AOM - Advanced Modeling and Optimization,Volume 17, Number 2, 2015

DUAL BOUNDS IN CONVEX AND NONCONVEX NONDIFFERENTIABLE OPTIMIZATION PROBLEMS AND APPLICATIONS

NOUREDDINE DAILI AND MERZAKA KHALDI
DEPARTMENT OF MATHEMATICS. CITÉ DES 300 LOTS, YAHIAOUI. 51 RUE HARRAG SENOUSSI, 19000 SÉTIF - ALGERIA

NOURDAILI_DZ@YAHOO.FR

ABSTRACT. In our work, the essential purpose of this paper was to answer the following question : Can we spread the used techniques in the differentiable case ([6]) for the nondifferentiable case by keeping the convergence as well as its rank ? This seems clear in our study. Indeed, always basing on the augmented Lagrangian method, we solved our problem by application of the proximal quasi-Newton technique.

Large linear and nonlinear systems of saddle point type arise in a wide variety of applications throughout computational science and engineering. Due to their indefiniteness and often poor spectral properties. The main purpose of this work is the study of the nondifferentiable problems of optimization and the determination of the dual bounds associated with these problems by exploiting the augmented Lagrangian method. We are interested in the study of a nondifferentiable problem of optimization in both cases: convex and nonconvex. For the method of resolution, we applied the augmented Lagrangian method. This technique has a lot of advantage for the regularization of the solution and numerical stability of the penalty method. We introduce the technique of the augmented Lagrangian. So, we obtain a sequence of unconstrained nondifferentiable subproblems the resolution of which is based, essentially, on a new method which is the proximal quasi-Newton method. The last section of this work implements this technique and handles numerical implementations.

Key words : Nondifferentiable optimization, Dual bounds, Algorithms
AMS Subject : (2010) Primary 90C25, 90C26; Secondary 49N15, 90C33, 90C90.

1. INTRODUCTION

The differentiable and nondifferentiable, convex and nonconvex optimization made the object of several studies. Let us quote as an example the works of Bazaraa-AL ([1]), Bertsekas ([2]), Fletcher ([15]), Dem'Yanov and Vasil'Ev ([13]) and others.

The nondifferentiable optimization is interested in the resolution of the problems of optimization when we lose the differentiability for the objective functions and the constraints, or for f only, either for the constraints.

The nondifferentiability of the objective function or constraints engenders problems of nondifferentiable optimization such as the problems of economic origin. Where hence, the major importance of these problems. Thus, the study of these problems became essential. Wolfe ([23]) was among the first ones in this domain. He gave an implementable descent method. Lemaréchal and Mifflin ([17]) proposed several techniques of descent based, essentially, on the subgradient.

The purpose of our work is the study of the dual bounds in nondifferentiable, convex and nonconvex optimization problems by exploiting the augmented Lagrangian method.

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

We saw ([7], [8]) that the calculation of the dual bound brought in the maximization of a concave function h. For the same reasons quoted previously, we still introduce the augmented Lagrangian method which remains valid in the nondifferentiable case ([9]; [10], [11]). The resolution of the dual problem for the augmented Lagrangian comes down, in this context, to the minimization of a nondifferentiable problem of unconstrained mathematical programming problems. This latter is a very active domain of search ([4], [13], [16], [17] and [18]). Then we transform this problem to the case of the resolution of a problem via a proximal quasi-Newton method ([4]). In fact, the principle of this method is a combination between the technique of proximal point algorithm of Rockafellar ([20]) and the quasi-Newton method. We were able to obtain a problem similar to the one that we studied in ([7]) in favor to the technique of the proximal point of which we used the cutting plane method. Results of convergence of these techniques are established. Then an implementation and numerical implementations are produced.

2. MAIN RESULTS

The main purpose of this work is the study of the nondifferentiable convex problems and the determination of the dual bounds associated to these problems.

More exactly, we study in the first section of this work the geometric interpretation of the optimality conditions. We give to it, having to raise our problem, necessary and sufficient conditions.

The second section is intended to the study of a nondifferentiable method of optimization and the calculation of the dual bounds by this technique.

In reality, we know well that in the nondifferentiable optimization, the use of the strongest slope direction obtained by the subgradient of the function which we want to minimize does not lead, inevitably, to the convergence towards the optimal point (Wolfe, [23]).

The implementable descent algorithms for the general nondifferentiable problems were presented by Wolfe ([23]), Lemaréchal and Mifflin ([17]).

First, we introduce the technique of augmented Lagrangian (Daili, [10], [11]). So, we obtain a sequence of nondifferentiable unconstrained sub-problems, the resolution of which is based, essentially, on a new method which is the proximal quasi-Newton method.

Indeed, this method is proposed by Chen and Fukushima ([4]) ; it is based on the proximal point algorithm of Rockafellar ([20]) and the cutting plane techniques. The latter establishes the most important part of this algorithm.

We give the results of convergence for each of the methods : proximal quasi–Newton and augmented Lagrangian.

2.1. Dual Bounds of Convex Nondifferentiable Optimization Problems.

Indeed, we can characterize an optimal point by a generalization of the Fermat condition of the differentiable case,

$$\bigtriangledown f(\bar{x}) = 0.$$

We use the subgradient which plays a similar role to that of the gradient.

2.1.1. *A Geometric Interpretation of Optimality Conditions*.

Consider the following optimization problem :

$$(\mathcal{P}) \qquad \begin{cases} \alpha := Inf f(x) \\ \text{subject to } x \in C, \end{cases} \tag{1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a convex function and $C \neq \varnothing$ is a convex set. Let

$$\gamma(x_0) = \{v = \lambda(z - x_0) : \lambda > 0, \ z \in C\}.$$

The set $\Gamma(x_0) = \overline{\gamma}(x_0)$ is called the cone of admissible directions of the set C to the point x_0.

We shall denote the conjugate cone of $\Gamma(x_0)$ by $\Gamma^+(x_0)$. We always have

$$0 \in \gamma(x_0) \cap \Gamma^+(x_0).$$

Theorem 1. *The function f reaches its optimal value on C at the point $\bar{x} \in C$, if and only if*

$$\partial f(\bar{x}) \cap \Gamma^+(\bar{x}) \neq \emptyset. \tag{2}$$

The subdifferential of the function f on the set C at the point x_0 is given by the following lemma :

Lemma 2. *If the function f is finite and convex on $\mathbb{R}^n$, then*

$$\partial^C f(x_0) = \partial f(x_0) - \Gamma^+(x_0).$$

According to this lemma the problem (1) can spell under the shape $0 \in \partial^C f(\bar{x})$.

Corollary 3. *If $C = \mathbb{R}^n$, the expression (2) is equivalent to the following condition : $0 \in \partial f(\bar{x})$.*

Corollary 4. *If $x_0 \in int(C)$ and f is finite and convex on C, then*

$$\partial^C f(x_0) = \partial f(x_0).$$

Lemma 5. *So that the expression (2) is satisfied, it is necessary that*

$$0 \in L_\eta(\bar{x}) \tag{3}$$

where

$$L_\eta(x) = co(\partial f(x)) \cup T_\eta(x)$$

and

$$T_\eta(x) = \left\{g \in -\Gamma^+(x) \ : \| \ g \ \| = \eta\right\}.$$

Besides, this condition is sufficient if $int(C) \neq \emptyset$.

Proof. **Necessity** : We have

$$\partial f(\bar{x}) \cap \Gamma^+(\bar{x}) \neq \emptyset,$$

then

$$\exists \ v \in \partial f(\bar{x}), \ \exists w \ \in \Gamma^+(\bar{x}) : \ v = w.$$

If $w = 0$, then $v = 0$ and thus $0 \in L_\eta(\bar{x})$.

If $w \neq 0$, then

$$g = -\eta w \parallel w \parallel^{-1} \in T\eta(\bar{x}).$$

As $v - w = 0$, then we have $\eta v \parallel w \parallel^{-1} + g = 0$.

For $\alpha = \eta \parallel w \parallel^{-1} (1 + \eta \parallel w \parallel^{-1})^{-1}$, we find

$$v_\alpha = \alpha v + (1 - \alpha) g = 0.$$

As $\alpha \in [0, 1]$, $v \in \partial f(\bar{x})$ and $g \in T_\eta(\bar{x})$, then $v_\alpha \in L_\eta(\bar{x})$ and thus $0 \in L_\eta(\bar{x})$.

Sufficiency : Suppose $T_\eta(\bar{x}) = \emptyset$, then $0 \in \partial f(\bar{x})$ and since $0 \in \Gamma^+(\bar{x})$, thus

$$\partial f(\bar{x}) \cap \Gamma^+(\bar{x}) \neq \emptyset.$$

Now, if $0 \in T_\eta(\bar{x})$, the condition $0 \in L_\eta(\bar{x})$ implies the existence of two vectors $v \in \partial f(\bar{x})$ and $g \in T_\eta(\bar{x})$ and a number $\alpha \in [0, 1]$ such that $\alpha v + (1-\alpha) g = 0$ and as $int(C) \neq \emptyset$, then $0 \notin co(T_\eta(\bar{x}))$. What implies $\alpha > 0$, and then $v = (\alpha - 1)\alpha^{-1} g$. We have

$$(\alpha - 1)\alpha^{-1} g = v \in \Gamma^+(\bar{x})$$

because $(\alpha - 1)\alpha^{-1} < 0$, thus the condition (2) is satisfied. $\square$

2.1.2. ***Necessary and Sufficient Optimality Conditions***.

Consider the following convex mathematical programming problem :

$$(\mathcal{PP}) \qquad \begin{cases} \alpha := Inf f(x) \\ \text{subject to} \quad \begin{cases} f_i(x) \leq 0, & 1 \leq i \leq p, \\ g_j(x) = 0, & 1 \leq j \leq q, \quad p, q \in \mathbb{N}^*, \\ x \in \mathbb{R}^n, \end{cases} \end{cases}$$

where $f, \;\; f_i, \;\; 1 \leq i \leq p$ and $g_j, \;\; 1 \leq j \leq q \; : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ are convex functions.

The Lagrangian of $(\mathcal{PP})$ is the function $L : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q \longrightarrow \overline{\mathbb{R}}$ defined by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{p} \lambda_i f_i(x) + \sum_{j=1}^{q} \mu_j g_j(x).$$

We associate to it the following dual problem :

$$(\mathcal{DP}) \qquad \Big\{ \beta := \underset{(\lambda,\,\mu)}{Sup}\; \underset{x}{Inf} L(x, \lambda, \mu).$$

Theorem 6. *Suppose the functions $f, \; f_i, \;\; 1 \leq i \leq p$, and $g_j, \;\; 1 \leq j \leq q$, are convex and there exists $x_0 \in \mathbb{R}^n$ such that $f_i(x_0) < 0, \;\; \forall i \; (1 \leq i \leq p)$. Let $(\bar{x}, \bar{\lambda}, \overline{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q$, then $(\bar{x}, \bar{\lambda}, \overline{\mu})$ is a saddle point for $L(x, \lambda, \mu)$, if, and only if, $(\bar{x}, \bar{\lambda}, \overline{\mu})$ satisfies*

(**a**) $0 \in (\partial f(x) + \sum\limits_{i=1}^{p} \lambda_i \partial f_i(x) + \sum\limits_{j=1}^{q} \mu_j \partial g_j(x))$;
(**b**) $f_i(\bar{x}) \leq 0, \quad 1 \leq i \leq p$;
(**c**) $g_j(\bar{x}) = 0, \quad 1 \leq j \leq q$;
(**d**) $\bar{\lambda}_i f_i(\bar{x}) = 0, \;\; \textit{for} \;\; 1 \leq i \leq p$.

Proof. According to ([6], Theorem 4.9, p. 141), $(\bar{x}, \bar{\lambda}, \overline{\mu})$ is a saddle point of $L(x, \lambda, \mu)$ if, and only if,

($\mathbf{a}_1$) $L(\bar{x}, \bar{\lambda}, \overline{\mu}) = \underset{x \in \mathbb{R}^n}{Inf} L(x, \bar{\lambda}, \overline{\mu})$;
($\mathbf{b}_1$) $f_i(\bar{x}) \leq 0, \quad 1 \leq i \leq p$;
($\mathbf{c}_1$) $g_j(\bar{x}) = 0, \quad 1 \leq j \leq q$;
($\mathbf{d}_1$) $\bar{\lambda}_i f_i(\bar{x}) = 0$, for $1 \leq i \leq p$.

To establish the proof of the Theorem 6, we notice that it is enough to show the equivalence ($\mathbf{a}$) $\Leftrightarrow$ ($\mathbf{a}_1$).

Using the Corollary 3, L reaches its infimum at $\overline{x}$ if, and only if, 0 is a subgradient of L at $\overline{x}$, that is $0 \in \partial L(\bar{x})$. As

$$\bigcap_{i=0}^{p} S_i = S \neq \emptyset,$$

where

$$S_i = \{x \in \mathbb{R}^n : \ f_i(x) \leq 0, \quad 1 \leq i \leq p\}.$$

We have

$$\partial L(\bar{x}) = \partial f(\bar{x}) + \sum_{i=1}^{p} \lambda_i \partial f_i(\bar{x}) + \sum_{j=1}^{q} \mu_j \partial g_j(\bar{x}).$$

Thus $(\mathbf{a}) \Leftrightarrow (\mathbf{a}_1)$. □

3. METHOD OF RESOLUTION

3.1. Motivation.

Once we characterized an optimum $\overline{x}$, we are interested in the problem of calculation of this optimum. As main rule and in a concern of coherence, the nondifferentiable methods are clearly connected to the algorithms developed for differentiable functions. We shall give an algorithm which finds a direction d_k and a step α_k of linear search to update the current iteration x_k.

However, the nondifferentiable optimization problem presents some difficulties that we summarize it :

1) The lack of an implementable stop test, because the condition

$$g_k \in \partial f(x_k) \ \text{ where } \ \| g_k \| \leq \varepsilon$$

translated directly of $\| \nabla f(\, x_k \,) \| \leq \varepsilon$, can be never verified ;

2) the calculation of the approached subgradients : if f is nondifferentiable at x^*, the classical methods (differentiable case) are not valid. Sometimes in the practice the gradient is not exactly calculated. It is often obtained by finite differences of the values of the function f ;

3) the curse of the nondifferentiability : we already know that the multi-application : $x \longmapsto \partial f(x)$ not being continuous, a small variation on x_k can give big variations at $\partial f(x_k)$, the calculation of d_k can give x_{k+1} very different.

As in our work ([7]), we also use the augmented Lagrangian method. Indeed, this technique remains valid in the nondifferentiable case.

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a convex function. Let g_i, $i = 1, ..., q : \mathbb{R}^n \to \mathbb{R}$ be convex functions and

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

Consider the following mathematical programming problem :

$$\alpha := Inf \{f(x), \ x \in S\}.$$

The augmented Lagrangian associated is

$$L_{r_k}(x, \lambda) = f(x) + \frac{1}{2r_k} \sum_{i=1}^{q} (\Psi^+((\lambda_i + 2r_k g_i(x))^2 - \lambda_i^2),$$

where $\Psi^{+}(t) = \max(t, 0)$, namely

$$L_{r_k}(x,\lambda) = f(x) + \begin{cases} \sum_{i=1}^{q} \lambda_i g_i(x) + r_k \sum_{i=1}^{q} g_i^2(x), & \text{if } \lambda_i + 2r_k g_i(x) > 0, \\ \\ -\frac{1}{4r_k} \sum_{i=1}^{q} \lambda_i^2, & \text{if } \lambda_i + 2r_k g_i(x) \leq 0. \end{cases}$$

Being given $\lambda_k \geq 0$ and $r_k > 0$ to determine x_k which minimizes $L_{r_k}(x, \lambda_k)$.

We generate the sequence $\{\lambda_k\}$ by resting

$$\lambda_{k+1}^{i} = \Psi^{+}(\lambda_k^{i} + 2r_k g_i\,(x_k)) \geq 0, \quad \text{for} \quad i = 1, ..., q,$$

where

$$\lambda_{k+1} = \lambda_k + 2r_k \nabla_\lambda L(x_k,\, \lambda_k).$$

For the resolution of the unconstrained nondifferentiable sub-problems

$$\underset{x \in \mathbb{R}^n}{Inf}\, L_{r_k}(x, \lambda_k)$$

we have choose the proximal quasi-Newton method whose principle we present in what follows :

3.1.1. ***Proximal quasi-Newton Method***.

In this section, we propose the proximal quasi-Newton method to minimize a convex nondifferentiable function in $\mathbb{R}^n$. This method is based on the Rockafellar's ([20]) proximal point algorithm and the cutting plane method. In every step, we approxime the proximal point by $p^{\alpha}(x_k)$ at the point x_k, to define one $v_k \in \partial_{\varepsilon_k} f(x_k)$ with $\varepsilon_k \leq \alpha \parallel v_k \parallel$, α constant. The quasi-Newton step is used to reduce the value of $\parallel v_k \parallel$.

Without the differentiability of f, the method converges globally.

a) The Concept of Proximal Point :

An idea to introduce the quasi-Newton method into the nondifferentiable optimization problem is to consider the proximal point as it is developed in Rockafellar ([20]).

Consider the following mathematical programming problem :

$$\alpha := \underset{x \in \mathbb{R}^n}{Inf}\, f(x), \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex non-necessarily differentiable function.

In particular we assure that f is finite and continuous on $\mathbb{R}^n$.

This problem can be to transform into a convex differentiable minimization problem

$$\overline{\alpha} := \underset{x \in \mathbb{R}^n}{Inf}\, F(x), \tag{5}$$

where

$$F(x) = \underset{y \in \mathbb{R}^n}{Inf} \left\{ f(y) + \frac{\mu}{2} \parallel y - x \parallel^2 \right\} \tag{6}$$

and μ is a positive real number. ($F(x)$ is called the Moreau-Yosida regularized of f)

Proposition 7. *The function F_μ defined by*

$$F_\mu \ : \ \ y \to F_\mu(y) = f(y) + \frac{\mu}{2}\ \|y - x\|^2$$

reaches its infimum.

Proof. The function

$$F_\mu \; : \;\; y \to F_\mu(y) = f(y) + \frac{\mu}{2} \; \|y - x\|^2$$

is continuous. Let us show that it is coercive on $\mathbb{R}^n$. As f is convex by hypothesis, it possesses an affine minorant, namely there exist $s_0 \in \mathbb{R}^n$ and $\alpha_0 \in \mathbb{R}$ such that

$$f(y) \geq < s_0,\, y > + \alpha_0, \quad\;\; \forall\; y \in \mathbb{R}^n$$

Consequently,

$$F_\mu(y) \geq < s_0,\, y > + \alpha_0 + \tfrac{\mu}{2} \|y - x\|^2$$

$$\geq \tfrac{\mu}{2} \left\| y - (\tfrac{s_0}{\mu} - x) \right\|^2 - \tfrac{\mu}{2} \left\| \tfrac{s_0}{\mu} - x \right\|^2 \; + \tfrac{\mu}{2} \|x\|^2 + \alpha_0,$$

thus, we deduct that

$$\lim_{\|y\| \to +\infty} F_\mu(y) = +\infty.$$

As a result there is well a point minimizing F_μ on $\mathbb{R}^n$. □

Proposition 8. *The infimum of F_μ is reached in an only point of $\mathbb{R}^n$.*

Proof. The function F_μ being strictly convex, as sum of a convex function and a strictly convex function, there is only a point minimizing F_μ on $\mathbb{R}^n$. □

Definition 1. *The infimum of F_μ in (6) is called the proximal point of x. We shall note it by $p(x)$ in all which follows.*

Proposition 9. *The function F is derivable in all $x \in \mathbb{R}^n$ and its derivative is given by*

$$\nabla F(x) = \mu(x - p(x)) \tag{7}$$

$$\mu(x - p(x)) \in \partial f(p(x)). \tag{8}$$

Proof. Let $N := \frac{\mu}{2} \|.\|^2$; then F is the inf-convolution of f and N :

$$F = (f \square N).$$

Furthermore, this inf-convolution is exact in every point of $\mathbb{R}^n$.

$$\forall x \in \mathbb{R}^n,\;\; F(x) = (f \square N)(x) = f(p(x)) + \frac{\mu}{2} \|x - p(x)\|^2 .$$

We use, then, the calculation rule giving the subdifferential of an inf-convolution of the function F to obtain

$$\forall x \in \mathbb{R}^n,\;\; \partial F(x) = \partial f(p(x)) \cap \partial N(x - p(x)),$$

or

$$\partial N(x - p(x)) = \{\mu(x - p(x))\}\,,$$

then

$$\partial F(x) = \{\mu(x - p(x))\} \;\;\text{ and }\; \{\mu(x - p(x))\} \in \partial f(p(x)).$$

The function F is convex from $\mathbb{R}^n$ to $\mathbb{R}$, thus locally Lipschitz on $\mathbb{R}^n$. Then the only element $\mu(x - p(x))$ of $\partial F(x)$ define a differential (in the Fré chet sens) of F at x as follows :

$$G(x) = \nabla F(x) = \mu(x - p(x)).$$

□

Lemma 10. *If f is a lower bounded function on $\mathbb{R}^n$, then, the function F is lower bounded on $\mathbb{R}^n$.*

Proof. As

$$F(x) \geq f(y) + \tfrac{\mu}{2}\left\|x - y\right\|^2$$

$$\geq f(p(x)) + \tfrac{\mu}{2}\left\|x - p(x)\right\|^2 \geq f(p(x)) \geq \underset{x\in\mathbb{R}^n}{Inf}\, f(x),$$

$F(x)$ is lower bounded on $\mathbb{R}^n$ as soon as f is it. □

Lemma 11. *The problems (4) and (5) are equivalent in the sense that all the sets of solutions of the problems (4) and (5) simultaneous.*

Proof. The conjugate N^* of N is

$$s \in \mathbb{R}^n \longrightarrow N^*(s) = \frac{1}{2\mu}\left\|s\right\|^2.$$

Because $F = f \square N$, we have

$$F^* = f^* + N^*.$$

Namely,

$$\forall s \in \mathbb{R}^n,\ F^*(s) = f^*(s) + N^*(s) = f^*(s) + \frac{1}{2\mu}\left\|s\right\|^2.$$

In particular, $F^*(0) = f^*(0)$. As

$$F^*(0) = -\underset{x\in\mathbb{R}^n}{Inf}\, F(x) \ \text{ and } \ f^*(0) = -\underset{x\in\mathbb{R}^n}{Inf}\, f(x),$$

then, the set of solutions of the problem (4) coincides with the set of solutions of the problem (5). □

Theorem 12. *Let f be a convex function with finite value at least in a point. Then*

$(\mathbf{a})$ *for all $\mu > 0$ we have*

$$f(p(x)) \leq F(x) \leq f(x),\ \ \forall x \in \mathbb{R}^n\,;$$

$(\mathbf{b})$ *the following properties are equivalent :*
$(\mathbf{p}_1)$ *x minimize f on $\mathbb{R}^n$;*
$(\mathbf{p}_2)$ *x minimize F on $\mathbb{R}^n$;*
$(\mathbf{p}_3)$ $x = p(x)$;
$(\mathbf{p}_4)$ $f(x) = f(p(x))$;
$(\mathbf{p}_5)$ $f(x) = F(x)$.

Proof. $(\mathbf{a})$ By definition we have

$$f(p(x)) \leq f(p(x)) + \frac{\mu}{2}\left\|x - p(x)\right\|^2 = F(x) \leq f(x), \quad \forall x \in \mathbb{R}^n.$$

$(\mathbf{b})$ We prove

$(\mathbf{p}_1) \Rightarrow (\mathbf{p}_2)$: (immediately) by the previous lemma.

$(\mathbf{p}_2) \Rightarrow (\mathbf{p}_3)$: We know that F is convex and differentiable, with

$$\nabla F(x) = \mu(x - p(x)) \quad \forall\, x,$$

consequently x minimize F on $\mathbb{R}^n$ if, and only if, $\nabla F(x) = 0$. Thus

$$\mu(x - p(x)) = 0,$$

then $p(x) = x$.

$(\mathbf{p}_3) \Rightarrow (\mathbf{p}_4) \Rightarrow (\mathbf{p}_5)$: immediately from $(\mathbf{a})$.
$(\mathbf{p}_5) \Rightarrow (\mathbf{p}_1)$: We have

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

It means that the lower bounded in the definition of $F(x)$ is reached at x , thus

$$x = p(x) \Rightarrow x - p(x) = 0 \Rightarrow 0 \in \partial f(x) \Rightarrow x \text{ minimize } f.$$

□

The proximal Newton type method was the object of search for several actors what gave good results ([4], [16] and [18]).

To solve the problem (4), we resort to the quasi-Newton method a shape of which is the following one :

$$x_{k+1} = x_k - \alpha_k B_k^{-1} G(x_k), \tag{9}$$

where α_k is the step of displacement and B_k is generated by the quasi-Newton formula. We use

$$s_k = x_{k+1} - x_k, \qquad y_k = G(x_{k+1}) - G(x_k). \tag{10}$$

For example, with Broyden-Fletcher-Goldfarb-Shanno $(BFGS)$ update formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^t B_k}{s_k^t B_k s_k} + \frac{y_k y_k^t}{y_k^t s_k} \tag{11}$$

the global convergence of the $BFGS$ method with a linear search and the convergence order are already studied in ([7]).

In the general case, it is impossible to obtain the exact proximal point $p(x_k)$ for x_k given. We need to approximate $p(x_k)$ by an approximative method of resolution of the problem (6) only.

b) Algorithm and Global Convergence :

In this algorithm, we look for an approximation of proximal point $p^\alpha(x_k)$ of x_k. The used technique is the cutting plane method such that

$$\mu(x_k - p^\alpha(x_k)) \in \partial_{\varepsilon_k} f(p^\alpha(x_k))$$

for $\varepsilon_k \in [0, \alpha \|v_k\|]$, where α is a constant. $\partial_\varepsilon f(x)$ is the subdifferential of f for any positive ε,

$$\partial_\varepsilon f(x) := \left\{ x \in \mathbb{R}^n \ : f(y) \geq f(x) + g^t(y-x) - \varepsilon, \ y \in \mathbb{R}^n \right\}.$$

Then, we put

$$G^\alpha(x_k) = v_k$$

the approximant of $G(x)$. We use this formula in the quasi-Newton method (10), then we build B_{k+1} by considering

$$y_k = G^\alpha(x_{k+1}) - G^\alpha(x_k)$$

in the formulae (11) and (12). The displacement step α_k is determined from the value of $\|G^\alpha(x_k)\|$ either by a linear search on the function f.

This algorithm works with double iterations, each serves to solve the following approximate problem :

$$\underset{y\in\mathbb{R}^n}{Inf} f(y) + \frac{1}{2}\mu \|y - x_k\|^2 . \tag{12}$$

We use the cutting plane method technique, the point x_k is the current iteration given to the iteration k.

More exactly, the internal iteration generates a sequence $\{y_j\}$ defined as follows :

for $\quad j = 1, 2, ..., \quad y_j$ is the only solution of following problem :

$$\underset{y \in \mathbb{R}^n}{Inf} f_{k,j}(y) + \frac{1}{2}\mu \|y - x_k\|^2 , \tag{13}$$

where $f_{k,j}$ is the polyhedral convex function defined by

$$f_{k,j} = \max_{i=0}^{j-1} \left\{ f(y_i) + g_i^t(y - y_i) \right\} \tag{14}$$

it is an approximation of f in the neighborhood of x_k and

$$g_i \in \partial f(y_i), \; i = 0, 1, ..., j - 1.$$

So there, y_j is the solution of the sub-problem in it j-th iteration.

The internal iteration stops if one of the following conditions is satisfied :

$$f(x_k) - f_{k,\,j}(y_j) < \rho \tag{15}$$

$$f(y_j) \leq f(x_k) - \sigma_k(f(x_k) - f_{k,\,j}(y_j)), \tag{16}$$

where $\rho > 0$ and $\sigma_k \in [0, 1]$ are parameters.

Remark 1. *The procedures* (13), (14) *form the pure cutting plane algorithm.*

The relation (14) of $f_{k,\,j}$ and the subgradient inequality imply :

$$f_{k,j}(y) \leq f(y), \quad \forall \; y \in \mathbb{R}^n. \tag{17}$$

We have the inequalities :

$$\begin{aligned} f_{k,j}(y_j) &\leq f_{k,j}(y_j) + \tfrac{1}{2}\mu \|y_j - x_k\|^2 = \underset{y \in \mathbb{R}^n}{Inf} \left\{ f_{k,j}(y) + \tfrac{1}{2}\mu \|y - x_k\|^2 \right\} \\ &\leq \underset{y \in \mathbb{R}^n}{Inf} \left\{ f(y) + \tfrac{1}{2}\mu \|y - x_k\|^2 \right\} \leq f(x_k), \quad \forall \; j = 1, 2, .. \end{aligned} \tag{18}$$

Consequently, if (16) is satisfied, from (18), it results that

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

This inequality means that, if ρ is enough small, then y_j is a good approximtion of the solution of the problem (12).

Consider the condition (16). We can show ([16], Proposition 3) that, when j increases, $f(y_j)$ and $f_{k,\,j}(y_j)$ approach one of the other one and y_j converges to the proximal point $p(x_k)$.

When (16) is satisfied, we put

$$p^{\alpha}(x_k) := y_j$$

to use it as an approximation of the proximal point of x_k.

If $\sigma_k \longrightarrow 1$, then $p^{\alpha}(x_k)$ is the best approximation of the proximal point $p(x_k)$.

Algorithm :

Step 0 : put $k = 0$

Choose an initial point $x_0 \in \mathbb{R}^n$ and parameters $\rho > 0$, σ, c, $\gamma \in [0, 1]$, a constant M enough large such that $M \geq f(x_0)$ and a sequence $\{\sigma_k\}_k$ such that $\sigma < \sigma_k < 1$.

Step 1 : solve the sub-problem (13) by the procedures (14) and (15) to obtain y_j there satisfying (16) or (17).

. If (16) is satisfied, then stop, end.

. If (17) is satisfied, then one put

$$j_k = j \quad \text{and} \quad p^{\alpha}(x_k) = y_j$$

and go to the step **2**.

Step 2 : Put

$$v_k = \mu(x_k - p^{\alpha}(x_k)).$$

Step 3 : Let $B_0 = (1+\mu)I$, for $k = 0$.

If $k \geq 1$, we build a definite positive and symmetric matrix $B_k \in \mathbb{R}^{n\times n}$, we use the quasi-Newton formula with

$$s_k = x_k - x_{k-1} \text{ and } \quad y_k = v_k - v_{k-1}$$

Step 4 : Calculate $d_k = -(B_k - \mu^{-1}I)v_k$.

. If $k = 0$ and $\eta_1 = \|v_0\|$, go to step **5**.

. If $k \geq 1$, $\|v_k\| \leq c\eta_k$ and $f(p^{\alpha}(x_k) + d_k) \leq M$, we put $\alpha_k = 1$ and go to step **6**.

. Else put $\eta = \eta_k$ and go to step **5**.

Step 5 : Let m_k be the smallest positive number of m such that

$$f(p^{\alpha}(x) + \gamma^m d_k) \leq f(x_k) - \frac{\gamma^m \sigma}{\mu} \|v_k\|^2 .$$

Put $\alpha_k = \gamma_{m_k}$.

Step 6 : Calculate $x_{k+1} = x_k + \alpha_k d_k$, $\quad k = k+1$ and return to the step **1**.

To assure that each of the steps of this algorithm is well defined, we need to establish some propositions. In what follows, we indicate by f_k the polyhedral function $f_{k,\,j}$ for which the condition of the formula (17) is satisfied. So $p^{\alpha}(x_k)$ satisfies the inequality

$$f(p^{\alpha}(x_k)) \leq f(x_k) - \sigma_k(f(x_k) - f_k(p^{\alpha}(x_k))). \tag{19}$$

The following proposition shows that the step **1** is always finitely feasible.

Proposition 13. *(*[4]*) In every iteration k of the algorithm, the step **1** is executed in a finite number of steps, and gives a sequence y_j which verifies* (16) *and* (17) *for all $j = 1, 2, ...$*

The following propositions give the error, from the approximation of the function f_k and the approximation of the proximal point $p^{\alpha}(x_k)$.

Proposition 14. *(*[4]*)For all $k \geq 0$, put*

$$F^{\alpha}(x_k) = f(p^{\alpha}(x_k)) + \tfrac{\mu}{2} \|p^{\alpha}(x_k) - x_k\|$$

$$G^{\alpha}(x_k) = v_k$$

$$\varepsilon_k = (1 - \sigma_k)(f(x_k) - f_k(p^{\alpha}(x_k))).$$

Then,

$$F(x_k) \leq F^{\alpha}(x_k) \leq F(x_k) + \varepsilon_k \tag{20}$$

$$\|p^{\alpha}(x_k) - p(x_k)\| \leq \sqrt{\frac{2\varepsilon_k}{\mu}} \tag{21}$$

$$\|G^{\alpha}(x_k) - G(x_k)\| \leq \sqrt{2\mu\varepsilon_k} \tag{22}$$

$$v_k \in \partial_{\varepsilon_k} f(p^{\alpha}(x_k)) \tag{23}$$

Moreover, if the sequences $\{x_k\}$ *and* $\{p^{\alpha}(x_k)\}$ *are bounded, then*

$$\varepsilon_k \leq \beta \|v_k\|, \tag{24}$$

where

$$\beta = (1-\sigma)\frac{L}{\sigma\mu}$$

and L *is a Lipschitz constant of* f.

Proposition 15. *For all* k*, there exists* $\overline{\tau}_k > 0$ *such that*

$$f(p^{\alpha}(x) + \tau d_k) \leq f(x_k) - \frac{\tau\sigma}{\mu}\|v_k\|^2, \quad \forall \tau \in [o, \overline{\tau}_k]. \tag{25}$$

We are going to show here that, under reasonable hypotheses, the proximal quasi-Newton method converges.

Theorem 16. ***(Convergence)*** *Let* f *be a convex function such that the bounded set of minimizers of* f *is non empty and* $\{\|B_k\|\}_k$ *is a bounded sequence, then the previous algorithm converges in a finite number of iterations.*

Proof. Let us indicate $K = \{1, 2, 3, \ ...\}$ and

$$K_0 = \{0\} \cup \{k \in K \ : \text{excluding step } \mathbf{5}\}.$$

The hypothesis f has a bounded set of minimizers implies the functions f and F have a bounded level set. By construction, for every k, x_k and $p^{\alpha}(x_k)$ remain in the bounded set

$$D = \{x : f(x) \leq M\}.$$

Let $\overline{f}$ be the minimum value of f, suppose that (14) is not satisfied, that is

$$f(x_k) - f_k(p^{\alpha}(x_k)) > \rho, \qquad \forall k. \tag{26}$$

According to the Proposition 13, the previous algorithm generates two sequences $\{x_k\}$ and $\{p^{\alpha}(x_k)\}$. We prove that every cluster point of $\{x_k\}$ and $\{p^{\alpha}(x_k)\}$ minimizes f.

Case 1 : Consider the case where K_0 is infinite. Let us put

$$K_0 = \{k_0 = 0 \leq k_1 \leq k_2 \leq ...\}.$$

By construction, we have

$$\|v_{k_l}\| \leq c\eta_{k_l} = c\left\| v_{k_{l-1}} \right\|, \quad \forall l = 0, 1, 2, ...$$

and as $v_k = G^{\alpha}(x_k)$, then

$$\|v_k\| \leq c^l \|G^{\alpha}(x_0)\|, \quad \forall\, l = 0, 1, 2, ...$$

Thus, we find

$$\lim_{(l \longrightarrow +\infty)} \|G^{\alpha}(x_{k_l})\| \leq \lim_{(l \longrightarrow +\infty)} c^l \|G^{\alpha}(x_0)\| = 0 \tag{27}$$

Let L be a Lipschitz constant of f. From the formula (25), of the Proposition 15, we have

$$\varepsilon_{k_l} \leq \beta \|G^{\alpha}(x_k)\|,$$

where

$$\beta = \frac{(1-\sigma)\, L}{\sigma\mu}.$$

By the formula (23) of the Proposition 15, we have then

$$\|G(x_{k_l})\| \leq \|G^{\alpha}(x_{k_l})\| + \sqrt{2\mu\varepsilon_{k_l}} \leq \|G^{\alpha}(x_{k_l})\| + \sqrt{2\mu\alpha\,\|G^{\alpha}(x_{k_l})\|} \tag{28}$$

It results from the formula (28) that

$$\lim_{(l\longrightarrow+\infty)} \|G(x_{k_l})\| = 0.$$

Because for $x_k,\ p^{\alpha}(x_k) \in D$ for all $l \geq 0$; there is an infinite subset $K' \subset K_0$ such that

$$\lim_{(k\longrightarrow\infty,\, k\in K')} x_k = \overline{x} \quad \text{and} \quad \lim_{(k\longrightarrow\infty,\, k\in K')} p^{\alpha}(x_k) = \overline{p}\,.$$

Thus, from

$$\mu\,\|\overline{x} - \overline{p}\| \leq \lim_{(k\longrightarrow\infty,\, k\in K')} \|G^{\alpha}(x_k)\| = 0$$

and the formula (28), we have $\overline{x} = \overline{p}$ and $G\,(\overline{x}) = 0$. Then the cluster point $\overline{x}$ of the subsequence $\{x_k\}_{k\in K'}$ is an optimal solution.

Let us show, now, that every cluster point of the sequence $\{x_k\}$ is an optimal solution of f.

Indeed; if there is a $\overline{k}$ such that $k \geq \overline{k}$ for all $k \in K_0$, then from

$$\|G^{\alpha}(x_{k+1})\| \leq c\,\|G^{\alpha}(x_k)\|\,,\ \ k \geq \overline{k},$$

it holds that every cluster point of $\{x_k\}$ is an optimal point.

On the other hand, if $\overline{k}$ do not exist, then by construction, for every $k \in K_0$ there exists a sufficiently large integer $k_l \in K_0$ such that $k_l \leq k$ and

$$\overline{f} \leq f(x_{k+1}) \leq f(x_{k_l} + 1) = f(x_{k_l} - B_k v_{k_l}).$$

Because $\{\|\, B_k\|\}_k$ is bounded, $\|\, v_k\| \longrightarrow 0$ and $f(x_k) \longrightarrow \overline{f}$ then every cluster point of $\{x_k\}_k$ is an optimal solution of f.

Case 2 : Consider the case where K_0 is finite, let $\hat{k} = \max\limits_{k\in K_0} k$.

As the level sets of f are bounded, the sequences $\{x_k\}_k$ and $\{p^{\alpha}(x_k)\}$ are bounded by construction, then

$$\frac{\sigma}{\mu}\sum_{k=\hat{k}+1}^{\infty} \tau_k\,\|v_k\|^2 \leq f(x_{\hat{k}+1}) - \overline{f} < \infty \tag{29}$$

which implies

$$\lim_{(k\longrightarrow\infty)} \alpha_k\,\|v_k\|^2 = 0. \tag{30}$$

Because

$$\sigma < \sigma_k \ \text{ and } \ \mu\,\|x_k - p^{\alpha}(x_k)\|^2 \leq f(x_k) - f_k(p^{\alpha}(x_k)).$$

We have, according to the Proposition 15,

$$\overline{f} \leq f(p^{\alpha}(x_k)) \leq f(x_k) - \frac{\sigma}{\mu}\,\|v_k\|^2\,. \tag{31}$$

Moreover, as $\{x_k\}_k$ and $\{p^{\alpha}(x_k)\}_k$ are bounded, there exists a subsequence $K_1 \subset K$ such that

$$\lim_{(k\longrightarrow\infty,\ k\in K_1)} p^{\alpha}(x_k) = \overline{p}\ ,\ \lim_{(k\longrightarrow\infty,\ k\in K_1)} x_k = \overline{x}\ ,\ \lim_{(k\longrightarrow\infty,\ k\in K_1)} v_k = \overline{v}.$$

If $\lim\limits_{(k\longrightarrow\infty,\ k\in K_1)} \alpha_k > 0$, then the formula (34) implies $\overline{v} = 0$. On the other hand, if $\lim\limits_{(k\longrightarrow\infty,\ k\in K_1)} \alpha_k = 0$, then the definition of m_k in step of linear search gives

$$f(p^{\alpha}(x_k) + \gamma^{m_{k-1}} d_k) > f(x_k) - \frac{\gamma^{m_k - 1}\sigma}{\mu} \|v_k\|^2 . \tag{32}$$

As

$$\gamma^{m_{k-1}} = \frac{\alpha_k}{\gamma} \longrightarrow 0,$$

the formula (31) implies

$$f(\overline{p}) \geq f(\overline{x}). \tag{33}$$

From the formula (32), we will have

$$f(\overline{p}) \leq f(\overline{x}) - \frac{\sigma}{\mu} \|\overline{v}\|^2 . \tag{34}$$

Both formulae (34) and (35) implie $\overline{v} = 0$ and $\overline{x} = \overline{p}$. So, we have

$$\overline{\varepsilon} = \lim(1-\sigma_k)(f(x_k) - f_k(p^{\alpha}(x_k))) \leq \frac{1-\sigma}{\sigma} \lim_{(k\longrightarrow\infty,\ k\in K_1)} (f(x_k) - f_k(p^{\alpha}(x_k))) = 0,$$

what implies $\overline{\varepsilon} = 0$.

We have $v_k \in \partial_{\varepsilon_k} f(p^{\alpha}(x_k))$, the formula of ε-subdifferential implies $0 \in \partial f(\overline{p})$. Then, $\overline{x} = \overline{p}$ minimize f and $f(\overline{x}) = \overline{f}$.

Knowing that

$$\overline{f} \leq f(x_{k+1}) \leq f(x_k) - \frac{\alpha_k \sigma}{\mu} \|v_k\|^2 , \quad \forall\, k > \hat{k}\, . \tag{35}$$

$\{f(x_k)\}_{k>\hat{k}}$ is a decreasing sequence thus has a limit f^*. In passing in the limit in (35) with $k \in K_1$, we obtain $f^* = \overline{f}$. Then, every cluster point of $\{x_k\}_k$ stays an optimal point. Consequently, if the formula (27) is verified for all k, it follows that every cluster point of $\{x_k\}_k$ and $\{p^{\alpha}(x_k)\}$ minimize f and

$$f(p^{\alpha}(x_k)) - f(x_k) \leq \sigma_k(f(x_k)) - f_k(p^{\alpha}(x_k))) \leq -\sigma(f(x_k) - f_k(p^{\alpha}(x_k))),\ \forall\, k.$$

However, this returns us to a contradiction with (2) because

$$f(p^{\alpha}(x_k)) - f(x_k) \longrightarrow 0.$$

Thus, the algorithm converges in a finite number of iterations. □

We have following both theorems :

Theorem 17. *Let us suppose that the asymptotic optimal value for the problem ($\mathcal{PP}$) is finite. Let $(\lambda_k)_k$ be a bounded sequence which maximizes the problem $(P^*_{r_k})$, $(r_k > 0)$, let $x_k \in \mathbb{R}^n$ be satisfying*

$$L_{r_k}(x_k, \lambda_k) - Inf L_{r_k}(., \lambda_k) = L_{r_k}(x_k, \lambda_k) - d_{r_k}(\lambda_k) \leq \varepsilon_k,$$

where $\varepsilon_k \longrightarrow 0$, as $(k \longrightarrow +\infty)$. Then $(x_k)_k$ is an asymptotically minimizing sequence of ($\mathcal{PP}$).

Theorem 18. *Let us suppose that the problem $(\mathcal{PP})$ admits a strict feasible solution (the Slater condition is verified) and that the optimal value of $(\mathcal{PP})$ is finite. Let $(\lambda_k)_k$ be a sequence solution of the problem $(P^*_{r_k})$, $(r_k > 0)$, and let $x_k \in \mathbb{R}^n$ and ε_k be satisfying*

$$L_{r_k}(x_k, \lambda_k) - Inf L_{r_k}(., \lambda_k) = L_{r_k}(x_k, \lambda_k) - d_{r_k}(\lambda_k) \leq \varepsilon_k,$$

where $\varepsilon_k \longrightarrow 0$, as $(k \longrightarrow +\infty)$. Then

$(\mathbf{a})$ $(x_k)_k$ *is asymptotic minimizing ;*
$(\mathbf{b})$ *there exists a sequence* $(\overline{x}_k)_k$ *of feasible solutions of the problem* $(\mathcal{PP})$ *such that*

$$\lim_{(k\longrightarrow+\infty)}(\overline{x}_k - x_k) = 0 \quad \textit{and} \quad \lim_{(k\longrightarrow+\infty)} f(\overline{x}_k) = \lim_{(k\longrightarrow+\infty)} f(x_k) = \alpha,$$

where α *is at the same time an optimal value and asymptotically optimal of* $(\mathcal{PP})$*;*
$(\mathbf{c})$ *furthermore, the sequence* $(\lambda_k)_k$ *is bounded and all cluster values are Kuhn-Tucker (K.T) vectors for the problem* $(\mathcal{PP})$.

Proof. For the proof of these theorems, we can see ([10]; [11]). □

4. Main Algorithm and Study of the Convergence

The purpose of this section is to show that sequences $\{x_k\}_k$ and $\{\lambda_k\}_k$ given by the algorithm above, converge globally for one fixed positive r_k and with at least the Slater condition is satisfied.

Algorithm :

Step 0 : $(k = 0)$ Choose a factor $r_0 > 0,$ a multiplier $\lambda_0 \in \mathbb{R}_+^m$ and a sequence $(\varepsilon_k)_k$ with

$$\varepsilon_k \geq 0 \ , \quad \lim_{(k\longrightarrow+\infty)} \varepsilon_k = 0.$$

Step 1 : Give $r_k > 0,\ \lambda_k \geq 0$, find one solution x_k of the problem

$$Inf\,\{L_{r_k}(x, \lambda_k),\ x \in \mathbb{R}^n\}$$

such that

$$L_{r_k}(x_k, \lambda_k) - Inf\,\{L_{r_k}(x, \lambda_k),\ x \in \mathbb{R}^n\} \leq \varepsilon_k$$

using proximal quasi-Newton method.

Step 2 : Put

$$\lambda_{k+1}^i = \max(\lambda_k^i + 2r_k g_i(x_k),\ 0), \qquad i = 1, ..., q,$$

or

$$\lambda_{k+1} = \lambda_k + 2r_k \nabla_\lambda L_{r_k}(x_k,\ \lambda_k)$$

Step 3 : If

$$\|\nabla L_{r_k}(x_k,\ \lambda_k)\| \leq \delta$$

stop and take x_k as solution of $(\mathcal{PP})$.
Choose $r_{k+1} \geq r_k$ (if necessary), and return to the step **1**.
The convergence result is given by the following theorem :

Theorem 19. *(*[10]*;* [7]*) Let us suppose the* $(\mathcal{PP})$ *problem has a Kuhn-Tucker (K.T.) vector and that*

$$\sum_{k\geq 1} \varepsilon_k < +\infty.$$

Then we have the following properties :
$(\mathbf{p}_1)$: $(\lambda_k)_k$ *converges towards to an K.T. vector ;*
$(\mathbf{p}_2)$: *the sequence* $\{x_k\}_k$ *satisfies*

$$\lim_{(k\longrightarrow+\infty)} \sup g_i(x_k) \leq 0\,,\ \textit{for}\ \ i = 1, ..., q\ ;\ \lim_{(k\longrightarrow+\infty)} f(x_k) = \alpha.$$

5. COMPUTATIONAL RESULTS

Example 1. $(n = 2;\ m = 2)$

Consider the following mathematical programming problem :

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

where

$$\begin{cases} f_1(x_1, x_2) = x_2^3 + x_1^2 + 1\ ; \\ f_2(x_1, x_2) = (2 - x_2)^2 - x_1. \end{cases}$$

We have

λ_0	x_k	*f(opt)*	*iterations*	r_k	ε_k	*time(s)*
$(2;2)^t$	$(-0.9076; 0.7413)^t$	2.35557133	30	1.5	10^{-3}	$0.11s$
$(2;2)^t$	$(-0.8216; 0.8045)^t$	2.22491671	16	5.5	10^{-3}	$0.05s$
$(1;3)^t$	$(-0.8039; 0.8138)^t$	2.21093811	33	1.5	10^{-3}	$0.06s$
$(1;3)^t$	$(-0.9579; 0.7854)^t$	2.43303965	20	10	10^{-3}	$0.05s$
$(1.5;2.5)^t$	$(-0.7780; 0.8216)^t$	2.19451953	61	5.5	10^{-3}	$0.05s$
$(1.5;2.5)^t$	$(-0.7784; 0.8227)^t$	2.19445110	74	5.5	10^{-4}	$0.06s$
$(1.5;2.5)^t$	$(-0.7790; 0.8226)^t$	2.19444251	89	5.5	10^{-5}	$0.11s$
$(4;6)^t$	$(-0.7258; 0.8342)^t$	2.17547419	57	5.5	10^{-4}	$0.05s$
$(10;1)^t$	$(-0.7666; 0.8344)^t$	2.16525885	41	5.5	10^{-4}	$0.06s$

$$\underline{\delta = 10^{-8}}$$

Example 2. $(n = 2;\ m = 2)$

Consider the following mathematical promming problem :

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

where

$$\begin{cases} f_1(x_1, x_2) = x_2^2 + x_1 x_2\,; \\ f_2(x_1, x_2) = (4 - x_1)^2 + (1 - x_2)^2 + 5\,; \\ f_3(x_1, x_2) = \exp(-x_1 + x_2). \end{cases}$$

We have

λ_0	x_k	$f(opt)$	*iterations*	r_k	ε_k	*time(s)*
$(5;4)^t$	$(1.1250;1.8750)^t$	21.53253229	10	1.5	10^{-5}	$0.05s$
$(5;4)^t$	$(1.0208;-1.9792)^t$	22.75086814	13	11.5	10^{-4}	$0.05s$
$(5;4)^t$	$(1.0208;-1.9792)^t$	22.75086814	14	11.5	10^{-6}	$0.05s$
$(5;4)^t$	$(1.0122;-1.9878)^t$	22.85395638	12	20	10^{-5}	$0.05s$
$(24;1)^t$	$(0.6547;-2.1304)^t$	24.76234123	186	11.5	10^{-4}	$0.11s$
$(9;1)^t$	$(0.9375;-2.0625)^t$	23.75812712	14	11.5	10^{-4}	$0.06s$
$(0;0)^t$	$(1.0732;-1.9569)^t$	22.63629699	17	20	10^{-5}	$0.05s$
$(5.5;7)^t$	$(1.0061;-1.9939)^t$	22.92690486	12	20	10^{-5}	$0.05s$
$(6;80)^t$	$(1.0000;-2.0000)^t$	23.00001645	7	20	10^{-5}	$0.00s$

$$\underline{\delta = 10^{-6}}$$

Example 3. $(n = 3;\ m = 3;\ \delta = 10^{-8})$

Consider the following mathematical programming problem :

$$(\mathcal{PP}) \quad \begin{cases} \alpha := Inf f(x) = Inf \max\limits_{i=1}^{3}(\frac{1}{2}x^t A_i x + b_i^t x \ + c_i) \\ \\ \text{subject to} \begin{cases} x_1^2 - 1 \leq 0, \\ \\ x_2^2 - 1 \leq 0, \\ \\ x_3^2 - 1 \leq 0, \end{cases} \end{cases}$$

where

$$A_1 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad c_1 = 2,$$

$$A_2 = \begin{pmatrix} 5 & 3 & 1 \\ 2 & 10 & 1 \\ 1 & 0 & 20 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad c_2 = 0,$$

$$A_3 = \begin{pmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 2 & 0.5 \\ 0.5 & 0.5 & 2 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad c_2 = 0.$$

We have

λ_0	x_k	f(opt)	iterations	r_k	ε_k	time(s)
$(1;1;1)^t$	$(-0.5987;0.1998;0.0008)^t$	1.70000180	48	1.5	10^{-4}	$0.05s$
$(1;4;7)^t$	$(-0.4457;0.0628;0.0019)^t$	1.73089727	44	0.5	10^{-4}	$0.06s$
$(1;4;7)^t$	$(-0.6000;0.1.999;0.0002)^t$	1.70000033	57	1.5	10^{-4}	$0.11s$
$(1;4;7)^t$	$(-0.5860;0.2073;0.0028)^t$	1.70038087	50	5	10^{-4}	$0.06s$
$(0.5;1;2)^t$	$(-0.6000;0.2000;0.0001)^t$	1.70000000	49	1.5	10^{-4}	$0.05s$
$(0.5;1;2)^t$	$(-0.5762;0.2138;0.0008)^t$	1.70118323	52	5.5	10^{-5}	$0.06s$
$(0.5;1;2)^t$	$(-0.6000;0.2000;0.0001)^t$	1.70000000	59	5.5	10^{-6}	$0.11s$
$(12;12;1)^t$	$(-0.2138;0.0712;0.0006)^t$	1.82426573	99	5.5	10^{-4}	$0.11s$
$(2;0;0)^t$	$(-0.5998;0.1998;0.0002)^t$	1.70000006	71	10.5	10^{-4}	$0.11s$

Example 4. $(n = 2;\ m = 2)$

Consider the following mathematical programming problem :

$$(\mathcal{PP}) \qquad \begin{cases} \alpha := \inf f(x) = Inf \exp(x_1 x_2) + |x_1| \\ \\ \textit{subject to} \quad \begin{cases} x_1 + x_2 + 1 \leq 0, \\ \\ x_2^2 - 1 \leq 0. \end{cases} \end{cases}$$

We have

λ_0	x_k	$f(opt)$	*iterations*	ε_k	r_k	*time(s)*
$(1;1)^t$	$(-0.0029;-0.9921)^t$	1.00577409	11	10^{-2}	1.5	$0.05s$
$(1;1)^t$	$(-0.0122;-08877)^t$	1.02305248	9	10^{-5}	2.5	$0.05s$
$(1;1)^t$	$(-0.0022;-0.7916)^t$	1.00386107	15	10^{-5}	1.5	$0.06s$
$(9;0)^t$	$(0;-1)^t$	1.00001757	6	10^{-5}	1.5	$0.05s$
$(9;0)^t$	$(-0.0008;-1.0000)$	1.00151714	5	10^{-4}	10	$0.00s$
$(1.6;2)^t$	$(-0.0023;-0.5006)^t$	1.00347098	13	10^{-5}	1.5	$0.05s$
$(4;4)^t$	$(-0.0121;-0.3612)^t$	1.01649251	14	10^{-5}	2.5	$0.06s$
$(4;4)^t$	$(-0.0103;-0.8906)^t$	1.01948055	8	10^{-5}	10.5	$0.05s$
$(4;4)^t$	$(-0.0037;-0.9887)$	1.00746318	6	10^{-5}	100	$0.05s$

$$\underline{\delta = 10^{-8}}$$

6. Comments and conclusions

In this work, we were interested in the calculation of the dual bounds of nondifferentiable optimization problems. A very important remark to indicate : the extreme difficulty of the calculation of sub-gradient.

Then, on one hand, we introduce the proximal quasi-Newton method to surmount this inconvenience. On the other hand, to spread the quasi–Newton method in the nondifferentiable case.

The obtained results of the numerical tests show that the convergence is global and that the number of iterations depends on three parameters : the used algorithm to solve the sub-problems, the initial points and the factor of penalty r_k.

The calculation of the dual bound for differentiable or nondifferentiable optimization problems is an interesting and vast domain of research. In spite of few attempts were made in the nondifferentiable case for these problems.

In our work, the essential purpose of this paper was to answer the following question :

Can we spread the used techniques in the differentiable case ([6]) for the nondifferentiable case by keeping the convergence as well as its rank ?

This seems clear in our study. Indeed, always basing on the augmented Lagrangian method, we solved our problem by application of the proximal quasi-Newton technique.

The development of other methods, in the nondifferentiable case, is a center of current searches and are open problems.

Acknowledgement 1. *The authors would like to thank the referees for some corrections which greatly improved the presentation of this paper.*

References

[1] M. S. Bazaraa Hanif, D. Sherali and C. M. Shety, Nonlinear programming. Theory and algorithms, John Wiley and Sons-Inc, Second edition, 1993.

[2] D. P. Bertsekas, Constrained optimization and Lagrangian multiplier methods, Academic Press, 1982.

[3] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal and C. Sagastizabal, Optimisation numérique. Aspects théoriques et pratiques, SAIM. N°. 27, Springer-Verlag, 1997.

[4] X. Chen and M. Fukushima, Proximal quasi-Newton methods for nondifferentiable convex optimization, Math. Prog. 85 (1999), pp. 313-334.

[5] J. C. Culioli, Introduction à l'optimisation, Ellipses, France, 1994.

[6] N. Daili and M. Khaldi, Duality Gap in Differentiable and Nondifferentiable Optimization Problems, Jour. of Comput. Maths. and Optim. (JCMO), 4(2) (2008), pp. 127-144.

[7] N. Daili and M. Khaldi, Dual Bounds in Convex and Nonconvex Differentiable Optimization Problems and Applications, Pacific Jour. of Appli. Maths. (PJAM), 2(1) (2009), pp. 1-26.

[8] N. Daili and M. Khaldi, Bornes Duales de Problèmes d'Optimisation Convexes, Non Convexes et Non Différentiables et Applications, Rapport de Recherche n°.CRM-3333, Centre de Recherches Mathématiques, Université de Montr éal, Québec, 2013.

[9] N. Daili and Kh. Saadi, Epsilon-Proximal Point Algorithms for Nondifferentiable Convex Optimization Problems and Applications, Adv. Model. Optim. (AMO), Vol. 14, Issue 1 (2012), pp. 175-195.

[10] N. Daili, Some Augmented Lagrangian Algorithms applied to Convex Nondifferentiable Optimization Problems, Jour. Inf.& Optim. Sci. (JIOS), Vol. 33, n°. 4-5 (2012), pp. 487-526.

[11] N. Daili, Applications de Lagrangien Augmenté à des Problèmes de Programmation Convexe non Nécessairement Diffé rentiables, Pub. 455/P261, IRMA - CNRS, Strasbourg I, 1986.

[12] J. P. Dedieu and J. C. Yakoubsohn, Localization of an algebraic hypersurface by the exclution algorithm, Applicable algebra in Engineering, Com. and Computing (AAECC) 2 (1992), pp. 239-256.

[13] V. F. Dem'Yanov and L. V. Vasil'Ev, Nondifferentiable Optimization, Optimization Software Inc, New York, 1985.

[14] Ch. Ferrier, Bornes duales de problèmes d'optimisation polynomiale, Thèse de Doctorat, Pub. Univ. Paul Sabatier, Toulouse, 1997.

[15] R. Fletcher, Pratical Methods of Optimization. Volume 2, Constrained optimization, SIAM, Philadelphie, 1980.

[16] M. Fukushima, A descent algorithm for nonsmooth convex minimization, Math. Prog. 30 (1984), pp. 163-175.

[17] C. Lemaréchal and R. Mifflin, Nonsmooth Optimization, Pergamon Press, Oxford, 1978.

[18] R. Mifflin, A quasi-second-order proximal bundle algorithm, Math. Prog. 73 (1996), pp. 51-72.

[19] M. Minoux, Programmation mathématique, Théorie et algorithmes, Tome 1, Dunod, Paris, 1983.

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

[21] R. T. Rockafellar, Convex analysis, Princeton Univ. Press, Princeton, 1970.

[22] Ph. D. Tao, Some new results in nonconvex nondifferentiable optimization, 6$^{\text{th}}$ German Conference on optimization. Lambrech, Germany, (1991), pp. 216-816.

[23] P. Wolfe, A method of conjugate subgradients for minimizing non-differentiable function, Math. Prog. (1975), pp.145-173.