AMO - Advanced Modeling and Optimization,Volume 14, Number 1, 2012

EPSILON-PROXIMAL POINT ALGORITHMS FOR NONDIFFERENTIABLE CONVEX OPTIMIZATION PROBLEMS AND APPLICATIONS

N. DAILI AND KH. SAADI

Cité des 300 Lots. Yahiaoui. 51, rue Chikh Senoussi. 19000 Sétif, Algeria.
email: nourdaili_dz@yahoo.fr

Abstract. In this work, we study the Penalty-proximal method and let us give some algorithms of resolution known as the ε-Proximal point algoritms. These results will be given within the framework of a nondifferentiable optimization. Implementable algorithms for constrained nonsmooth convex programs are given.

Key words : nondifferentiable convex optimization, ε-Proximal penalty methods, duality, convergence of algorithms.
AMS subject classification : Primary 90C25, 49M29; Secondary 90C56

1. Introduction

The purpose of this work is double. On one hand, we are going to present in a detailed enough way, the general principles of penalty methods. On the other hand, we are going to study the works of A. Auslender et Al ([1]) and let us give algorithms of resolution known as ε-Proximal penality. These results will be given within the framework of a nondifferentiable optimization.

Results of diverse numerical essays will illustrate the behavior of the algorithm and finalize its efficiency will be afterward presented.

Consider the following optimization problem:

$$(\mathcal{P}) \quad \begin{cases} \alpha := Inf f(x) \\ \text{subject to } x \in C, \end{cases}$$

where

. $C := \{x \in \mathbb{R}^n : g_i(x) \leq 0, \;\; i = 1, ..., m\}$

. f is a finite value and not necessarly differentiable convex function;

. $g_i, i = 1, ..., m$, are convex functions of $\mathcal{C}^1$.

Suppose that $\lim\limits_{(\|x\| \longrightarrow +\infty)} f(x) = +\infty$ (i.e., f is inf-compact).

The penalty methods constitute a family of particularly interesting algorithms of the double point of view of the principal simplicity and the practical efficiency. There are two variants for these methods, the most used: the exterior penalty methods and the interior penalty methods.

The ε-Proximal penalty method, be going to lead us to introduce the exterior penalty methods. For it we content with presenting the principle of these methods. Then, their principle comes because the problem

$$(\mathcal{P}) \quad \begin{cases} \alpha := Inf f(x) \\ \text{subject to } x \in C \end{cases}$$

is equivalent to the following unconstrained problem:

$$(\mathcal{EP}) \quad \alpha_e := \underset{x \in \mathbb{R}^n}{Inf} \left\{ \varphi(x) = f(x) + \Psi_C(x) \right\},$$

*AMO - Advanced Modeling and Optimization. ISSN: 1841-4311.

where

$$\Psi_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ \\ +\infty & \text{if } x \notin C \end{cases}$$

is the indicator function of the set C.

So, the general principle of these methods consists in replacing the probblem $(\mathcal{P})$ by the following unconstrained problem:

$$(\mathcal{P}_r) \quad \alpha_r := \underset{x \in \mathbb{R}^n}{Inf} \left\{ \varphi(x, r) = f(x) + rh(x) \right\},$$

where $r > 0$ and $h(x)$ is a function defined on $\mathbb{R}^n$ such that

$$\begin{cases} (i) \;\; h \text{ is continuous;} \\ \\ (ii) \;\; h(x) \geq 0 \;\; \forall x \in \mathbb{R}^n; \\ \\ (iii) \;\; h(x) = 0 \iff x \in C \end{cases}$$

so that when $(r \longrightarrow +\infty)$ the obtained solution $\overline{x}(r)$ tends to $\overline{x} \in S$ (solutions set of $(\mathcal{P})$).

r is called penalty coefficient, h is called exterior penalty function. It is not difficult to build functions $h(x)$. For example:

$$h(x) := \sum_{i=1}^{m} (g_i(x))^2$$

called classical penalty function;

$$h(x) := \sum_{i=1}^{m} g_i^+(x)$$

called interior penalty function, where

$$g_i^+(x) = Max(0, g_i(x)).$$

The choice of an appropriate value of the penalty coefficient r results from a compromise:

. on one hand, r must be chosen large enough so that the point $\overline{x}(r)$ obtained is close to all the solutions;

. on the other hand, if r is chosen too large, the function φ can be ill-conditioned where from numerical difficulties in the search for the optimum without constraint.

This explains why the penalty methods are generally implemented under iterative shape in the following way:

We begin by choosing a penalty coefficient r_1 of not too much raised value (to avoid the numerical difficulties) then we resolve the problem without constraints:

$$(\mathcal{P}_{r_1}) \quad \alpha_{r_1} := \underset{x \in \mathbb{R}^n}{Inf} \left\{ \varphi(x, r_1) = f(x) + r_1 h(x) \right\}.$$

Let $\overline{x}(r_1)$ be the obtained point.

If the quantity $r_1 h(\overline{x}(r_1))$ is enough weak, $\overline{x}(r_1)$ is a good approximation of the optimum, and the calculations are ended. Should the opposite occur, we shall thus choose a penalty coefficient $r_2 > r_1$ (for example: $r_2 = 10r_1$) and we shall resolve the new problem without constraint:

$$(\mathcal{P}_{r_2}) \quad \alpha_{r_2} := \underset{x \in \mathbb{R}^n}{Inf} \left\{ \varphi(x, r_2) = f(x) + r_2 h(x) \right\},$$

we shall obtain a new point $\overline{x}(r_2)$, and so on.

The following algorithm shows the necessary steps of the resolution:

Algorithm 1:

Step 0: $(k = 0)$

We begin by choosing a penalty coefficient r_0, a precision $\delta > 0$, $(k = 0)$.

Step 1: $(k \geq 0)$

We resolve the problem without constraints

$$\alpha_{r_k} := \underset{x\in\mathbb{R}^n}{Inf}\ \{\varphi(x, r_k) = f(x) + r_k h(x)\}\,.$$

Step 2:
Let $\overline{x}(r_k)$ be the obtained solution. If

$$r_k h(\overline{x}(r_k)) < \delta$$

then $\overline{x}(r_k)$ is a good approximation of the optimum and the calculations are ended.
Should the opposite occur, we choose a penalty coefficient $r_{k+1} > r_k$ (for example: $r_{k+1} = 10r_k$), $k \longrightarrow k+1$, we return to the step 1.

Remark 1. In every step k of the previous process, it is advantageous to use the point $\overline{x}(r_{k-1})$ obtained in the step $k-1$ as an initial point of the used algorithm of optimization.

2. Main Results

2.1. **The Proximal Regularization.**

A method allowing to find the minimum of a non necessarily differentiable convex function is the proximal method of J.J. Moreau ([3], [6]).

Its principle is the following one: to the problem

$$\alpha := \underset{x\in\mathbb{R}^n}{Inf}\, f(x) \tag{1}$$

we associate the following problem:

$$\alpha_y := \underset{x\in\mathbb{R}^n,\ y\in\mathbb{R}^n}{Inf} \left\{F(x,y) = f(x) + \frac{1}{2}\|x-y\|^2\right\}. \tag{2}$$

The relaxation algorithm applied to this problem is transformed and engenders a sequence $\{x^k, y^k\}$ such that x^{k+1} be a solution of the problem

$$\alpha := \underset{x\in\mathbb{R}^n}{Inf} \left\{F(x,y^k) = f(x) + \frac{1}{2}\left\|x-y^k\right\|^2\right\}$$

and y^{k+1} be a solution of the problem

$$\alpha := \underset{y\in\mathbb{R}^n}{Inf} \left\{F(x^{k+1},y) = f(x^{k+1}) + \frac{1}{2}\left\|x^{k+1}-y\right\|^2\right\} = x^{k+1}.$$

Thus a simpler iteration : x^{k+1} is a solution of the problem

$$\underset{x\in\mathbb{R}^n}{Inf} \left\{f(x) + \frac{1}{2}\left\|x-x^k\right\|^2\right\}.$$

The following theorem summarizes the most remarkable properties of this method.

Theorem 1. *([3]) Let us suppose f a convex function. Denote φ the following function:*

$$\varphi(y) := \underset{x\in\mathbb{R}^n}{Inf} \left\{f(x) + \frac{1}{2}\|x-y\|^2\right\}. \tag{3}$$

(a) For all $y \in \mathbb{R}^n$, there exists an unique solution $\overline{x}(y)$ of

$$\underset{x\in\mathbb{R}^n}{Inf} \left\{f(x) + \frac{1}{2}\|x-y\|^2\right\}.$$

(b) $\varphi(y)$ is a convex differentiable function of gradient

$$\nabla\varphi(y) = y - \overline{x}(y) \tag{4}$$

and one has

$$\nabla\varphi(y) \in \partial f(\overline{x}(y)).$$

(c) The sequence defined by

$$x^{k+1} = \overline{x}(x^k) \tag{5}$$

converges to an optimal solution of $\underset{x\in\mathbb{R}^n}{Inf} f(x)$.

(*d*) *The iteration* $x^{k+1} = \overline{x}(x^k)$ *also spells*

$$x^{k+1} = x^k + \nabla\varphi(\overline{x}(y)) \tag{6}$$

Definition 1. The function $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$\varphi(y) := \underset{x\in\mathbb{R}^n}{Inf} \left\{ f(x) + \frac{1}{2}\|x-y\|^2 \right\}$$

is called regularized of the function f on $\mathbb{R}^n$.

According to the previous theorem, we indeed understand why φ one is regularized. It preserves the good properties of f (convexity, even minimum) but moreover, it is differentiable.

Proof. (**Theorem 1**)
(*a*) The function

$$\varphi(x) = f(x) + \frac{1}{2}\|x-y\|^2$$

is strictly convex at x, where from the uniqueness of the solution.
(*b*) Let y_1, y_2 and $\overline{x}_1$, $\overline{x}_2$ be the associated solutions.
For $t \in [0,1]$, denote $\overline{x} = t\overline{x}_1 + (1-t)\overline{x}_2$. We have

$$\begin{aligned} t\varphi(y_1) + (1-t)\varphi(y_2) &= tf(\overline{x}_1) + (1-t)f(\overline{x}_2) + \tfrac{1}{2}\|\overline{x}_1 - y_1\|^2 + \tfrac{1-t}{2}\|\overline{x}_2 - y_2\|^2 \\ &\geq f(t\overline{x}_1 + (1-t)\overline{x}_2) + \tfrac{1}{2}\|t(\overline{x}_1 - y_1) + (1-t)(\overline{x}_2 - y_2)\|^2 \\ &= f(\overline{x}) + \tfrac{1}{2}\|\overline{x} - ty_1 - (1-t)y_2\|^2 \\ &\geq \underset{x\in\mathbb{R}^n}{Inf}\left\{ f(x) + \tfrac{1}{2}\|x - ty_1 - (1-t)y_2\|^2 \right\} = \varphi(ty_1 + (1-t)y_2). \end{aligned}$$

For the calculation of the gradient, we choose some direction $w \in \mathbb{R}^n$, and a real $t > 0$. We have

$$\varphi'(y,w) = \lim_{(t\longrightarrow 0)} \frac{\varphi(y+tw) - \varphi(y)}{t}$$

the directional derivative of φ in the direction w. We have

$$\begin{aligned} \varphi(y+tw) - \varphi(y) &\leq f(\overline{x}(y)) + \tfrac{1}{2}\|\overline{x}(y) - (y+tw)\|^2 - \varphi(y) \\ &= \tfrac{1}{2}\|\overline{x}(y) - (y+tw)\|^2 - \tfrac{1}{2}\|\overline{x}(y) - y\|^2 \\ &= \tfrac{t^2}{2}\|w\|^2 + t(y - \overline{x}(y))^t w. \end{aligned}$$

then

$$\frac{\varphi(y+tw) - \varphi(y)}{t} \leq \frac{t}{2}\|w\|^2 + (y - \overline{x}(y))^t w,$$

which implies

$$\varphi'(y,w) \leq (y - \overline{x}(y))^t w.$$

Since

$$\varphi'(y,-w) \geq -\varphi'(y,w)$$

then,

$$\varphi'(y,w) \geq -\varphi'(y,-w) \geq (y - \overline{x}(y))^t w.$$

Thus

$$\varphi'(y,w) = (y - \overline{x}(y))^t w, \text{ for all } w,$$

hence

$$\nabla\varphi(y) = y - \overline{x}(y).$$

Since $\overline{x}(y)$ realize the minimum in the expression (3), we have

$$0 \in \partial(f(\overline{x}(y)) + \frac{1}{2}\|\overline{x}(y) - y\|^2) \Longrightarrow y - \overline{x}(y) = \nabla\varphi(y) \in \partial f(\overline{x}(y)).$$

(c) Denote $\overline{x}$ some solution of the problem $\underset{x\in\mathbb{R}^n}{Inf}\, f(x)$.

Given that $x^{k+1} = \overline{x}(x^k)$, we have necessarily

$$f(x^{k+1}) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2$$

what implies that the sequence $\left\{f(x^k)\right\}_{k\in\mathbb{N}}$ is decraising and since it is lower bounded by $f(\overline{x})$, it converges to $\overline{f}$. Besides, we also deduct that from it the series

$$M = \sum_{k\in\mathbb{N}} \left\|x^{k+1} - x^k\right\|^2 \tag{7}$$

is convergent, and thus

$$\lim_{(k\longrightarrow+\infty)} \left\|x^{k+1} - x^k\right\| = 0. \tag{8}$$

Show that $\left\{x^k\right\}_{k\in\mathbb{N}}$ is bounded. Since φ is convex, we can deduct that

$$\varphi(\overline{x}) \geq \varphi(x^k) + (\nabla\varphi(x^k))^t(\overline{x} - x^k).$$

Hence

$$\varphi(\overline{x}) \geq \varphi(x^k) + (x^k - x^{k+1})^t(\overline{x} - x^k).$$

Besides, $f(x^k) \geq f(\overline{x})$, what leads to

$$(x^k - x^{k+1})^t(\overline{x} - x^k) \leq 0.$$

We can then write

$$\begin{aligned}\left\|x^{k+1} - \overline{x}\right\|^2 &= \left\|x^{k+1} - x^k\right\|^2 + \left\|x^k - \overline{x}\right\|^2 + 2(x^{k+1} - x^k)^t(x^k - \overline{x}) \\ &\leq \left\|x^{k+1} - x^k\right\|^2 + \left\|x^k - \overline{x}\right\|^2.\end{aligned}$$

It holds that for all k,

$$\left\|x^{k+1} - \overline{x}\right\|^2 \leq \left\|x^0 - \overline{x}\right\|^2 + M < +\infty.$$

Thus the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ is bounded, it admits, thus, cluster points. Let x^* one of them. Let $t \in [0,1]$, we can write

$$f(x^{k+1}) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2 \leq f(tx^{k+1} + (1-t)\overline{x}) + \frac{1}{2}\left\|tx^{k+1} - (1-t)\overline{x} - x^k\right\|^2$$

or

$$f(x^{k+1}) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2 \leq tf(x^{k+1}) + (1-t)f(\overline{x}) + \frac{1}{2}\left\|tx^{k+1} - (1-t)\overline{x} - x^k\right\|^2.$$

Let $\left\{x^{k'}\right\}_{k'}$ be a subsequence of $\left\{x^k\right\}_{k\in\mathbb{N}}$ convergent to x^*. The subsequence $\left\{x^{k'+1}\right\}_{k'}$ converges so to x^* because

$$\lim_{(k'\longrightarrow+\infty)} \left\|x^{k'+1} - x^{k'}\right\| = 0.$$

Taking the limit in the previous inequality, we obtain

$$f(x^*) \leq tf(x^*) + (1-t)f(\overline{x}) + \frac{1}{2}\|(1-t)(\overline{x} - x^*)\|^2.$$

Hence

$$f(x^*) - f(\overline{x}) \leq \frac{1-t}{2}\|\overline{x} - x^*\|^2.$$

If we take a limit when t tends to 1, we find that $f(\overline{x}) = f(x^*)$.

(d) We have

$$\nabla\varphi(x^k) = x^k - x^{k+1} \Longrightarrow x^{k+1} = x^k + \nabla\varphi(x^k).$$

An iteration of the Proximal algorithm is thus equivalent to one step of gradient on the regularized function. □

Remark 2. For a concave function h, we deduct the function φ in the following way:

$$\varphi(y) = \underset{z}{Sup}\left\{h(z) - \frac{1}{2}\|y-z\|^2\right\}. \tag{9}$$

Thus

$$\nabla\varphi(y) = (z_y - y),$$

where y realize the Sup in the expression (9).

We thus have the following Proximal algorithm:
Algorithm 2:
Step 0: $(k = 0)$
Let $x^0 \in \mathbb{R}^n$, a precision $\delta > 0$, $(k = 0)$.
Step 1:
Find x^{k+1} solution of the problem

$$\underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \frac{1}{2}\left\|x - x^k\right\|^2\right\}.$$

Step 2:
If $\left\|x^{k+1} - x^k\right\| < \delta$ then x^{k+1} is a good approximation of the optimum and the calculations are ended.
Else, make $k \longrightarrow k+1$ and return to step 1.

Remark 3. The same remark as the Remark 1. In every step k, it is advantageous to use the point x^{k-1} obtained in the step $k-1$ as initial point of the used algorithm of optimization.

To resolve the problem of optimization $(\mathcal{P})$, A. Auslender et Al ([1]) have proposed an algorithm of resolution allowing to find a minimum of $(\mathcal{P})$. They coupled the Proximal method with that of type exterior penalties and also suggested resolving a sequence of unconstrained problems of the following shape:

$$(\mathcal{P}) \quad \alpha := \underset{x\in\mathbb{R}^n}{Inf}\left\{f(x^k) + r_k h(x^k) + \frac{1}{2}\left\|x - x^k\right\|^2\right\} \tag{10}$$

whose every solution is calculated with a precision ε.
Hence, we obtain the following ε-Proximal penalty algorithm:
Algorithm 3:
Step 0: $(k = 0)$
Let us given $x^0 \in \mathbb{R}^n$ a precision $\delta > 0$, ε^0 be given, $(k = 0)$.
Step 1: $(k \geq 0)$
Find x^{k+1} solution of the problem

$$f(x^{k+1}) + r_k h(x^{k+1}) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2 \leq f(x) + r_k h(x) + \frac{1}{2}\left\|x - x^k\right\|^2 + \varepsilon^k \tag{11}$$

Step 2:
Stopping test.
The convergence of this last sequence will be shown in the last paragraph.

2.2. Resolution of The Problem $(\mathcal{P})$.

Consider the following optimisation problem:

$$(\mathcal{P}) \quad \left\{\begin{array}{c} \alpha := Inf f(x) \\ \text{subject to } x \in C, \end{array}\right.$$

where

$$C := \{x \in \mathbb{R}^n : \ g_i(x) \leq 0, \ i = 1, ..., m\}.$$

We introduce the penalty functions h which satisfie the following conditions:

$$\begin{cases} (i) \;\; h \text{ is convex of a } \mathcal{C}^1; \\ (ii) \;\; h(x) \geq 0, \; \forall x \in \mathbb{R}^n; \\ (iii) \;\; h(x) = 0 \iff x \in C. \end{cases}$$

Set

$$S := \arg\min_{x \in C} f(x) \quad ; \quad \alpha := Inf \{f(x) : x \in C\},$$

and $f_k := f + r_k h$. Denote S^k the following set:

$$S^k := \{x \in \mathbb{R}^n : f_k(x) \leq \alpha\}$$

and S^k_δ the following set:

$$S^k_\delta := \{x \in \mathbb{R}^n : f_k(x) \leq \alpha + \delta\}$$

and

$$S^\infty_\delta := \{x \in C : f(x) \leq \alpha + \delta\}.$$

We need also to the following elementary result:

Lemma 1. *Let C be a convex set, then C is bounded $\iff D_C = \emptyset$, where D_C is the directions set of C defined by*

$$D_C := \{d \in \mathbb{R}^n : d \neq 0, \; x + \lambda d \in C, \; \forall x \in C, \; \forall \lambda \geq 0\}.$$

We have the following lemma:

Lemma 2. *If the function f is inf-compact, then f_k is also inf-compact for $k \geq \overline{k}$.*

Proof. We have

$$S \subset S^{k+1} \subset S^k \quad \forall k \quad \text{and} \quad S = \bigcap_k S^k.$$

Indeed;

$$S \subset S^k \quad \forall k \Longrightarrow S \subset \bigcap_k S^k.$$

On the other hand, let $x \in \bigcap_k S^k$, then

$$f_k(x) \leq \alpha, \;\; \forall k \Longrightarrow f(x) + r_k h(x) \leq \alpha, \;\; \forall k.$$

Thus

$$f(x) + \lim_{(k \longrightarrow +\infty)} r_k h(x) \leq \alpha \Longrightarrow \lim_{(k \longrightarrow +\infty)} r_k h(x) = 0.$$

Since $\lim_{(k \longrightarrow +\infty)} r_k = +\infty$, we have $h(x) = 0$. This implies $x \in C$. Hence

$$(f(x) \leq \alpha \;\text{ and }\; x \in C) \Longrightarrow f(x) = \alpha \Longrightarrow x \in S \Longrightarrow \bigcap_k S^k \subset S.$$

Then, it holds that $S = \bigcap_k S^k$.

S is a nonempty compact set because f is inf-compact. Show that there exists $\overline{k} \in \mathbb{N}$ such that $S^{\overline{k}}$ be compact.

Suppose S^k is not compact $\forall k \in \mathbb{N}$. Let $\overline{x} \in S$. We define

$$K_k := \left\{ d \in \mathbb{R}^n : \;\; \|d\| = 1, \; \overline{x} + td \in S^k, \; \forall t > 0 \right\}.$$

We have $K_k \neq \emptyset$. Indeed; S^k is not a bounded closed convex set, then according to the Lemma 2 there exists at least $d \neq 0$ such that

$$\overline{x} + td \in S^k, \;\; \forall t > 0,$$

take $\frac{d}{\|d\|} \in K_k$. Since K_k is compact and $K_{k+1} \subset K_k$, $\forall k$, we have $\underset{k}{\cap} K_k \neq \emptyset$, where

$$\underset{k}{\cap} K_k = \left\{ d \in \mathbb{R}^n : \ \|d\| = 1, \ \overline{x} + td \in \underset{k}{\cap} S^k, \ \forall t > 0 \right\}$$

$$= \{ d \in \mathbb{R}^n : \ \|d\| = 1, \ \overline{x} + td \in S, \ \forall t > 0 \}$$

this implies that S is not a bounded set, thus a contradiction.

Then, $\exists \overline{k} \in \mathbb{N}$ such that $S^{\overline{k}}$ be a compact set (consequently S^k is a compact set for all $k \geq \overline{k}$).

But, we know (classical result of convex analysis) that f is inf-compact $\Longleftrightarrow \exists \lambda_0$ such that $S_{\lambda_0}(f)$ is compact, where $S_{\lambda_0}(f)$ is a level set of f. It holds, according to this result, that f_k is inf-compact. □

Definition 2. Let $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a convex function. We call Prox mapping of h the mapping defined as follows:

$$y \longrightarrow \Psi(y, x) = h(y) + \frac{1}{2} \|y - x\|^2 \quad (x \in \mathbb{R}^n).$$

Remark 4. The mapping Ψ is strongly convex.

Put

$$\operatorname{Prox}(h; x) = \arg \min_{y \in \mathbb{R}^n} \Psi(y, x).$$

We have, according to R.T. Rockafellar ([7], Theo.31.5, p.340),

$$\|\operatorname{Prox}(h; x_1) - \operatorname{Prox}(h; x_2)\| \leq \|x_1 - x_2\| \tag{12}$$

We have the following lemma:

Lemma 3. *If* $\overline{y} = \underset{y \in \mathbb{R}^n}{\arg\min} \Psi(y, x)$ *and* $\Psi(z, x) \leq \Psi(\overline{y}, x) + \varepsilon$ *then*

$$\|z - \overline{y}\| \leq \sqrt{2\varepsilon}.$$

Proof. Since Ψ is strongly convex, we have

$$\Psi(\overline{y}, x) \leq \Psi(y, x) - \frac{1}{2} \|y - \overline{y}\|^2, \text{ for all } y.$$

For $y = z$, we have

$$\Psi(\overline{y}, x) \leq \Psi(z, x) - \frac{1}{2} \|z - \overline{y}\|^2.$$

Then

$$\Psi(\overline{y}, x) \leq \Psi(\overline{y}, x) + \varepsilon - \tfrac{1}{2} \|z - \overline{y}\|^2$$

$$\Longrightarrow \|z - \overline{y}\|^2 \leq 2\varepsilon \Longrightarrow \|z - \overline{y}\| \leq \sqrt{2\varepsilon}.$$

□

2.3. Study of the Convergence.

Let $\left\{x^k\right\}_k$ be a sequence satisfying:

$$f(x^{k+1}) + r_k h(x^{k+1}) + \frac{1}{2} \left\|x^{k+1} - x^k\right\|^2 \leq f(x) + r_k h(x) + \frac{1}{2} \left\|x - x^k\right\|^2 + \varepsilon^k,$$

where

$$\varepsilon^n \geq 0 \quad \text{and} \quad \sum_n \varepsilon^n < +\infty.$$

Retake the Algorithm 3. We have the following theorem:

Theorem 2. *([1]) Let $\{\varepsilon^k\}$ be a sequence satisfying $\varepsilon^n \geq 0$ and $\sum_n \varepsilon^n < +\infty$. Let $\{r_k\}$ be a sequence such that*

$$r_{k+1} > r_k > 0 \ \text{ and } \lim_k r_k = +\infty.$$

Then the sequence $\{x^k\}_k$ which satisfies

$$f(x^{k+1}) + r_k h(x^{k+1}) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2 \leq f(x) + r_k h(x) + \frac{1}{2}\left\|x - x^k\right\|^2 + \varepsilon^k$$

is bounded.

Proof. First of all $S \neq \emptyset$ (f is inf-compact). Let $\overline{x} \in S$. Put

$$y^k = \arg\min_{y \in \mathbb{R}^n}\left\{\Psi(y, x^k) = f_k(y) + \frac{1}{2}\left\|y - x^k\right\|^2\right\};$$

we remark that

$$\Psi(x^{k+1}, x^k) \leq \min_{y \in \mathbb{R}^n} \Psi(y, x^k) + \varepsilon^k.$$

Then, according to the Lemma 4, we have

$$\left\|x^{k+1} - y^k\right\| \leq \sqrt{2\varepsilon^k}.$$

On the other hand, $y^k = Prox(f_k; x^k)$ and $\overline{x} = Prox(f_k; \overline{x})$. According to the expression (12), we have

$$\left\|y^k - \overline{x}\right\| \leq \left\|x^k - \overline{x}\right\|.$$

Then

$$\|x^{k+1} - \overline{x}\| = \|x^{k+1} - y^k + y^k - \overline{x}\| \leq \|x^{k+1} - y^k\| + \|y^k - \overline{x}\|$$

$$\leq \sqrt{2\varepsilon^k} + \|x^k - \overline{x}\| < M + \sum_k \sqrt{2\varepsilon^k} < +\infty.$$

Consequently $\{x^k\}$ is a bounded sequence. □

In the following theorem, we shall show that every limit point of this sequence is an element of S.

Theorem 3. *Let $\{x^s\}_s$ be a convergent subsequence which converges to $\overline{x}$, then $\overline{x}$ is an optimal solution of $(\mathcal{P})$.*

To show this theorem we have need to the following lemma:

Lemma 4. *Let S_δ^k be the set*

$$S_\delta^k := \{x \in \mathbb{R}^n : \ f_k(x) \leq \alpha + \delta\}.$$

There exists $M_\delta^{\overline{k}} > 0$ ($\overline{k}$ is defined in the Lemma 3) such that

$$< g, x - y > \geq M_\delta^{\overline{k}} \|x - y\| \tag{13}$$

for all $x \notin S_\delta^k$, $y \in S$, $g \in \partial f_k(x)$, $k \geq \overline{k}$.

Proof. Let z be the intersection point of $[x, y]$ with the boundary of $S_\delta^{\overline{k}}$ such that $x \notin S_\delta^k$ and $y \in S$. Then $f_k(z) = \alpha + \delta$. On the other hand,

$$f_k(z) \geq \alpha + \delta \ \ (k \geq \overline{k}) \Longrightarrow f_k(z) - f_k(y) \geq \alpha + \delta - \alpha = \delta.$$

But we know that

$$f'(x, y - x) \leq f(y) - f(x) \ \text{ and } \ f(x) - f(y) \leq f'(x, x - y)$$

(classical result of derivation), then we have

$$f_k(z) - f_k(y) \leq f_k'(z, z - y) \leq f_k'(x, z - y).$$

But

$$z = y + \frac{\|z-y\|}{\|x-y\|}(x-y) \Longrightarrow \delta \leq f_k^{'}(x, z-y) = f_k^{'}(x, \frac{\|z-y\|}{\|x-y\|}(x-y))$$

$$\Longrightarrow f_k^{'}(x, z-y) \geq \frac{\delta}{\|z-y\|} \|x-y\| \Longrightarrow f_k^{'}(x, z-y) \geq M_\delta^k \|x-y\|$$

with

$$M_\delta^k := \inf \left\{ \frac{\delta}{\|z^{'} - y^{'}\|},\ y^{'} \in S,\ z^{'} \in Fr(S_\delta^k) \right\} > 0.$$

Since $S_\delta^k \subset S_\delta^{\overline{k}}$, we have $M_\delta^k \geq M_\delta^{\overline{k}}$, thus

$$f_k^{'}(x, z-y) \geq M_\delta^{\overline{k}} \|x-y\|, \text{ for all } k \geq \overline{k}. \tag{14}$$

If f_k is a differentiable function at x, we have

$$f_k^{'}(x, z-y) = < \nabla f_k(x), z-y > .$$

Else, let $g \in \partial f_k(x)$, then $g = \lim_i \nabla f_k(x_i)$, where $x_i \longrightarrow x$.

Then

$$f_k^{'}(x_i, z-y) = < \nabla f_k(x_i), z-y > \geq M_\delta^{\overline{k}} \|x_i - y\|$$

$$\Longrightarrow \lim_i < \nabla f_k(x_i), z-y > \geq \lim_i M_\delta^{\overline{k}} \|x_i - y\|$$

$$\Longrightarrow < g, z-y > \geq M_\delta^{\overline{k}} \|x-y\|$$

thus wich shows the lemma. □

For the proof of the previous theorem, we consider the following notations inspired of ([1]):

$$r_k(\delta) := \sup_{y \in S_\delta^k} \inf_{x \in S} \|x-y\|; \quad T_\delta^\infty := \left\{ x \in \mathbb{R}^n : f_k(x) \leq \alpha + \frac{1}{2} r_k^2(\delta) \right\};$$

$$A_\delta^k := S_\delta^k \cup T_\delta^k; \ \ q_k(\delta) := \sup_{y \in A_\delta^k} \inf_{x \in S} \|x-y\|; \ \ \varepsilon_{p(k)} := \max_{j>k} \varepsilon_j$$

$$A_\delta^{kp} := A_\delta^k + B(0; \sqrt{2\varepsilon_p}); \ \ q_{kp}(\delta) = \sup_{y \in A_\delta^{kp}} \inf_{x \in S} \|x-y\|;$$

and

$$W_{\delta k} := S + B(0; q_{k\ p(k)}(\delta)).$$

We have

$$\lim_{(k \longrightarrow +\infty, \delta \longrightarrow 0)} q_{k\ p(k)}(\delta) = 0. \tag{15}$$

To show the Theorem 6, we have need also to the following definition:

Definition 3. Let $C \subseteq \mathbb{R}^n$. The relative interior of the set C, denoted $ri(C)$, is defined as follows:

$$ri(C) := \{x \in C : \exists \varepsilon > 0,\ (x + \varepsilon C) \cap aff(C) \subseteq C\} \tag{16}$$

with

$$x + \varepsilon C := \{y \in \mathbb{R}^n : \|x-y\| \leq \varepsilon\}$$

and $aff(C)$ is the affine hull of C.

Proof. **(of Theorem 6)** Denote

$$K := \left\{ s \geq \overline{k} :\ x^s \notin S_\delta^k,\ y^s \notin S_\delta^k \right\}, \quad K^c := \left\{ s \geq \overline{k} : s \notin K \right\}.$$

We shall show that there exists $\overline{s}$ such that $x^s \in K^c$, for all $s \geq \overline{s}$.

Indeed; if $x^s \notin S_\delta^k$, $y^s \notin S_\delta^k$, we have $(x^s - y^s) \in \partial f_k(y^s)$.

Indeed; according to the definition of y^s we have for all y,

$$f_k(y) \geq f_k(y^s) + \tfrac{1}{2}(\|y^s - x^s\|^2 - \|y - x^s\|^2)$$

$$\geq f_k(y^s) + \tfrac{1}{2}(< y^s - x^s, y^s - x^s > - < y - x^s, y - x^s >)$$

$$\geq f_k(y^s) + \tfrac{1}{2}(< y^s - x^s, y^s - x^s > - < y + y^s - y^s - x^s, y - x^s >)$$

$$\geq f_k(y^s) + \tfrac{1}{2}(< y^s - x^s, y^s - y > + < y - y^s, y - x^s >)$$

$$\geq f_k(y^s) + \tfrac{1}{2}(< y^s - x^s, y^s - y > + < y - y^s, y - y^s + y^s - x^s >)$$

$$\geq f_k(y^s) + \tfrac{1}{2}(2 < y^s - x^s, y^s - y > + \left\|y - y^2\right\|^2),$$

which implies

$$f_k(y) \geq f_k(y^s) + < x^s - y^s, y - y^s > .$$

According to the Lemma 7, we have

$$< x^s - y^s, y^s - y > \geq M_\delta^{\overline{k}} \|y^s - y\|, \text{ for all } y \in S,$$

on the other hand, owing to this formula

$$< x^s - y, y^s - y > = < x^s - y^s, y^s - y > + < y^s - y, y^s - y >$$

$$\geq M_\delta^{\overline{k}} \|y^s - y\| + \|y^s - y\|^2 .$$

Thus

$$< x^s - y, y^s - y > \leq \|x^s - y\| \, \|y^s - y\|$$

$$\Longrightarrow \|x^s - y\| \, \|y^s - y\| \geq M_\delta^{\overline{k}} \|y^s - y\| + \|y^s - y\|^2 ,$$

it holds that

$$\|x^s - y\| \geq M_\delta^{\overline{k}} + \|y^s - y\| \tag{17}$$

The Lemma 4 and the formula (17) give

$$\begin{aligned} \left\|x^{s+1} - y\right\| &\leq \left\|x^{s+1} - y^s\right\| + \|y^s - y\| \\ &\leq \sqrt{2\varepsilon_s} + \|x^s - y\| - M_\delta^{\overline{k}} < \sqrt{2\varepsilon_s} + \|x^s - y\| \end{aligned} \tag{18}$$

We remark that if s is too large, the formula (18) does not satisfied; then there exists $\overline{s}$ such that $x^s \in K^c$, for all $s \geq \overline{s}$.

We notice that if $s \geq \overline{s}$, then we have two cases:

$$x^s \notin S_\delta^k \text{ and } y^s \in S_\delta^k \quad \text{or} \quad x^s \in S_\delta^k \text{ and some } y^s.$$

. If $x^s \notin S_\delta^k$ and $y^s \in S_\delta^k$, according to the Lemma 4, we have

$$\left\|x^{s+1} - y^s\right\| \leq \sqrt{2\varepsilon_s} \Longrightarrow x^{s+1} \in S_\delta^k + B(0; \sqrt{2\varepsilon_s})$$

$$\Longrightarrow x^{s+1} \in A_\delta^k + B(0; \sqrt{2\varepsilon_s}) = A_\delta^{kp}$$

then $x^{s+1} \in W_{\delta k}$.

. If $x^s \in S_\delta^k$ and some y^s, we have $x^{s+1} \in W_{\delta k}$.

Indeed;

$$f(y^s) + r_s h(y^s) + \tfrac{1}{2}\|y^s - x^s\|^2 \leq f(x) + \tfrac{1}{2}\|x - x^s\|^2 , \ \forall x \in S$$

$$\Longrightarrow f(y^s) + r_s h(y^s) \leq f(x) + \tfrac{1}{2}\|x - x^s\|^2 , \ \forall x \in S$$

$$\leq \alpha + \tfrac{1}{2} r_s^2(\delta).$$

Hence

$$f_s(y^s) \leq \alpha + \frac{1}{2} r_s^2(\delta) \Longrightarrow y^s \in T_\delta^s .$$

According to the Lemma 4, we have

$$\left\|x^{s+1}-y^{s}\right\| \leq \sqrt{2\varepsilon_s} \Longrightarrow x^{s+1} \in T_\delta^s + B(0;\sqrt{2\varepsilon_s})$$

$$\Longrightarrow x^{s+1} \in A_\delta^s + B(0;\sqrt{2\varepsilon_s}) \Longrightarrow x^{s+1} \in A_\delta^{sp} \Longrightarrow x^{s+1} \in W_{\delta k}.$$

Where from the following conclusion: $x^{s+1} \in W_{\delta k}$, for all $s \geq \overline{s}$, then

$$d(x^s, S) \leq q_{sp(s)}(\delta), \text{ for } s \geq \overline{s}+1.$$

If we make the limit on s when s tends to $+\infty$ and δ tends to 0, we find a result according to the expression (16) of the Definition 3. □

Stopping Test:

We notice, first of all, according to the Theorem 1 that

$$(x^k - x^{k+1}) \in \partial(f + r_k h)(x^{k+1}),$$

then,

$$\xi \in \partial h(x^{k+1}) \Longrightarrow g_{k+1} = ((x^k - x^{k+1}) - r_k \xi) \in \partial f(x^{k+1}).$$

For the stopping test in the Algorithm 3, we propose the following test:

$$\|g_{k+1}\| \left\|x^{k+1} - x^k\right\| \leq \varepsilon^k,$$

where $g_{k+1} \in \partial f(x^{k+1})$.

We have the following proposition:

Proposition 1. *Let $\delta > 0$, if $\|g_{k+1}\| \left\|x^{k+1} - x^k\right\| < \delta$ then $f(x^{k+1}) \leq f(\overline{x}) + \delta$, where $\overline{x}$ is an optimal solution of $(\mathcal{P})$.*

Proof. Admitting that

$$\left\|x^{k+1} - x^k\right\| \simeq \left\|\overline{x} - x^{k+1}\right\|,$$

then, we have

$$f(\overline{x}) \geq f(x^{k+1}) + < g_{k+1}, \overline{x} - x^{k+1} >$$

$$\Longrightarrow f(\overline{x}) \geq f(x^{k+1}) - \|g_{k+1}\| \left\|\overline{x} - x^{k+1}\right\|$$

$$\Longrightarrow f(\overline{x}) \geq f(x^{k+1}) - \|g_{k+1}\| \left\|x^{k+1} - x^k\right\|$$

$$\Longrightarrow f(\overline{x}) \geq f(x^{k+1}) - \delta \Longrightarrow f(x^{k+1}) \leq f(\overline{x}) + \delta.$$

□

2.4. Numerical Experiments.

In this paragraph, we propose some numerical examples illustrating the convex nondifferentiable programming algorithms that we displayed in this work.

Let us remind that the previous algorithms consist in resolving a sequence of unconstrained problems every problem of which must be resolved by the Algorithm 4 below by making the linesearch given by the expression (20) below.

Example 1. *Consider the following problem:*

$$(\mathcal{P}) \qquad \begin{cases} \alpha := Inf \left\{ f(x) = \max_{i=1}^{3} (x^t A_i x + b_i^t x + c_i) \right\} \\ \qquad \textit{subject to} \quad x_1^2 + 3x_2 + 2x_1 \leq 0, \end{cases}$$

where

$$A_1 = \begin{pmatrix} 2 & -1 \\ 1 & 4 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad c_1 = 4;$$

$$A_2 = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad c_2 = -5;$$

$$A_3 = \begin{pmatrix} 2.5 & 2 \\ 0.5 & 2 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 4 \\ -3 \end{pmatrix}, \quad c_3 = 3;$$

x^0 *initial*	k	k *total*	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \lVert g_k \rVert \cdot \lVert x^{k+1} - x^k \rVert$	*time* s
$(2,0)$	4	196	$(-0.391, 0.210)$	3.49	10^4	10^{-4}	$10^{-8}6.0$	$10^{-7}2.0$	0.17
$(4,3)$	6	268	$(-0.460, 0.236)$	3.49	10^6	10^{-6}	10^{-8}	$10^{-2}2.0$	0.22
$(-2,1)$	5	379	$(-0.404, 0.215)$	3.48	10^5	10^{-5}	$10^{-9}2.0$	$10^{-8}7.0$	0.28

Table 1
ε-***Proximal Penalty algorithm:*** $(\delta = 10^{-6})$

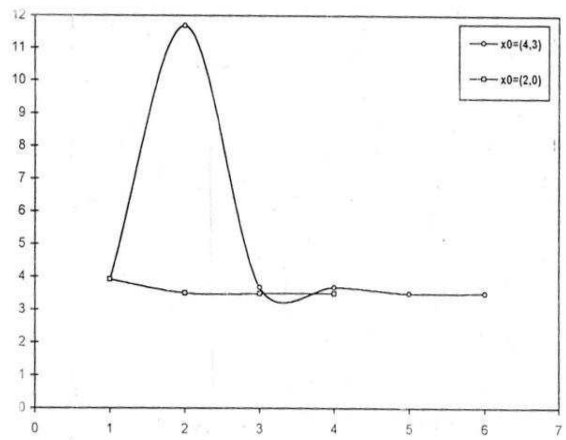

FIGURE 1. The objective function value at each step

Example 2. *Consider the following optimization problem:*

$$(\mathcal{P}) \qquad \begin{cases} \alpha := Inf\,\{f(x) = \max(2x+2,\ (x+1)^2,\ x^2+1)\} \\ \qquad\qquad \textit{subject to} \quad 2x+3 \leq 0. \end{cases}$$

x^0 *initial*	k	k *total*	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \lVert g_k \rVert \, \lVert x^{k+1} - x^k \rVert$	*time* s
5	4	15	-1.5	3.25	10^4	10^{-4}	$10^{-5}5.6$	$10^{-12}5.5$	0.06
62	7	26	-1.5	3.25	10^7	10^{-7}	$10^{-8}5.6$	$10^{-12}5.5$	0.06
-412	4	15	-1.5	3.25	10^4	10^{-4}	$10^{-5}5.6$	$10^{-12}5.5$	0.05

Table 2
ε-***Proximal Penalty algorithm:*** $(\delta = 10^{-11})$

Example 3. *Consider the following optimization problem:*

$$(\mathcal{P}) \qquad \begin{cases} \alpha := Inf\,\{f(x) = \max(f_1(x), f_2(x))\} \\ \qquad \textit{subject to} \ \begin{cases} x_1 + 2x_2 \leq 0 \\ x_2 + 1 \leq 0 \end{cases}, \end{cases}$$

where

$$f_1(x) = x_1^2 + x_2^2 - x_2 - x_1 - 1;\ f_2(x) = 3x_1^2 + 2x_2^2 + 2x_1x_2 - 16x_1 - 14x_2 + 22$$

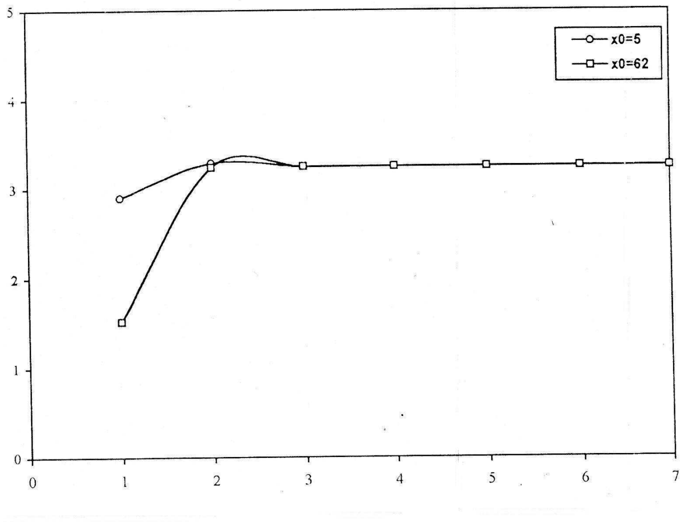

FIGURE 2. The objective function value at each step

x^0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k =$ $\|g_k\| \, \|x^{k+1} - x^k\|$	time s
$(2,0)$	6	20	$(2,-1)$	14	10^6	10^{-6}	$10^{-6}9.0$	$10^{-9}5.0$	0.05
$(-4,3)$	6	24	$(2,-1)$	14	10^6	10^{-6}	$10^{-6}9.0$	$10^{-9}3.0$	0.05
$(6,-7)$	6	24	$(2,-1)$	14	10^6	10^{-6}	$10^{-6}9.0$	$10^{-9}3.0$	0.06

Table 3
ε-***Proximal Penalty algorithm***:$(\delta = 10^{-8})$

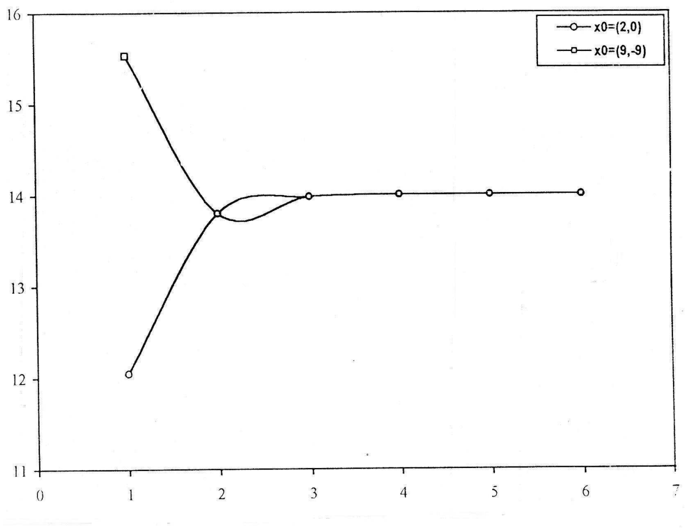

FIGURE 3. The objective function value at each step

Example 4. *Consider the following optimization problem:*

$$(\mathcal{P}) \quad \begin{cases} \alpha := Inf\,\{f(x) = \max(f_1(x), f_2(x), f_3(x))\} \\ \qquad \text{subject to} \begin{cases} x_1 - x_2 + 1 \leq 0 \\ 2x_2 - 1 \leq 0 \end{cases}, \end{cases}$$

where

$$f_1(x) = x_1^2 + x_2^2; \;\; f_2(x) = (x_1 + x_2)^2; \;\; f_3(x) = (2x_1 + 3x_2)^2.$$

x^0 *initial*	k	k *total*	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \lVert g_k \rVert \cdot \lVert x^{k+1} - x^k \rVert$	*time* s
$(3,2)$	6	35	$(-0.5, 0.5)$	0.5	10^6	10^{-6}	$10^{-7}6.7$	$10^{-7}5.7$	0.06
$(5,4)$	6	33	$(-0.5, 0.5)$	0.5	10^6	10^{-6}	$10^{-7}9.1$	$10^{-7}9.8$	0.05
$(-2,-4)$	6	27	$(-0.5, 05)$	0.5	10^6	10^{-6}	10^{-6}	$10^{-6}1.2$	0.05

Table 4
ε-***Proximal Penalty algorithm***: $(\delta = 10^{-5})$

Example 5. *Consider the following optimization problem:*

$$(\mathcal{P}) \qquad \begin{cases} \alpha := Inf \left\{ f(x) = \max_{i=1}^{3} (x^t A_i x + b_i^t x + c_i) \right\} \\ \qquad subject\ to \ \begin{cases} x_1 + x_3 \leq 0 \\ 2x_1 + 1 \leq 0 \end{cases}, \end{cases}$$

where

$$A_1 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad c_1 = 0;$$

$$A_2 = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad c_2 = -2;$$

$$A_3 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad c_3 = 2;$$

x^0 *initial*	k	k *total*	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \lVert g_k \rVert \cdot \lVert x^{k+1} - x^k \rVert$	*time* s
$(1,2,4)$	8	36	$(0,-0.5,0)$	2.25	10^9	10^{-9}	$10^{-11}6.3$	$10^{-10}3.4$	0.11
$(2,8,0)$	3	19	$(0,-0.5,0)$	2.25	10^4	10^{-4}	$10^{-6}6.3$	$10^{-10}1.4$	0.06
$(-2,-1,5)$	9	37	$(0,-0.5,0)$	2.25	10^{10}	10^{-10}	$10^{-10}1.4$	$10^{-13}6.9$	0.11

Table 5
ε-***Proximal Penalty algorithm***: $(\delta = 10^{-9})$

Example 6. *Consider the following optimization problem:*

$$(\mathcal{P}) \qquad \begin{cases} \alpha := Inf \left\{ f(x) = \begin{cases} -x + |x| + e^{|x|} & if\ x \leq 0 \\ x^2 + |x| + e^{|x|} & else \end{cases} \right\} \\ \qquad subject\ to \ \ x + 1 \leq 0. \end{cases}$$

x^0 *initial*	k	k *total*	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \lVert g_k \rVert \lVert x^{k+1} - x^k \rVert$	*time* s
1	11	33	-1	4.718	10^{11}	10^{-11}	$10^{-11}5.6$	$10^{-12}4.3$	0.05
-2	5	14	-1	4.718	10^5	10^{-5}	$10^{-5}5.6$	$10^{-12}4.3$	0.06
-1.5	5	14	-1	4.718	10^5	10^{-5}	$10^{-5}5.6$	$10^{-12}4.3$	0.05

Table 6
ε-***Proximal Penalty algorithm***: $(\delta = 10^{-11})$

2.5. **Comments and Conclusions.**

Basing itself on the results obtained in the previous experiments, we can make the following remarks:

for the ε-Proximal penalty algorithms, we have used the classical penalty functions

$$h(x) = \sum_{i=1}^{m} (g_i(x))^2$$

and the sequence $(r_k)_k$ such that $r_{k+1} = 10r_k$.

Generally, the obtained solutions are enough precise.

The number of iterations depends, on one hand of the algorithm used to resolve the unconstrained subproblems, on the other hand on initial points. This approach possesses the property of the global convergence.

From a theoretical point of view, this approach uses the regularization of Prox. It makes the regularity for the subproblems. So the idea to bring the resolution of primal problem to a sequence of auxiliary problems.

The algorithm which we had used requiet the knowledge at least of a subgradient in every step, and the value of the function to be minimized, then a difficulty concerning the determination of a subgradient which is, generally, difficult in practice.

3. Appendix

For the application of the algorithms of the results obtained previously in the case of nondifferentiable problems, we need to introduce the $BFGS$ method within the framework of nondifferentiable optimization for the resolution of unconstrained subproblems. Then we give an algorithm concerning the line-search.

We saw in the previous algorithms that we need, in every step, to resolve a problem without constraints whose objective function is nondifferentiable. For it we propose the $BFGS$ algorithm. The calculation of a descent direction at x_k, in this algorithm, requests the knowledge of the value of $f(x_k)$ and an arbitrary subgradient $g \in \partial f(x_k)$.

Then we have the necessary steps of this algorithm:

Algorithm 4: ([5])

Step 0:

Let $x^0 \in \mathbb{R}^n$, $k = 0$, $g_0 \in \partial f(x_0)$, $B_0 = Id$, $\delta > 0$.

Compute t_0 by a line-approach along the direction $d_0 = -B_0 g_0$.

Compute

$$x^1 = x^0 + d_0.$$

If $\|g_0\| \left\|x^1 - x^0\right\| \leq \delta$ stop, else go to step 1:

Step 1:

Compute t_k by a line-approach along the direction

$$d_k = -B_k g_k. \tag{19}$$

Compute x^{k+1} by the following iterative process:

$$x^{k+1} = x^k + t_k d_k, \tag{20}$$

where

$$\begin{cases} B_0 = Id \\ B_{k+1} = B_k + (1 + \frac{\gamma_k^t . B_k . \gamma_k}{\delta_k^t . \gamma_k}) \frac{(\delta_k . \delta_k^t)}{(\delta_k^t . \gamma_k)} - \frac{\delta_k . \gamma_k^t . B_k + B_k . \gamma_k . \delta_k^t}{\delta^t . \gamma_k}, \end{cases} \tag{21}$$

with

$$\delta_k = x^{k+1} - x^k, \quad \gamma_k = g_{k+1} - g_k.$$

Step 2:

If

$$\|g_k\| \left\|x^{k+1} - x^k\right\| \leq \delta$$

stop, else $k \longrightarrow k + 1$ and return to the step 1.

3.1. Linesearch.

Before giving the procedure of the linesearch, we need to the following lemmas:

Lemma 5. *([4]) If $\exists t_0 \geq 0$ and $g \in \partial f(x+t_0 d)$ such that $< g, d > \geq 0$, then there exists $\bar{t} \geq 0$ such that*

$$(i)\ \ t > \bar{t} \Longrightarrow < g, d > \geq 0, \ \ \forall g \in \partial f(x+td)$$

$$(ii)\ \ t < \bar{t} \Longrightarrow < g, d > \leq 0, \ \ \forall g \in \partial f(x+td).$$

Moreover, $\bar{t}$ minimizes f in the direction d and

$$\exists \bar{g} \in \partial f(x+\bar{t}d) \ \ \textit{such that} \ \ \ < d, \bar{g} >= 0.$$

Proof. Consider the following convex function $h(t) = f(x+td)$.

Let us notice at first that $h(t) \geq h(0)$ for $t \leq 0$, and $h(0) \geq h(t)$ for $t \geq 0$. Let $t > 0$.

We have

$$\partial h(t) = \{< d, g >: g \in \partial f(x+td)\} \tag{22}$$

For $t \geq t_0$, we have

$$f(x+td) \geq f(x+t_0 d) + (t-t_0) < d, g > \geq f(x+t_0 d) \Longrightarrow h(t) \geq h(t_0).$$

The function $h(t)$ attains its minimum at a point $\bar{t} \in [0, t_0]$. Then, $\forall t < \bar{t}$, $\forall g \in \partial f(x+td)$, we have

$$h(t) \geq h(\bar{t}) \geq h(t) + (\bar{t}-t) < d, g >,$$

hence

$$< d, g > \leq 0.$$

On the other hand, $\forall t > \bar{t}$, $\forall g \in \partial f(x+td)$, we have

$$h(t) \geq h(\bar{t}) \geq h(t) + (\bar{t}-t) < d, g >,$$

thus

$$< d, g > \geq 0.$$

Finally, because $\bar{t}$ minimize $h(t)$ whithout constraints (because $h(\bar{t}) \leq h(0) \leq h(t)$ for $t < 0$), then $0 \in \partial h(\bar{t})$, hence

$$\exists \bar{g} \in \partial f(x+\bar{t}d) \ \ \text{such that} \ \ \ < d, \bar{g} >= 0.$$

□

To make our linesearch, we are going to be inspired by this lemma. Let us call $\bar{t}$ the optimal step in the direction of d.

We are going to try hard to frame $\bar{t}$ by two values t_1 and t_2, such that:

(a) at t_1 we shall have a subgradient $g_1 \in \partial_\varepsilon f(x)$;

(b) at t_2 we shall have a subgradient $g_2 \in \partial_\varepsilon f(x+t_2 d)$ such that

$$< g_2, d > \geq -m \|d\|^2,$$

where $0 \leq m \leq 1$.

Calculating a convex combination

$$g = \lambda g_1 + (1-\lambda) g_2, \quad 0 \leq \lambda \leq 1$$

suitably chosen, we can arrive at the following conclusion:

$$g \in \partial_\varepsilon f(x) \ \ \text{and} \ \ \ < g, d > \geq -m \|d\|^2. \tag{23}$$

Remark 5. 1) The ε-subgradient g in the expression (20) is a subgradient nowhere, but it has no importance.

2) The properties

$$g \in \partial_\varepsilon f(x) \ \ \text{and} \ \ \ < g, d > \geq -m \|d\|^2$$

mean that:

on one hand

$$g \in \partial_\varepsilon f(x) \ \Longrightarrow f(x+td) \geq f(x) + t < g, d > -\varepsilon \Longrightarrow < g, d > \leq \frac{\varepsilon}{t},$$

thus t is big enough (Lemma 9);

on the other hand,

$$< g, d > \geq -m \|d\|^2$$

implies t is enough small (Lemma 9). In this case we can take t as an optimal step.

Let us envisage now the most current case where $\overline{t} \in]0, +\infty[$. Suppose

$$g_1 \in \partial f(x + t_1 d)$$

such that

$$f(x + t_1 d) \geq f(x) - \varepsilon$$

and

$$g_2 \in \partial f(x + t_2 d)$$

such that

$$f(x + t_2 d) \geq f(x) - \varepsilon.$$

We have the following lemma:

Lemma 6. *In order that*

$$g = \lambda g_1 + (1 - \lambda) g_2, \quad 0 \leq \lambda \leq 1$$

is an element of $\partial_\varepsilon f(x)$, *it is enough that*

$$\lambda t_1 < g_1, d > + (1 - \lambda) t_2 < g_2, d > \leq 0 \tag{24}$$

Proof. We have

$$f(y) \geq f(x + t_1 d) + < g_1, y - x - t_1 d >, \quad \forall y$$

$$f(y) \geq f(x + t_2 d) + < g_2, y - x - t_2 d >, \quad \forall y.$$

By convex combination we obtain

$$\forall y \in \mathbb{R}^n, \quad f(y) \geq \lambda f(x + t_1 d) + (1 - \lambda) f(x + t_2 d) + \lambda < g_1, y - x - t_1 d >$$

$$+ (1 - \lambda) < g_2, y - x - t_2 d > .$$

Since

$$f(x + t_i d) \geq f(x) - \varepsilon \quad (i = 1, 2),$$

this implies

$$\forall y \in \mathbb{R}^n, \quad f(y) \geq f(x) - \varepsilon + < \lambda g_1 + (1 - \lambda) g_2, y - x >$$

$$-(\lambda t_1 < g_1, d > + (1 - \lambda) t_2 < g_2, d >).$$

According to the expression (21), we have

$$f(y) \geq f(x) + < \lambda g_1 + (1 - \lambda) g_2, y - x > -\varepsilon,$$

hence the lemma is shown. □

Consequently the linesearch can be stopped with

$$g = \lambda g_1 + (1 - \lambda) g_2$$

as soon as $\lambda \in [0, 1]$ satisfy simultaneously

$$\begin{cases} < \lambda g_1 + (1 - \lambda) g_2, d > \geq -m \|d\|^2 \\ \\ \lambda t_1 < g_1, d > + (1 - \lambda) t_2 < g_2, d > \leq 0 \end{cases} \tag{25}$$

Let us synthetize the previous results. We are going to execute a linesearch in two phases. The first one consists in extrapolating t_1 so as to find one t_2 if it exists there. The second phase will be a sequence of interpolation between t_1 and t_2.

Algorithm 5: ([4])

1st Phase:

(a) Let $t_1 \geq 0$ and $g_1 \in \partial f(x + t_1 d)$ such that $< g_1, d > \leq -m \|d\|^2$.

(b) Let $t > t_1$ be given. Compute $f(x+td)$ and $g \in \partial f(x+td)$. We test

$$f(x+td) < f(x) - \varepsilon.$$

If yes, end (we take t as an optimal step).
Else, we go to the following step:
(c) Test

$$< g, d >< -m \|d\|^2 .$$

If yes, t is enough small. We shall put $t_1 = t$, $g_1 = g$; we lock up at (b).
Else, we go to the following step:
(d) Test

$$f(x+td) \geq f(x) + t < g, d > -\varepsilon.$$

If yes, stop, we set t_1 as an optimal step.
Else,
(e) set $t_2 = t$ and $g_2 = g$ and go to the second phase.
2^d Phase:
$(a^{'})$ We test if there exists λ satisfying simultaneously:

$$\begin{cases} 0 \leq \lambda \leq 1 \\ \\ \lambda < g_1, d > +(1-\lambda) < g_2, d > \geq -m \|d\|^2 \\ \\ \lambda t_1 < g_1, d > +(1-\lambda) t_2 < g_2, d > \leq 0. \end{cases}$$

If yes, end and we have

$$g = \lambda g_1 + (1-\lambda) g_2$$

where

$$< g, d > \geq -m \|d\|^2 \quad \text{and} \quad g \in \partial_\varepsilon f(x).$$

We take

$$t = \lambda t_1 + (1-\lambda) t_2.$$

Else, we go to the next step: $(b^{'})$ We interpolate between t_1 and t_2: let $t \in]t_1, t_2[$. We compute $f(x+td)$ and $g \in \partial f(x+td)$. We test

$$f(x+td) < f(x) - \varepsilon.$$

If yes, stop (we take t as an optimal step). Else, we go to the next step: $(c^{'})$ Compute $< g, d >$. If

$$< g, d > \leq -m \|d\|^2 ,$$

t is enough small: we proclaim $t_1 = t$ and $g_1 = g$ and we lock up at $(a^{'})$. Else, we go to the next step: $(d^{'})$ If $< g, d > \geq 0$, t is too large: we proclaim $t_2 = t$ and $g_2 = g$ and we lock up at $(a^{'})$. Else: $(e^{'})$ end, we take t as an optimal step. We summarize this paragraph by giving a possible organigram for the linesearch. Generally, we choose $m = 0.1$ and $t_1 = 0$.

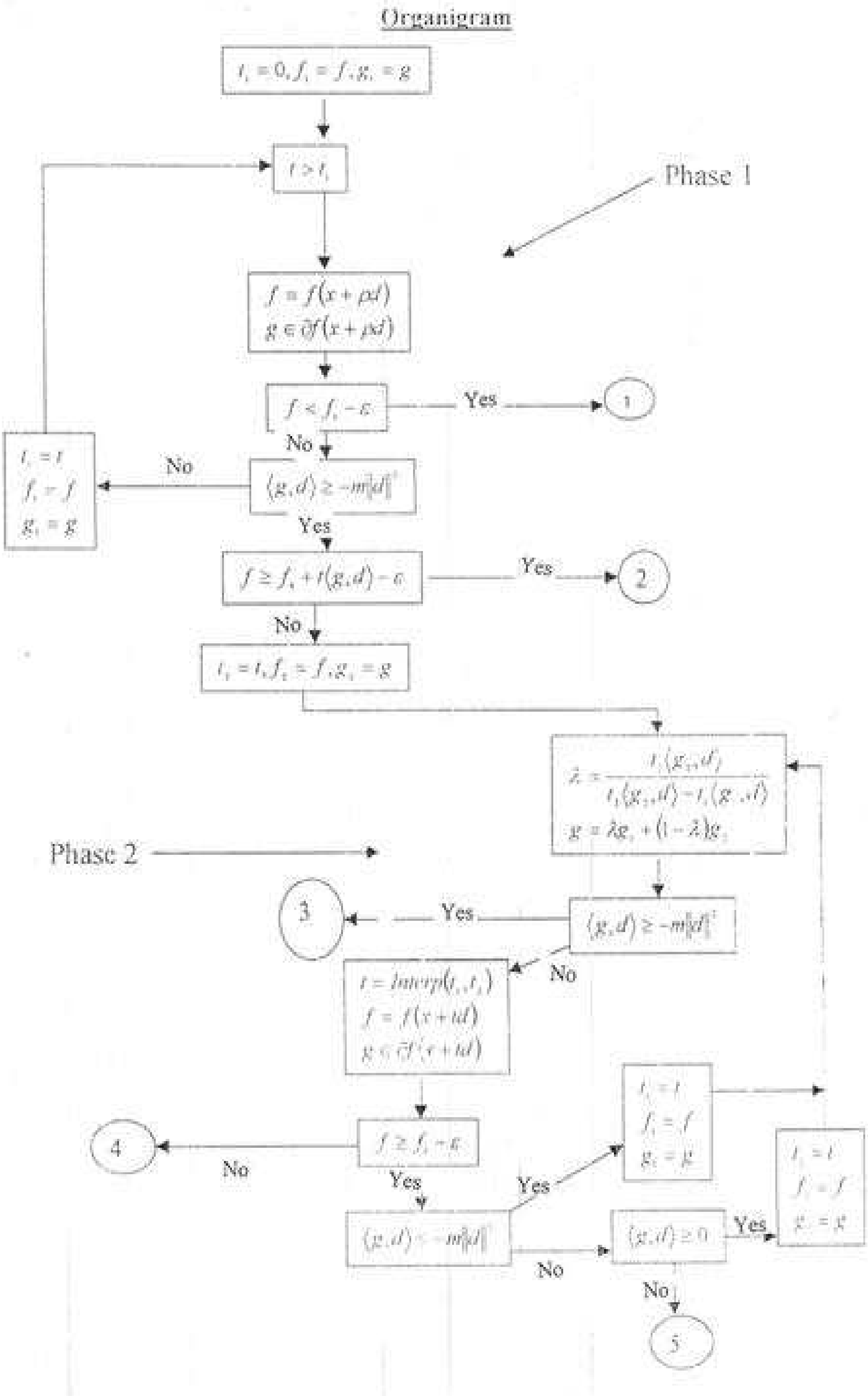

FIGURE 4.

3.2. **Comments.**

In the previous algorithm, we notice that:

- if (a) is not verified we go to (2);
- at (d) we have

$$g \in \partial_\varepsilon f(x) \quad \text{and} \quad < g, d > \geq -m \|d\|^2 ;$$

- in the second phase, we have

$$g_1 \in \partial f(x + t_1 d) \ \text{ and } \ g_2 \in \partial f(x + t_2 d)$$

$$< g_1, d > < -m \|d\|^2 \quad \text{and} \quad < g_2, d > \geq -m \|d\|^2$$

and also

$$f(x + t_1 d) \geq f(x) - \varepsilon \ \text{ and } \ f(x + t_2 d) \geq f(x) - \varepsilon,$$

thus, according to the Lemma 10, we can find λ satisfying $(a^{'})$. Then, if the value λ, as in the organigram, satisfies $(a^{'})$ we stop, otherwise we go to $(b^{'})$;

- at $(e^{'})$, we have $-m \|d\|^2 << g, d >< 0$, then, it is finished because

$$< g, d > \geq -m \|d\|^2$$

and that

$$\left. \begin{array}{l} f(x + td) \geq f(x) - \varepsilon \\ \\ 0 \geq t < g, d > \end{array} \right\} \Longrightarrow g \in \partial_\varepsilon f(x).$$

In the previous organigram:

- if (1) take place then we are in (b);
- if (2) take place then we are in (d);
- if (3) take place then we are in $(a^{'})$;
- if (4) take place then we are in $(e^{'})$.

4. Acknowledgements

Authors would like to thank anonymous referees for their very careful reading of the manuscript and their valuable suggestions.

References

[1] A. Auslender, J.P. Crouzeix and P. Fedit, Penalty-Proximal in convex programming, J.O.T.A., Vol. 55 (1987), pp. 1-21.

[2] A. Auslender, On the differential properties of the support function of the ε-subdifferential of a convex function, Math. Prog. 24 (1982), pp. 257-268.

[3] J.Ch. Culioli, Introduction à l'optimisation, Ellipse édition, France, 1994.

[4] C. Lemarechal, Extensions diverses des méthodes de gradient et applications, Thèse de Doctorat, Pub. Université de Paris IX, 1980.

[5] C. Lemarechal, Numerical experiments, In Nonsmooth Optimization, Pub. I.N.R.I.A., France, (1979), pp. 61-84.

[6] J.J. Moreau, Proximité et dualité dans un espace hilbertien, Bull. Soc. Math. France, (BSMF) 93(1965), pp. 273-299.

[7] R.T. Rockafellar, Convex Analysis, Princeton Univ. Press. Princeton, 2010.

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