{n-1}), u \in \overline{S}\}$. □

Remark 3.2. For $\varepsilon_n = 0, \forall n \in N^*$ the algorithm (IEP) is written :

$$x^n = argmin\{f(u) + \frac{1}{\lambda_n} D_h(u, x^{n-1}), u \in \overline{S}\},$$

which have nothing but the entropic proximal algorithm (EP) studied by $[4, 18]$. (IEP) is then an inexact version of the algorithm (EP) . The convergence results developed below cover the convergence properties given in $[18]$.

Proposition 3.3. (Summability) If $f^* := \inf_{y \in \overline{S}} > -\infty$,$\sum_{n=1}^{\infty} \lambda_n \varepsilon_n < +\infty$ and $0 < \lambda_n \leq \overline{\lambda}$, then,

$$\sum_{n=1}^{\infty} D_h(x^n, x^{n-1}) < +\infty \text{ and } \sum_{n=1}^{\infty} D_h(x^{n-1}, x^n) < +\infty$$

Proof. From (3) , we have :

$$\lambda_n(f(x^n) - f(u)) \leq [D_h(u, x^{n-1}) - D_h(u, x^n) - D_h(x^n, x^{n-1})] + \varepsilon_n \lambda_n \quad (4)$$

Put $u = x^{n-1}$ in (4), we have

$$D_h(x^n, x^{n-1}) + D_h(x^{n-1}, x^n) \leq \overline{\lambda}(f(x^{n-1}) - f(x^n)) + \lambda_n \varepsilon_n \quad (5)$$

$$\Rightarrow \sum_{n=1}^{n=p} D_h(x^n, x^{n-1}) + D_h(x^{n-1}, x^n) \le \overline{\lambda} \sum_{n=1}^{n=p} (f(x^{n-1}) - f(x^n)) + \sum_{n=1}^{n=p} \lambda_n \varepsilon_n$$
$$\Rightarrow \sum_{n=1}^{n=p} D_h(x^n, x^{n-1}) + D_h(x^{n-1}, x^n) \le \overline{\lambda}(f(x^0) - f^*) + \sum_{n=1}^{n=p} \lambda_n \varepsilon_n.$$

$\sum_{n=1}^{\infty} \lambda_n \varepsilon_n < +\infty \Rightarrow \sum_{n=1}^{\infty} D_h(x^n, x^{n-1}) < +\infty$ and $\sum_{n=1}^{\infty} D_h(x^{n-1}, x^n) < +\infty.$ □

Proposition 3.4. (Global estimate in function values)
We suppose that $\arg\min_{\overline{S}} f \neq \emptyset$. Let $t_p := \sum_{n=1}^{p} \lambda_n$.

(a) $\forall u \in \overline{S}, \forall p \in N^*$, for $x^* \in \arg\min_{\overline{S}} f$

$$f(x^p) - f^* \le \frac{1}{t_p}[D_h(x^*, x^0) + \sum_{n=1}^{\infty} \lambda_n \varepsilon_n] \tag{6}$$

(b) If $\lambda_n = \lambda, \forall n \in N^*$,

$$f(x^p) - f^* = O(\frac{1}{p}) \tag{7}$$

Proof. (a) We have :

$$\lambda_n(f(x^n) - f(u)) \le D_h(u, x^{n-1}) - D_h(u, x^n) + \varepsilon_n \lambda_n \tag{8}$$

$$\Rightarrow \sum_{n=1}^{n=p} \lambda_n(f(x^n) - f(u)) \le \sum_{n=1}^{n=p} D_h(u, x^{n-1}) - D_h(u, x^n) + \sum_{n=1}^{n=p} \lambda_n \varepsilon_n$$
$$\Rightarrow \sum_{n=1}^{n=p} \lambda_n(f(x^n) - f(u)) \le D_h(u, x^0) - D_h(u, x^p) + \sum_{n=1}^{n=p} \lambda_n \varepsilon_n$$

Since $\{f(x^n)\}$ is decreasing, we have :

$$(f(x^p) - f(u)) \sum_{n=1}^{n=p} \lambda_n \le D_h(u, x^0) - D_h(u, x^p) + \sum_{n=1}^{n=p} \lambda_n \varepsilon_n \tag{9}$$

then,

$$f(x^p) - f(u) \le \frac{1}{t_p}[D_h(u, x^0) + \sum_{n=1}^{\infty} \lambda_n \varepsilon_n]. \tag{10}$$

(b)
$t_p = \lambda p \Rightarrow (7).$. □

Now , we derive a global convergence of the sequence generated by (IEP) to a minimizer of (P) .

Theorem 3.5.
(a) If $\sum \lambda_n = +\infty$ and $\varepsilon_n \to 0$ then $f(x^n) \to \inf\limits_{\overline{S}}$
(b) If, moreover $\arg\min\limits_{\overline{S}} f \neq \emptyset$ and $\sum \lambda_n \varepsilon_n < +\infty$ then $x^n \to x^* \in \arg\min\limits_{\overline{S}} f$

Proof. (a)
$(10) \Rightarrow f(x^p) \leq f(u) + \frac{D_h(u,x^0)}{t_p} + \frac{\sum\limits_{n=1}^{n=p} \lambda_n \varepsilon_n}{t_p}$, when $t_p \to +\infty$,

$$\overline{lim} f(x^p) \leq \inf_{\overline{S}} . \tag{11}$$

As $\underline{lim} f(x^p) \geq \inf\limits_{\overline{S}}$, we deduce

$$lim f(x^p) = \inf_{\overline{S}} .$$

(b) Let $x^* \in \arg\min\limits_{\overline{S}} f$, we put $u = x^*$ in (8), we have

$$D_h(x^*, x^n) \leq D_h(x^*, x^{n-1}) + \lambda_n \varepsilon_n \tag{12}$$

Since $\sum \lambda_n \varepsilon_n < +\infty$, we have

$$D_h(x^*, x^n) \to l \geq 0 \tag{13}$$

From H_3, $\{x^n\}_n$ is bounded . Let $u^* \in Adh\{x^n\}$, it exists then a sub-sequence $\{x^{n_i}\}$ of $\{x^n\}_n$ such as $x^{n_i} \longrightarrow u^*$. $u^* \in \arg\min\limits_{\overline{S}} f$. Indeed :

$$\inf_{\overline{S}} \leq f(u^*) \leq f(x^{n_i}) = \inf_{\overline{S}} \Rightarrow \inf_{\overline{S}} = f(u^*),$$

which shows that $u^* \in \arg\min\limits_{\overline{S}} f$, therefore $D_h(u^*, x^n) \to l'$.

Since $x^{n_i} \longrightarrow u^* \in \overline{S}$, from H_4 , we are getting :

$$D_h(u^*, x^{n_i}) \to 0,$$

then $l' = 0$, i.e., $D_h(u^*, x^n) \to 0$. The convergence of x^n to u^* results immediately from the following lemma :

Lemma 3.6. If $\{x^n\} \subset S$ and $D_h(u^*, x^n) \to 0$ then $x^n \to u^*$.

Proof : From H_3, $\{x^n\}$ is bounded. Let $x^* \in Adh\{x^n\}$, it exists then the sub-sequence $\{x^{n_i}\}$ of $\{x^n\}$ such that $x^{n_i} \to x^*$. From H_5 :

$D_h(u^*, x^{n_i}) \to 0$ and $x^{n_i} \to x^* \in \overline{S} \Rightarrow x^* = u^*$

$$\Rightarrow Adh\{x^n\} = \{x^*\}$$
$$\Rightarrow x^n \to u^*. \qquad \square$$

Corollary 3.7. we suppose that $ri(domf) \subset S$.
(a) If $\sum \lambda_n = +\infty$ and $\varepsilon_n \to 0$ then $f(x^n) \to inff$.
(b) If, moreover $argminf \neq \emptyset$ and $\sum \lambda_n \varepsilon_n < +\infty$ then $x^n \to x^* \in argminf$.

Proof. Is a simple consequence of Theorem 3.5. $\square$

4. ANALYSIS OF THE ENTROPIC BUNDLE ALGORITHMS

In what follows, we will deduce the results of convergence (Theorem 4.5. - 4.6.) for Entropic Bundle Algorithms . First, we need to establish some technical results.

Let $h \in B(S)$. Let $f \in \Gamma_0(R^d)$ and F a convex function with a finite values . We suppose that f verifies :

(A_1) : $$f = F + \Psi_{\overline{S}}$$

Let $\{\varphi^k\}_k$ a sequence of functions belonging to $\Gamma_0(R^d)$ such as :

(A_2) : $\forall\, x \in S, \forall\, \lambda > 0,\ y^k := prox^h_{\lambda\varphi^k}(x)$, exists and belongs to S.

(The conditions of realisation of (A_2) are given in the Remark 4.7.). From (2)

$$\Upsilon^k := \frac{\nabla h(x) - \nabla h(y^k)}{\lambda} \in \partial \varphi^k(y^k).$$

Let's put :

$$\ell^k(y) := \varphi^k(y^k) + < \Upsilon^k, y - y^k >$$
$$\overline{\ell}^k(y) := \ell^k(y) + \lambda^{-1} D_h(y, x)$$
$$\overline{\varphi}^k(y) := \varphi^k(y) + \lambda^{-1} D_h(y, x).$$

Furthermore we suppose :

(A_3) : $\varphi^k \leq f, \ \forall \ k = 1, ...$

(A_4) : $\ell^k \leq \varphi^{k+1}.$

(A_5) : $f(y^k) + < g^k, y - y^k > \leq \varphi^{k+1}(y)$ where : $g^k \in \partial F(y^k).$

Proposition 4.1. Let $h \in B(S)$. Let $\{x^k\} \subset S$ bounded and $\{y^k\} \subset \overline{S}$ bounded, such as $D_h(y^k, x^k) \to 0$. Then $\| y^k - x^k \| \to 0$.

Proof. Since $\{y^k\}$ and $\{x^k\}$ are bounded , the sequence $\{\| y^k - x^k \|\}_k$ is bounded too. Let's show that $Adh\{\|y^k - x^k\|\} = \{0\}$. Let $\delta \in Adh\{\|y^k - x^k\|\}$, it exists a sub-sequence $\{\|y^{k_i} - x^{k_i}\|\}_{k_i}$ such as:

$$\|y^{k_i} - x^{k_i}\| \to \delta.$$

$\{x^{k_i}\}$ is bounded , it exists $\{x^{k_j}\}$ such that $x^{k_j} \to x^*$. Otherwise

$$\{x^k\} \subset S \Rightarrow x^* \in \overline{S}.$$

From H_5 ,

$$D_h(y^{k_j}, x^{k_j}) \to 0 \text{ and } x^{k_j} \to x^* \Rightarrow y^{k_j} \to x^*.$$

It follows that $\| y^{k_j} - x^{k_j} \| \to 0$, which means : $\delta = 0$. □

Proposition 4.2. If f verifies A_1 and $\{\varphi^k\}_{k\geq 1}$ verifies A_2, A_3, A_4 and A_5 , then :

(i) $0 \leq f(y^k) - \varphi^k(y^k) \to 0,$

(ii) $y^k \to prox^h_{\lambda f} x.$

Proof . (i)

$$\overline{\ell}^k(y) - \overline{\ell}^k(y^k) = \ell^k(y) + \lambda^{-1}D_h(y,x) - \ell^k(y^k) - \lambda^{-1}D_h(y^k,x)$$

$$\begin{aligned} &=< \Upsilon^k, y - y^k > +\lambda^{-1}D_h(y,x) - \lambda^{-1}D_h(y^k,x) \\ &= \lambda^{-1} < \nabla h(x) - \nabla h(y^k), y - y^k > +\lambda^{-1}D_h(y,x) - \\ &\quad \lambda^{-1}D_h(y^k,x). \end{aligned}$$

From the Lemme 2.2.,

$$< \nabla h(x) - \nabla h(y^k), y - y^k >= D_h(y,y^k) + D_h(y^k,x) - D_h(y,x)$$

It follows that :

$$\overline{\ell}^k(y) = \overline{\ell}^k(y^k) + \lambda^{-1}D_h(y,y^k) \tag{14}$$

For $x \in S$,

$$\begin{aligned} f(x) \geq \varphi^{k+1}(x) &= \overline{\varphi}^{k+1}(x) \\ \geq \overline{\varphi}^{k+1}(y^{k+1}) &= \overline{\ell}^{k+1}(y^{k+1}) \\ \geq \overline{\ell}^k(y^{k+1}) &= \overline{\ell}^k(y^k) + \lambda^{-1}D_h(y^{k+1},y^k). \end{aligned} \tag{15}$$

We deduce that

$$f(x) \geq \overline{\ell}^k(y^k) \tag{16}$$

$$\overline{\ell}^{k+1}(y^{k+1}) \geq \overline{\ell}^k(y^k) + \lambda^{-1}D_h(y^{k+1},y^k) \tag{17}$$

Since $x \in dom f$, from (16) and (17), we can deduce that

$$\overline{\ell}^k(y^k) \to \ell \in R. \tag{18}$$

On the other hand

$$(17) \Rightarrow D_h(y^{k+1}, y^k) \to 0.$$

For y fixed in (14) , and due to (18), we deduce that the sequence $\{D_h(y, y^k)\}_k$ is bounded. From H_3, $\{y^k\}$ is bounded . By vertue of the Proposition 4.1. ,

$$\|y^{k+1} - y^k\| \to 0. \tag{19}$$

From the other hand ,

$$f(y^{k+1}) - f(y^k) \geq \varphi^{k+1}(y^{k+1}) - f(y^k) \geq < g^k, y^{k+1} - y^k >,$$

and from (A_1) and (A_2) ,

$$F(y^{k+1}) - F(y^k) \geq \varphi^{k+1}(y^{k+1}) - f(y^k) \geq < g^k, y^{k+1} - y^k >,$$

F is at finite values , so locally Lipschitz , which leads , with (19),

$$F(y^{k+1}) - F(y^k) \to 0.$$

Since $g^k \in \partial F(y^k)$ and $\{y^k\}$ is bounded, $\{g^k\}$ is bounded, then

$$\varphi^{k+1}(y^{k+1}) - f(y^k) \to 0.$$

Otherwise ,

$$0 \leq f(y^k) - \varphi^k(y^k) = F(y^k) - F(y^{k-1}) + F(y^{k-1}) - \varphi^k(y^k).$$

It follows that

$$f(y^k) - \varphi^k(y^k) \to 0.$$

ii. the sequence $\{y^k\}_k$ verifies $\frac{\nabla h(x) - \nabla h(y^k)}{\lambda} \in \partial_{\varepsilon_k} f(y^k)$ where $\varepsilon_k = f(y^k) - \varphi^k(y^k)$, thus the convergence of y^k towards $prox^h_{\lambda f}$ results immediatly from the following Lemme:

Lemma 4.3. Let $\{y^k\}_k \subset S$ such as:

i- $\frac{\nabla h(x) - \nabla h(y^k)}{\lambda} \in \partial_{\varepsilon_k} f(y^k)$,

ii- $\varepsilon_k \to 0$.

Then :

$(a)\ D_h(y^k, prox^h_{\lambda f}x) + D_h(prox^h_{\lambda f}x, y^k) \leq \lambda\varepsilon_k,$
$(b)\ y^k \to prox^h_{\lambda f}x.$

Proof. (a) $\frac{\nabla h(x)-\nabla h(y^k)}{\lambda} \in \partial_{\varepsilon_k} f(y^k)$ and $\frac{\nabla h(x)-\nabla h(prox^h_{\lambda f}x)}{\lambda} \in \partial_{\varepsilon_k} f(prox^h_{\lambda f}x)$, it follows that

$$< \nabla h(y^k) - \nabla h(prox^h_{\lambda f}x), y^k - prox^h_{\lambda f}x > \leq \lambda\varepsilon_k,$$

Let again ,

$$D_h(y^k, prox^h_{\lambda f}x) + D_h(prox^h_{\lambda f}x, y^k) \leq \lambda\varepsilon_k,$$

(b) $D_h(y^k, prox^h_{\lambda f}x) \to 0$, so from H_5, $y^k \to prox^h_{\lambda f}x.$ □

Example 4.4. If h verifies the conditions of the proposition 2.6., Then (A_2) is verified, we give below two exemples of sequences $\{\varphi^k\}$ verifying (A_3), (A_4) and (A_5) :

$$E_1 : \varphi^{k+1}(y) = max\{f(y^i) + < g^i, y - y^i >, 1 \leq i \leq k >\}$$
$$E_2 : \varphi^{k+1}(y) = max\{\ell^k(y), f(y^k) + < g^k, y - y^k >\}.$$

Now we represent below the Entropic Bundle Algorithms obtained by replacing the quadratic kernel in the base method with entropic distances.

Algorithm 2 : Entropic Bundle Algorithm (EBA)

1: input: $m \in]0, 1[$, $x_1 \in S$ and $k(1) := 0$
2: for $n = 1, 2, ...$ with $\lambda_n > 0$ and $\varepsilon_n \geq 0$, do
3: for $k = k(n) + 1, ..., k(n + 1)$, cumpute

$$y^k := prox^h_{\lambda_n \varphi^k} x_n \ ,$$

where $k(n+1)$ is the first natural integer $k > k(n)$ for which an improvement is obtained , precisely :

$$(*) : f(x_n) - f(y^k) \geq m\left[f(x_n) - \varphi^k(y^k)\right].$$

4: (i) If the condition (*) is never verified , we put $k(n+1) := +\infty$ and stop
(ii) If not , we put $x_{n+1} = y^{k(n+1)}$ go to step 3.

Theorem 4.5. Let $\{x_n\}$ the generated sequence by the Entropic Bundle Algorithms . From two things the one :

- (a) It exists n such as $k(n+1) = +\infty$ then x_n minimizes f.

- (b) The sequence k(n) is infinite , then if $\sum_n \lambda_n = +\infty$, $f(x_n) \to inf f$. If furthermore , $0 < \lambda_k \leq \overline{\lambda} < +\infty$ and argmin f $\neq \phi$, then

$$x_n \to \overline{x} \in argminf.$$

Proof . (a) (*) is not verified , it exists then a rank n such as :

$$f(x_n) - f(y^k) < m\left[f(x_n) - \varphi^k(y^k)\right], \ \ \forall\, k \geq \ k(n)+1.$$

Which leads :

$$(1-m)\left[f(x_n) - f(y^k)\right] < m\left[f(y^k) - \varphi^k(y^k)\right].$$

According to the Proposion 4.2., $f(y^k) - \varphi^k(y^k) \to 0$, then $f(x_n) \leq f(prox^h_{\lambda_{nf}} x_n)$. Otherwise $f(prox^h_{\lambda_{nf}} x_n) + \lambda_n^{-1} D_h(prox^h_{\lambda_{nf}} x_n, x_n) \leq f(x_n)$, then

$$f(prox^h_{\lambda_{nf}} x_n) + \lambda_n^{-1} D_h(prox^h_{\lambda_{nf}} x_n, x_n) \leq f(prox^h_{\lambda_{nf}} x_n)$$

Since $prox^h_{\lambda_{nf}} x_n \in \overline{S} = domf$, we have $D_h(prox^h_{\lambda_{nf}} x_n, x_n) = 0$, i.e. $prox^h_{\lambda_{nf}} x_n = x_n \in S$. From (2), $0 \in \partial f(x_n)$, then $x_n \in Argminf$.

(b) $\frac{\nabla h(x_n) - \nabla h(x_{n+1})}{\lambda_n} \in \partial \varphi^{k(n+1)}(x_{n+1}) \Rightarrow \frac{\nabla h(x_n) - \nabla h(x_{n+1})}{\lambda_n} \in \partial_{\varepsilon_{n+1}} f(x_{n+1})$, where $\varepsilon_{n+1} = f(x_{n+1}) - \varphi^{k(n+1)}(x_{n+1})$.

From (*) , we have

$$0 \leq \varepsilon_{n+1} \leq \frac{1-m}{m}[f(x_n) - f(x_{n+1})],$$

therefore, $f(x_n)$ is decreasing and $\sum_n \lambda_n \varepsilon_n < +\infty$. as ri(domf) = S, therefore by application of the Corollary 3.7. , we deduce the result . □

In order to avoid the calculation of y^k indefinitely, and since $f(y^k) - \varphi^k(y^k) \to 0$, we propose the following algorithm :

Algorithm 3: A version of the Entropic Bundle Algorithm (EBA)

1: input: $x_1 \in S$ and $k(1) := 0$
2: for $n = 1, 2, ...$ with $\lambda_n > 0$ and $\alpha_n \geq 0$, do
3: for $k = k(n) + 1, ..., k(n + 1)$, cumpute

$$y^k := prox^h_{\lambda_n \varphi^k} x_n \ ,$$

where $k(n + 1)$ is the first natural integer $k > k(n)$ such that :

(**) $$f(y^k) - \varphi^k(y^k) \leq \alpha_{n+1}$$

4: put : $x_{n+1} = y^{k(n+1)}$ go to step 3.

As before, we have $\frac{\nabla h(x_n) - \nabla h(x_{n+1})}{\lambda_n} \in \partial_{\varepsilon_{n+1}} f(x_{n+1})$, where

$$\varepsilon_{n+1} = f(x_{n+1}) - \varphi^{k(n+1)}(x_{n+1}).$$

Since $\varepsilon_{n+1} \leq \alpha_{n+1}$, the convergence result of this algorithm is given by :

Theorem 4.6. Let $\{x_n\}$ the generated sequence by Algorithm 3
(i) If $\sum_n \lambda_n = +\infty$ and $\alpha_k \to 0$ then $lim\ inf f(x_n) = inf f$.
(ii) If, moreover , $0 < \lambda_k \leq \overline{\lambda} < +\infty$, $\sum_k \alpha_k < +\infty$ and Argmin f $\neq \phi$ then

$$lim f(x_n) = inf f \text{ and } x_n \to \overline{x} \in argmin \text{ f.}$$

Proof. Immediate consequence of Corollary 3.7. □

Remark 4.7.

(i) If $Im\nabla h = R^d$ and $ri(dom\varphi^k) \cap S \neq \emptyset$, then according to the Proposition 2.6. (A_2) is always verified

(ii) If $S = R^d$, then (A_2) is checked and the Algorithms (2) and (3) solve (P) where f is finite values

5. CONCLUSION

The proposed class of methods (EBA) contains the bundle method (BM) described by Correa-Lemaréchal [13] and solves convex optimization problems under positive constraints:
(i) If $h = h_0$ then $(EBA) \Leftrightarrow (BM)$
(ii) For F: $R^d \to R$ convex , let the problem of convex optimization:

$$(P') : min\{F(x), x \in \Omega\} \text{ where } \Omega \text{ is the positive cone of } R^d$$

For $h \in B(S)$, with $S = int(\Omega)$, (for example $h = h_1$) , (EBA) solves (P') without using the dual problem, like Teboulle [17] and Eckstein [4] .

REFERENCES

[1] A. Auslender, Numrrical methods for non-differentiable convex optimization. Mathematical Programming Studies , n 30, p.102-126, Ed. Vial , Cornet, Nguyen, 1987.

[2] L. M. Bregman, The relaxation method of nding the common point of convex sets and its application to the solution of problems in conve programming, USSR Computational Mathematics and Mathematical Physics 7 (3) (1967) 200-217.

[3] C.L. Chiriac, On the splitting methods and the proximal point algorithm for maximal monotone operators, Analele Stiintifice ale Universitatii Ovidius Constanta, Vol. 12(2), 2004, 83–92

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

[5] O. Gouler, Ergodic convergence in proximal point algorithms with Bregman functions in Advancesin Optimization and Approximation, vol. 1 of Nonconvex Optimization and Its Applications, pp. 155–165, Kluwer Academic, Dordrecht, The Netherlands, 1994.

[6] S.Kabbadj, , Méthodes proximales entropiques, Thèse Université Montpellier II ,(1994).

[7] H.Khatibzadeh , S.Ranjbar, Inexact inertial proximal algorithm for maximal monotone operator , Analele Stiintifice ale Universitatii Ovidius Constanta, Vol. 23(2),2015, 133–146.

[8] K.C.Kiwiel, Proximity control in bundle methods for convex nondifferentiable minimization . Mathematical Programming , vol 46, 1, p. 105-122, 1990.

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

[10] B.Lemaire, Coupling optimization methods and variational convergence. Trends in Mathematical Optimization International Series of Numerical Mathematics , Birkhauser, Basel, p.163-179, 1988.

[11] B.Lemaire,The proximal algorithm . International Series Of Numerical Mathematics, vol. 87, p.73-87, 1989.

[12] C.Lemaréchal, An of Davidon methods to non differentiable problems , in nondifferentiable optimization, M.L.Balinski , P.Wolfe. Mathematical Programming Study , vol.3, p.95-109, 1975.

[13] C.Lemaréchal and R.Correa, Convergence of some algorithms for convex minimization , Mathematical Programming , vol. 62 p. 261-275, 1993.

[14] C.Lemaréchal , J.J .Strodiot and A.Bihain, , On a bundle algorithm for nonsmooth optimization , in nonlinear programming 4, O.L. Mangasarian, R.R.Meyer, S.M Robinson, Academic Press, p. 245-282, 1981.

[15] B.Martinet, Perturbation des méthodes d'optimisation - Application , R.A.I.R.O. Analyse numérique, vol.12, n 2, p. 153-171,1976.

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

[17] M.Teboulle, Entropic proximal mappings with applications to nonlinear programming .Mathematics of Operations Research, vol. 17, n 3, p.670-690,1992.

[18] M.Teboulle and Gong Chen, Convergence analysis of a proximal-like minimization algorithm using Bregman function. SIAM journal on Optimization, Vol.3, n 3 , p. 538-543,1993.