Malaya J. Mat. 4 (4)(2016) 534–555

Malaya Journal of Matematik

MJM
an international journal of mathematical sciences with computer applications...

www.malayajournal.org

ISSN : 2319-3786

Nondifferentiable Augmented Lagrangian, ε-Proximal penalty methods and Applications

Noureddine. Daili[a]* and K. Saadi[b]

[a] Department of Mathematics

[a] *Cité des 300 Lots. Yahiaoui. 51, rue Harrag Senoussi. 19000 Sétif, Algeria.*

[b] Laboratoire d'Analyse Fonctionnelle et Gomtrie des Espaces, University of M'sila, BP 166 Ichbilia, M'sila.

Abstract

The purpose of this work is to prove results concerning the duality theory and to give detailed study on the augmented Lagrangian algorithms and ε-proximal penalty method which are considered, today, as the most strong algorithms to solve nonlinear differentiable and nondifferentiable problems of optimization. We give an algorithm of primal-dual type, where we show that sequences $\left\{\lambda^k\right\}_k$ and $\left\{x^k\right\}_k$ generated by this algorithm converge globally, with at least the Slater condition, to $\overline{\lambda}$ and $\overline{x}$. Numerical simulations are given.

Keywords: Convex programming, augmented Lagrangian, ε-proximal penalty method, duality, Perturbation, Convergence of algorithms.

2010 MSC: 90C25, 49M29, 49N15.

1 Introduction

The augmented Lagrangian methods present a large inconvenience of point of view stability. If we have a sequence $\left\{\lambda^k\right\}_k$ who converges to an optimum $\overline{\lambda}$ of the dual function, the successive solutions x^k obtained converge to an optimal solution only if $L(x, \lambda)$ has an unique minimum at x in a neighborhood of $\overline{\lambda}$ (it will be the case for example if $L(x, \lambda)$ is strictly convex at x).

So the methods of exterior penalties present the inconvenience that, to obtain a feasible point, we make tighten the coefficient of penalty towards the infinity, then the penalized function becomes badly conditioned for which the methods of gradients have a slow convergence

In the case of the equality constraints, Hestenes (1969) and Powell (1969) suggested combining previous both approaches (penalties and dualities), and suggested solving a sequence of unconstrained problems of the following shape:

$$L_r(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + r \sum_{i=1}^{m} (g_i(x))^2 \tag{1.1}$$

A generalization of Hestenes and Powell function to inequality constraints will be after given.

So, the general principle of these methods consists in determining a saddle point of L_r instead of solving $(\mathcal{P})$. The first component of the saddle point is, also, an optimal solution of the problem $(\mathcal{P})$.

The augmented Lagrangian method can be considered as an improvement of the penalty methods, because it avoids having to use coefficients of penalties too big.

*Corresponding author.

E-mail address: nourdaili_dz@yahoo.fr (N. Daili), kh_saadi@yahoo.fr (K.Saadi)

Besides, the fact of adding the quadratic term $r(g(x))^2$ in the classical Lagrangian will improve the properties of convergence of Lagrangian algorithms because the augmented Lagrangian is strictly convex at x. It is the case where we find an unique primal solution in the neighborhood of the dual solution.

We can say that the augmented Lagrangian has a much more fundamental interest. Today, it is widely recognized that the algorithms of optimization based on the use of the augmented Lagrangian, are a part of the most effective general methods to solve differentiable and nondifferentiable mathematical programming problems.

The purpose of this work is to prove results concerning the duality theory and to give detailed study on the augmented Lagrangian algorithms and ε-proximal penalty methods which are considered, today, as the most strong algorithms to solve nonlinear differentiable and nondifferentiable problems of optimization. Numerical experiments are given.

2 Main Results

2.1 Results on the Augmented Lagrangian

Consider the following mathematical programming problem :

$$(\mathcal{P}) \quad \begin{cases} \alpha := Inf f(x) \\ \text{ subject to } \ x \in C \end{cases} \tag{2.2}$$

where

- f is a convex function with finite values and non necessarily differentiable.
- $C := \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ i = 1, ..., m\}$, g_i $(i = 1, ..., m)$ are $\mathcal{C}^1$-convex functions.

Suppose that

$$\lim_{(\|x\| \longrightarrow +\infty)} f(x) = +\infty \text{ (i.e., } f \text{ is inf-compact)} \tag{2.3}$$

and there exists x_0 such that

$$g_i(x_0) < 0, \quad (i = 1, ..., m) \tag{2.4}$$

Definition 2.1. *The augmented Lagrangian associated to the problem $(\mathcal{P})$ is defined as follows*

$$L_r(x, \lambda) := f(x) + \frac{1}{2r}\sum_{i=1}^{m}(\psi^+(\lambda_i + rg_i(x))^2 - \lambda_i^2) \text{ for all } x \in \mathbb{R}^n,\ \lambda \in \mathbb{R}^m_+, \tag{2.5}$$

where $\psi^+(t) = Max(0, t)$. Or still

$$L_r(x, \lambda) := f(x) + \sum_{i=1}^{m} \begin{cases} \frac{r}{2}g_i^2(x) + \lambda_i g_i(x) & \text{if } \ g_i(x) > -\frac{\lambda_i}{r} \\ \\ -\frac{1}{2r}\lambda_i^2 & \text{if } \ g_i(x) \leq -\frac{\lambda_i}{r}. \end{cases} \tag{2.6}$$

Remark 2.1. *Put*

$$L_r(x, \lambda) = f(x) + \varphi(g(x), \lambda, r),$$

where

$$\varphi(u, \lambda, r) = \frac{1}{2r}\sum_{i=1}^{m}(\psi^+(\lambda_i + ru_i)^2 - \lambda_i^2), \quad u \in \mathbb{R}^m, \quad \lambda \in \mathbb{R}^m_+, \quad r > 0.$$

We notice well that

- *if $u \leq 0$, then $\varphi(u, \lambda, r) \leq 0$;*
- *if $u = 0$, then $\varphi(u, \lambda, r) = 0$.*

Corollary 2.1. *We have* $\lim_{(r \longrightarrow 0)} L_r(x, \lambda) = L(x, \lambda).$

We have the following lemma :

Lemma 2.1. *We have*

$$\underset{x\in\mathbb{R}^n}{Inf}\ \underset{(\lambda,\,r)\in T}{Sup}\ L_r(x,\lambda)=\alpha,$$

where $T=\mathbb{R}^m_+\times\mathbb{R}_+$.

Proof. At first, we notice that for all u and $c\geq 0$ there is a couple $(\lambda,r)\in T$ such that $\varphi(u,\lambda,r)>c$.
Indeed, we distinguish two cases :

- Ca **se** 1**:** If $u\nleq 0$, there exists at least one component $u_i>0$. We note by

$$I:=\left\{i\in\{1,...,m\}:u_i>-\frac{\lambda_i}{r}\right\}.$$

$I\neq\emptyset$. Then

$$\varphi(u,\lambda,r)=\sum_{i\in I}(\frac{r}{2}u_i^2+\lambda_i u_i)-\sum_{i\notin I}\frac{\lambda_i^2}{2r}.$$

If $I=\{1,...,m\}$ then $\varphi(u,\lambda,r)\longrightarrow+\infty$, as $(\lambda,r)\longrightarrow+\infty$.

Else, we have $\varphi(u,0,r)\longrightarrow+\infty$, as $(r\longrightarrow+\infty)$.

Then, in both cases there existe $(\lambda,r)\in T$ such that

$$\varphi(u,\lambda,r)>c. \tag{2.7}$$

- **Case 2:** If $u_i\leq 0$, for all $i\in\{1,...,m\}$, one has

$$\frac{1}{2r}(\psi^+(\lambda_i+ru_i)^2-\lambda_i^2)=\left\{\begin{array}{ll}\frac{r}{2}u_i^2+\lambda_i u_i & \text{if}\ \ u_i>-\frac{\lambda_i}{r}\\ \\ -\frac{1}{2r}\lambda_i^2 & \text{if}\ \ u_i\leq-\frac{\lambda_i}{r}\end{array}\right\}\leq 0$$

then

$$\underset{(\lambda,\,r)\in T}{Sup}\ \varphi(u,\lambda,r)=0 \tag{2.8}$$

By means of formulae (2.6) and (2.7), one has

$$\underset{(\lambda,\,r)\in T}{Sup}\ L_r(x,\lambda)=\left\{\begin{array}{l}f(x)\ \ \text{if}\ x\in C\\ \\ +\infty\ \ \text{else;}\end{array}\right.$$

thus

$$\underset{x\in\mathbb{R}^n}{Inf}\ \underset{(\lambda,\,r)\in T}{Sup}\ L_r(x,\lambda)=\underset{x\in C}{Inf}f(x)=\alpha.$$

□

By definition, we put

$$d_r(\lambda):=\underset{x\in\mathbb{R}^n}{Inf}L_r(x,\lambda),\ \text{for all}\ \lambda\in\mathbb{R}^m_+.$$

We have the following Lemma:

Lemma 2.2. *For all* $r>0$*, we have*

$$d_r(\lambda):=\underset{z\geq 0}{Sup}\left\{d(z)-\frac{1}{2r}\left\|z-\lambda\right\|^2\right\}\ \textit{for all}\ \lambda\in\mathbb{R}^m_+. \tag{2.9}$$

Proof. We have

$$d_r(\lambda) := \underset{z\geq 0}{Sup}\left\{d(z) - \tfrac{1}{2r}\|z-\lambda\|^2\right\}$$

$$= \underset{z\geq 0}{Sup}\left\{\underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \sum_{i=1}^{m} z_i g_i(x)\right\} - \tfrac{1}{2r}\|z-\lambda\|^2\right\}$$

$$= \underset{z\geq 0}{Sup}\left\{\underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \sum_{i=1}^{m} z_i g_i(x) - \tfrac{1}{2r}\|z-\lambda\|^2\right\}\right\}.$$

The function

$$(x,z) \longrightarrow \delta(x,z) = f(x) + \sum_{i=1}^{m} z_i g_i(x) - \frac{1}{2r}\|z-\lambda\|^2$$

admits a saddle point because it verifies the following conditions :

. $\delta(x,z)$ is convex for x an concave for z;

. $\delta(x,z)$ tends to $+\infty$ as $\|x\| \longrightarrow +\infty$ (at a point $z=0$);

. $\delta(x,z)$ tends to $-\infty$ as $\|z\| \longrightarrow +\infty$ (at a point $x_0 : g(x_0) < 0$).

Then, we can invert $SupInf$ by $InfSup$ and we have

$$d_r(\lambda) = \underset{z\geq 0}{Sup}\,\underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \sum_{i=1}^{m} z_i g_i(x) - \tfrac{1}{2r}\|z-\lambda\|^2\right\}$$

$$= \underset{x\in\mathbb{R}^n}{Inf}\,\underset{z\geq 0}{Sup}\left\{f(x) + \sum_{i=1}^{m} z_i g_i(x) - \tfrac{1}{2r}\|z-\lambda\|^2\right\}.$$

The Sup is reached at $\bar{z}$ where

$$\bar{z}_i = \left\{\begin{array}{ll} rg_i(x) + \lambda_i & \text{if } g_i(x) > -\frac{\lambda_i}{r} \\ 0 & \text{if } g_i(x) \leq -\frac{\lambda_i}{r} \end{array}\right\} = \psi^+(rg_i(x) + \lambda_i).$$

For this notation, then the function $d_r(\lambda)$ spells

$$d_r(\lambda) := \underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \sum_{i=1}^{m}\psi^+(rg_i(x)+\lambda_i)g_i(x) - \tfrac{1}{2r}\sum_{i=1}^{m}(\psi^+(rg_i(x)+\lambda_i) - \lambda_i)^2\right\}$$

$$= \underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \sum_{i=1}^{m}(\psi^+(rg_i(x)+\lambda_i)g_i(x) - \tfrac{1}{2r}(\psi^+(rg_i(x)+\lambda_i) - \lambda_i)^2)\right\}$$

$$= \underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \sum_{i=1}^{m}\left\{\begin{array}{ll} \lambda_i g_i(x) + \frac{r}{2}g_i^2(x) & \text{if } g_i(x) > -\frac{\lambda_i}{r} \\ -\frac{1}{2r}\lambda_i^2 & \text{if } g_i(x) \leq -\frac{\lambda_i}{r} \end{array}\right\}\right.$$

$$= \underset{x\in\mathbb{R}^n}{Inf}\left\{f(x) + \tfrac{1}{2r}\sum_{i=1}^{m}(\psi^+(rg_i(x)+\lambda_i)^2 - \lambda_i^2))\right\} = \underset{x\in\mathbb{R}^n}{Inf}\, L_r(x,\lambda).$$

According to ([3], remark 2.1), d_r is the regularized function of d. It is, thus, differentiable at λ and we have

$$\nabla d_r(\lambda) = -\frac{1}{r}(\lambda - z_\lambda)$$

where z_λ realizes the Sup in the expression (2.8). We note, also, that d_r has the same optimal solutions as d. □

Definition 2.2. *The dual problem associated to the problem* $(\mathcal{P})$ *is the following one :*

$$(\mathcal{D}) \quad \beta := \underset{(\lambda,\, r)\in T}{Sup}\; d_r(\lambda), \tag{2.10}$$

where $T = \mathbb{R}_+^m \times \mathbb{R}_+$.

Definition 2.3. *We call perturbation function of* $(\mathcal{P})$ *the function* p *defined by*

$$p(u) := \underset{x \in \mathbb{R}^n}{Inf}\, F(x,u),$$

where

$$F(x,u) := \begin{cases} f(x) \ \ if \ \ g(x) \le u, \\ \\ +\infty \quad else \end{cases} \tag{2.11}$$

Remark 2.2. . *If* $u = 0$ *then* $p(0) = \alpha$
. *If* $u_1 \ge u_2$ *then* $p(u_2) \ge p(u_1)$.

The following lemma shows the relation which exists between L_r and F.

Lemma 2.3. *We have*

$$L_r(x,\lambda) = \underset{u \in \mathbb{R}^m}{Inf}\ \{F(x,u) + \varphi(u,\lambda,r)\} \tag{2.12}$$

for all $x \in \mathbb{R}^n$ *and* $(\lambda, r) \in T$.

Proof. Let $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$.
. If $g(x) \le u$ we have

$$F(x,u) = f(x) \ \text{ and } \ \varphi(g(x),\lambda,r) \le \varphi(u,\lambda,r), \ \ \forall(\lambda,r) \in T.$$

. If $g(x) \nleq u$ we have $F(x,u) = +\infty$. Then

$$L_r(x,\lambda) = f(x) + \varphi(g(x),\lambda,r) \le F(x,u) + \varphi(u,\lambda,r), \ \ \forall u \in \mathbb{R}^m,$$

thus

$$L_r(x,\lambda) \le \underset{u \in \mathbb{R}^m}{Inf}\ \{F(x,u) + \varphi(u,\lambda,r)\},$$

but

$$L_r(x,\lambda) = F(x,g(x)) + \varphi(g(x),\lambda,r) \ge \underset{u \in \mathbb{R}^m}{Inf}\ \{F(x,u) + \varphi(u,\lambda,r)\}.$$

Then both inequalities give the expression (2.11). □

Lemma 2.4. *We have*

$$L_r(x,\lambda) = \underset{u \in \mathbb{R}^m}{Inf}\ \left\{F(x,u) + < \lambda, u > + \frac{r}{2}\|u\|^2\right\}, \tag{2.13}$$

where F *is given by the expression* (2.11).

Proof. Indeed, let us put

$$\varphi_r(x^k,\lambda) = \underset{u \in \mathbb{R}^m}{Inf}\ \left\{F(x^k,u) + < \lambda, u > + \frac{r}{2}\|u\|^2\right\}$$

The Inf in the expression of $\varphi_r(x^k,\lambda)$ exists and unique (the function at u is strongly convex). For every x, we indicate by C_x the following set :

$$C_x := \{u \in \mathbb{R}^m : u \ge g(x)\}.$$

Then, the expression (2.12) becomes

$$\varphi_r(x^k,\lambda) = \underset{u \in C_x}{Inf}\ \left\{F(x,u) + < \lambda, u > + \tfrac{r}{2}\|u\|^2\right\} = \underset{u \in C_x}{Inf}\ \left\{f(x) + < \lambda, u > + \tfrac{r}{2}\|u\|^2\right\}$$

$$= f(x) + \underset{u \in C_x}{Inf}\ \left\{< \lambda, u > + \tfrac{r}{2}\|u\|^2\right\}.$$

To calculate the solution of $\underset{u \in C_x}{Inf}\ \left\{< \lambda, u > + \tfrac{r}{2}\|u\|^2\right\}$ we look for a minimization according to every i. Let us put

$$w(u) = < \lambda, u > + \frac{r}{2}\|u\|^2,$$

then, $\nabla w(u) = \lambda + ru$.

For all i, if $g_i(x) < -\frac{\lambda_i}{r}$ then, $\overline{u}_i = -\frac{\lambda_i}{r}$, else $\overline{u}_i = g_i(x)$. Thus

$$\underset{x\in C_x}{Inf}\, w(x) = \sum_{i=1}^{m} \begin{cases} \lambda_i g_i(x) + \frac{1}{2r}(g_i(x))^2 & \text{if } g_i(x) \geq -\frac{\lambda_i}{r} \\ -\frac{1}{2r}\lambda_i^2 & \text{if } g_i(x) < -\frac{\lambda_i}{r} \end{cases} = \varphi(g(x), \lambda, r).$$

Then $\quad \varphi_r(x^k, \lambda) = L_r(x, \lambda)$. □

We notice that L_r is convex at x and concave at (λ, r), consequently d_r is concave.

We have the following weak duality theorem :

Theorem 2.1. *(Weak duality) We have*

$$\beta \leq \alpha. \tag{2.14}$$

Proof. We always have

$$\underset{(\lambda,r)\in T}{Sup}\ \underset{x\in\mathbb{R}^n}{Inf}\, L_r(x, \lambda) \leq \underset{x\in\mathbb{R}^n}{Inf}\ \underset{(\lambda,r)\in T}{Sup}\ L_r(x, \lambda),$$

thus

$$\beta \leq \alpha.$$

□

Another relation exists between d_r and p is given by the following lemma :

Lemma 2.5. *We have*

$$d_r(\lambda) = \underset{u\in\mathbb{R}^n}{Inf}\, \{p(u) + \varphi(u, \lambda, r)\}, \ \ \forall\, (\lambda, r) \in T. \tag{2.15}$$

Proof. We have, according to the Lemma 2.4,

$$d_r(\lambda) = \underset{x\in\mathbb{R}^n}{Inf}\, L_r(x, \lambda) = \underset{x\in\mathbb{R}^n}{Inf}\ \underset{u\in\mathbb{R}^n}{Inf}\, \{F(x, u) + \varphi(u, \lambda, r)\}$$

$$= \underset{u\in\mathbb{R}^n}{Inf}\ \underset{x\in\mathbb{R}^n}{Inf}\, \{F(x, u) + \varphi(u, \lambda, r)\} = \underset{u\in\mathbb{R}^n}{Inf}\, \{p(u) + \varphi(u, \lambda, r)\}.$$

□

Lemma 2.6. *There is a function Φ such that for all $(\lambda, r) \in T,\ (z, s) \in T,\ r > s$, we have*

$$\varphi(u, \lambda, r) - \varphi(u, z, s) \geq -\Phi(\lambda, z, s, r)$$

with

$$\lim_{(r\longrightarrow+\infty)} \Phi(\lambda, z, s, r) = 0. \tag{2.16}$$

Proof. We have

$$\varphi(u, \lambda, r) - \varphi(u, z, s) = \frac{1}{2r}\sum_{i=1}^{m}\Psi^+(ru_i + \lambda_i) - \frac{1}{2s}\sum_{i=1}^{m}\Psi^+(su_i + z_i).$$

We distinguish two cases :

Case 1:

. If $u_i \leq -\frac{\lambda_i}{r}$, then $\Psi^+(u_i + \lambda_i) = 0$.

. If $u_i \leq -\frac{z_i}{s}$, then $\Psi^+(su_i + z_i) = 0$, thus

$$\frac{1}{2r}\Psi^+(ru_i + \lambda_i) - \frac{1}{2s}\Psi^+(su_i + z_i) = 0.$$

. If $u_i > -\frac{z_i}{s}$, then $-\frac{z_i}{s} < u_i \leq -\frac{\lambda_i}{r}$, from hence

$$\frac{1}{2s}\Psi^+(su_i + z_i) = \frac{s}{2}u_i^2 + z_i u_i \leq \frac{s}{2}(-\frac{\lambda_i}{r})^2 + z_i(-\frac{\lambda_i}{r}) \leq \frac{s}{2}\frac{\lambda_i^2}{r^2} - \frac{z_i\lambda_i}{r}.$$

As $r > s$, then, we have

$$\frac{1}{2s}\Psi^{+}(su+z) \leq \frac{\lambda_i^2}{2r} - \frac{z_i\lambda_i}{r}.$$

It holds that

$$\frac{1}{2r}\Psi^{+}(ru_i+\lambda_i) - \frac{1}{2s}\Psi^{+}(su_i+z_i) \geq -(-\frac{\lambda_i^2}{2r} + \frac{z_i\lambda_i}{r}) \longrightarrow 0, \text{ as } r \longrightarrow +\infty.$$

Case 2:

. If $u_i > -\frac{\lambda_i}{r}$, then

$$\frac{1}{2r}\Psi^{+}(ru_i+\lambda_i) = \frac{r}{2}u_i^2 + \lambda_i u_i.$$

. If $u_i \leq -\frac{z_i}{s}$, then

$$\frac{1}{2s}\Psi^{+}(su_i+z_i) = 0,$$

thus $-\frac{\lambda_i}{r} < u_i \leq -\frac{z_i}{s}$, it holds that

$$\frac{1}{2s}\Psi^{+}(ru_i+\lambda_i) - \frac{1}{2s}\Psi^{+}(su_i+z_i) = \frac{r}{2}u_i^2 + \lambda_i u_i \geq -(-\frac{r}{2}(-\frac{z_i}{s})^2 - \lambda_i(-\frac{z_i}{s}))$$

$$\geq -(-\frac{z_i^2}{2r} + \lambda_i\frac{z_i}{r}) \longrightarrow 0, \text{ as } r \longrightarrow +\infty.$$

. If $u_i > -\frac{z_i}{s}$, then, we have

$$\frac{1}{2r}\Psi^{+}(ru_i+\lambda_i) - \frac{1}{2s}\Psi^{+}(su_i+z_i) = \frac{1}{2}u_i^2(r-s) + (\lambda_i - z_i)u_i$$

$$\geq \frac{1}{2}(\frac{z_i-\lambda_i}{r-s})^2 + (\lambda_i - z_i)\frac{(z_i-\lambda_i)}{r-s}$$

$$\geq \frac{1}{2}(\frac{z_i-\lambda_i}{r-s})^2 + \frac{(z_i-\lambda_i)^2}{r-s} \longrightarrow 0, \text{ as } r \longrightarrow +\infty.$$

Finally, in every cases there is a function Φ verifying

$$\varphi(u,\lambda,r) - \varphi(u,z,s) \geq -\Phi(\lambda,z,s,r), \text{ for all } (\lambda,r) \in T,\ (z,s) \in T,\ \ r > s$$

with

$$\lim_{(r\longrightarrow+\infty)} \Phi(\lambda,z,s,r) = 0.$$

□

It results from this lemma the following result :

Lemma 2.7. *For all $(\lambda,r) \in T,\ (r > 0)$, we have*

$$d_r(\lambda) \geq \underset{(z,s)\in T,\ (r>s>0)}{Sup} (d_s(z) - \Phi(\lambda,z,s,r)).$$

Proof. According to the Lemma 2.8, we have

$$\varphi(u,\lambda,r) \geq \varphi(u,z,s) - \Phi(\lambda,z,s,r).$$

Hence

$$p(u) + \varphi(u,\lambda,r) \geq p(u) + \varphi(u,z,s) - \Phi(\lambda,z,s,r)\ \ \forall\, u \in \mathbb{R}^m,$$

then

$$\underset{u\in\mathbb{R}^m}{Inf}\,(p(u) + \varphi(u,\lambda,r)) \geq \underset{u\in\mathbb{R}^m}{Inf}\,(p(u) + \varphi(u,z,s) - \Phi(\lambda,z,s,r)).$$

It holds that

$$d_r(\lambda) \geq d_s(z) - \Phi(\lambda,z,s,r),\ \ \forall (z,s) \in T,\ \forall r > s > 0$$

$$\Longrightarrow d_r(\lambda) \geq \underset{(z,s)\in T,\ (r>s>0)}{Sup} (d_s(z) - \Phi(\lambda,z,s,r)).$$

□

We have the following theorem :

Theorem 2.2. *We have*

$$\beta = \underset{(z,s)\in T}{Sup}\ d_s(z) = \lim_{(r\longrightarrow+\infty)} d_r(\lambda),\ \textit{for all}\ \lambda \in \mathbb{R}^m_+. \tag{2.17}$$

Proof. For all $(z,s) \in T,\ \varepsilon > 0$ and $\lambda \in \mathbb{R}^m_+$, it exists, according to the Lemma 2.8, r enough large, with $(r > s)$ such that $\Phi(\lambda, z, s, r) < \varepsilon$. Then

$$d_r(\lambda) \geq d_s(z) - \varepsilon,\ \ \forall\varepsilon > 0,$$

thus

$$\lim_{(r\longrightarrow+\infty)} d_r(\lambda) \geq d_s(z) - \varepsilon,\ \ \forall\varepsilon > 0,\ \forall(z,s) \in T.$$

And then, for every $\varepsilon > 0$

$$\lim_{(r\longrightarrow+\infty)} d_r(\lambda) \geq \underset{(z,s)\in T}{Sup}\ d_s(z) - \varepsilon,$$

thus

$$\lim_{(r\longrightarrow+\infty)} d_r(\lambda) \geq \underset{(z,s)\in T}{Sup}\ d_s(z).$$

On the other hand,

$$\underset{(z,s)\in T}{Sup}\ d_s(z) \geq d_r(\lambda),\ \ \forall\lambda \in \mathbb{R}^m_+$$

$$\underset{(z,s)\in T}{Sup}\ d_s(z) \geq \lim_{(r\longrightarrow+\infty)} d_r(\lambda),$$

where holds the result. □

This theorem gives a technique of resolution of $(\mathcal{D})$. Indeed; if we penalize the function d, by using the term of penalty $(-\frac{1}{2r}\|z-\lambda\|)$, then by making the resolution when $(r \longrightarrow +\infty)$, we are in front of a said penalty method.

The following algorithm shows the necessary steps for the resolution :

Algorithm 1:

Step 1: $(k = 0)$

Fixe λ and we choose a factor of penalty $r_0 > 0$ and $z_0 \in \mathbb{R}^m_+,\ (k = 0)$.

Step 2: $(k \geq 0)$

Find z_k solution of

$$d_{r_k}(\lambda) = \underset{z\geq 0}{Sup}\left\{d(z) - \frac{1}{2r_k}\|z-\lambda\|^2\right\}.$$

Step 3:

If z_k do not verify the stop test one makes $r_{k+1} > r_k$, $k \longrightarrow k+1$ and we return to the step 1.

2.2 Augmented Lagrangian Algorithms

Let $(\mathcal{P})$ be the following constrained mathematical programming problem :

$$(\mathcal{P}) \qquad \alpha := \underset{x\in C}{Inf} f(x),$$

where

. f is a non necessarely differentiable convex function with finite value ;

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

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

Suppose that $\lim_{(\|x\|\longrightarrow+\infty)} f(x) = +\infty$ and there exists x_0 such that

$$g_i(x_0) \leq 0,\ i = 1, ..., m.$$

We give an algorithm with which we can calculate optimal solutions of $(\mathcal{P})$.

Algorithm 2:
Step 1: $(k=0)$
Let us fix $r>0$, let us determine one $\overline{\lambda}$ in which the function d_r reaches its maximum on $\mathbb{R}^m$.
Step 2:
Let us look for any point $\overline{x}$ which minimizes the convex function $L_r(.,\overline{\lambda})$ on $\mathbb{R}^n$.

Remark 2.3. *The essential difficulty in the previous method lies in the calculation of $(\overline{x},\overline{\lambda})$. This couple is not calculable with accuracy. But, if $\overline{\lambda}$ and $\overline{x}$ are approximately determined by the previous method, can we be sure that $\overline{x}$ is, approximately, an optimal solution of $(\mathcal{P})$?*

Another complication appears because of the non direct clarified of d_r at the wished way.

However, we can calculate $d_r(\lambda)$ and $\nabla d_r(\lambda)$, for every λ, by determining a point x which minimizes $L_r(.,\lambda)$ on $\mathbb{R}^n$. This operation is too expensive (from point of view cost) by repeating, every time, the process of iteration.

To by-pass this difficulty, let us suppose that for one λ given, we have one $x\in\mathbb{R}^n$ minimizing $L_r(.,\lambda)$ on $\mathbb{R}^n$ with a precision $\varepsilon\geq 0$, that is

$$L_r(x,\lambda)-d_r(\lambda)\leq\varepsilon.$$

We see that

$$d_r(\lambda')\leq L_r(x,\lambda')\leq L_r(x,\lambda)+<\lambda'-\lambda,\nabla_\lambda L_r(x,\lambda)>\quad\forall x\in\mathbb{R}^n,\ \lambda'\in\mathbb{R}^m$$

$$\Longrightarrow d_r(\lambda')\leq d_r(\lambda)+<\lambda'-\lambda,\nabla_\lambda L_r(x,\lambda)>+\varepsilon.$$

It holds that $\nabla_\lambda L_r(x,\lambda)$ is an ε -subgradient of d_r at λ.

Definition 2.4. *. A sequence $\left\{x^k\right\}_k$ of $\mathbb{R}^n$ is called asymptotically feasible for the problem $(\mathcal{P})$ if*

$$\lim_{(k\longrightarrow+\infty)} g_i(x^k)\leq 0,\ i=1,...,m.$$

. A sequence which realizes the Sup of the problem $(\mathcal{D})$ is a sequence $\left\{\lambda^k\right\}_k$ of $\mathbb{R}^m$ such that

$$d_r(\lambda^k)\longrightarrow Supd_r,\ as\ (k\longrightarrow+\infty).$$

. An asymptotically minimizing sequence of $(\mathcal{P})$ is a sequence $\left\{x^k\right\}_k$ asymptotically feasible and such that

$$\lim_{(k\longrightarrow+\infty)} f(x^k)=\alpha.$$

Theorem 2.3. *Let $\left\{\lambda^k\right\}_k$ be a bounded sequence wich maximizes $(\mathcal{D})$, let $\left\{x^k\right\}_k$ be a sequence satisfying*

$$L_r(x^k,\lambda^k)-\underset{x\in\mathbb{R}^n}{Inf}\,L_r(x,\lambda^k)=L_r(x^k,\lambda^k)-d_r(\lambda^k)\leq\varepsilon_k,$$

where $\varepsilon_k\longrightarrow 0$ as $k\longrightarrow+\infty$.

Then $\left\{x^k\right\}_k$ is an asymptotically minimizing sequence of $(\mathcal{P})$.

For the proof of this theorem, we need to the following three lemmas:

Lemma 2.8. *The function d_r satisfies, for all $\lambda,\ \lambda'\in\mathbb{R}^m_+$*

$$\begin{aligned}&d_r(\lambda')\leq d_r(\lambda)+<\lambda'-\lambda,\nabla d_r(\lambda)>\\&d_r(\lambda')\geq d_r(\lambda)+<\lambda'-\lambda,\nabla d_r(\lambda)>-\tfrac{1}{2r}\left\|\lambda'-\lambda\right\|^2\end{aligned}\tag{2.18}$$

Proof. The first inequality is immediate from the concavity of $d_r(\lambda)$.

For the second inequality, we have

$$d_r(\lambda)=\underset{z\in\mathbb{R}^n_+}{Sup}\left\{d(z)-\frac{1}{2r}\left\|\lambda-z\right\|^2\right\}.$$

It exists an unique z_λ such that

$$d_r(\lambda) = d(z_\lambda) - \frac{1}{2r}\|\lambda - z_\lambda\|^2.$$

Let us put

$$q(\lambda') := d(z_\lambda) - \frac{1}{2r}\left\|\lambda' - z_\lambda\right\|^2.$$

Or $q(\lambda')$ is quadratic, we shall have

$$q(\lambda') = q(\lambda) + <\lambda' - \lambda, \nabla q(\lambda)> + \frac{1}{2}(\lambda' - \lambda)^t \nabla q(\lambda)(\lambda' - \lambda).$$

Because

$$q(\lambda) = d_r(\lambda) \quad \text{and} \quad q(\lambda') \le d_r(\lambda'), \ \forall \lambda',$$

it holds that

$$\nabla q(\lambda) = \nabla d_r(\lambda).$$

On the other hand,

$$\nabla^2 q(\lambda) = -\frac{1}{r} Id \ \text{ (where } Id \text{ is an identity matrix),}$$

then

$$q(\lambda') = d_r(\lambda) + <\lambda' - \lambda, \nabla d_r(\lambda)> - \frac{1}{2r}\left\|\lambda' - \lambda\right\|^2.$$

So

$$d_r(\lambda) + <\lambda' - \lambda, \nabla d_r(\lambda)> - \frac{1}{2r}\left\|\lambda' - \lambda\right\|^2 \le d_r(\lambda').$$

□

Lemma 2.9. *We have*

$$\frac{r}{2}\left\|\nabla d_r(\lambda^k)\right\|^2 \le Supd_r - d_r(\lambda^k). \tag{2.19}$$

Proof. According to Lemma 2.12, it holds that

$$Supd_r \ge \underset{\lambda' \in \mathbb{R}^m_+}{Sup}\left\{d_r(\lambda^k) + <\lambda' - \lambda, \nabla d_r(\lambda^k)> - \tfrac{1}{2r}\left\|\lambda' - \lambda\right\|^2\right\}$$

$$= d_r(\lambda^k) + \underset{\lambda' \in \mathbb{R}^m_+}{Sup}\left\{<\lambda' - \lambda, \nabla d_r(\lambda^k)> - \tfrac{1}{2r}\left\|\lambda' - \lambda\right\|^2\right\}$$

$$= d_r(\lambda^k) + \tfrac{r}{2}\left\|\nabla d_r(\lambda^k)\right\|^2$$

what gives

$$\frac{r}{2}\left\|\nabla d_r(\lambda^k)\right\|^2 \le Supd_r - d_r(\lambda^k).$$

□

Lemma 2.10. *Let us consider following both properties:*
(**a**) $L_r(x^k, \lambda^k) - \underset{x \in \mathbb{R}^n}{Inf}\, L_r(x, \lambda^k) = L_r(x^k, \lambda^k) - d_r(\lambda^k) \le \varepsilon_k$,
where $\varepsilon_k \longrightarrow 0$, *as* $k \longrightarrow +\infty$;
(**b**) $\frac{r}{2}\left\|\nabla_\lambda L_r(x^k, \lambda^k) - \nabla d_r(\lambda^k)\right\|^2 \le \varepsilon_k$.
Then (**a**) $\Longrightarrow$ (**b**).

Proof. We use the Lemma 2.12 and the concavity of $L_r(x^k, .)$ then, we shall have for every $w \in \mathbb{R}^m$

$$d_r(w) \le L_r(x^k, w) \le L_r(x^k, \lambda^k) + <w - \lambda^k, \nabla_\lambda L_r(x^k, \lambda^k)>$$

and

$$d_r(w) \ge d_r(\lambda^k) + <w - \lambda^k, \nabla d_r(\lambda^k)> - \frac{1}{2r}\left\|w - \lambda^k\right\|^2$$

what gives

$$L_r(x^k,\lambda^k) - d_r(\lambda^k) \geq < w-\lambda^k, \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) > -\frac{1}{2r}\left\| w-\lambda^k \right\|^2.$$

That is

$$L_r(x^k,\lambda^k) - d_r(\lambda^k) \geq \underset{w\in R_+^m}{Sup}\left\{ < w-\lambda^k, \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) > -\tfrac{1}{2r}\left\| w-\lambda^k \right\|^2 \right\}$$

$$= r\left\| \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) \right\|^2 - \tfrac{r}{2}\left\| \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) \right\|^2$$

$$= \tfrac{r}{2}\left\| \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) \right\|^2.$$

Where, according to (**a**), we have

$$\frac{r}{2}\left\| \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) \right\|^2 \leq \varepsilon_k.$$

□

Proof. (**Theorem 2.11**) According to the Lemma 2.14 we have

$$L_r(x^k,\lambda) = Inf\left\{ F(x^k,u) + < x,u > + \frac{r}{2}\|u\|^2 \right\},$$

where F is given by

$$F(x,u) = \begin{cases} f(x) \ \text{ if } g_i(x) \leq u_i, \ i=1,...,m \\ \\ +\infty \ \text{ else.} \end{cases}$$

For $\lambda = \lambda^k$, there is an unique point u^k such that

$$L_r(x^k,\lambda^k) = F(x^k,u^k) + < \lambda^k, u^k > + \frac{r}{2}\left\| u^k \right\|^2.$$

Let us put

$$q(\lambda) = F(x^k,u^k) + < \lambda, u^k > + \frac{r}{2}\left\| u^k \right\|^2.$$

We notice that

$$q(\lambda) \geq L_r(x^k,\lambda) \ \ \forall\lambda, \ \text{ and } \ q(\lambda^k) = L_r(x^k,\lambda^k),$$

thus

$$\nabla q(\lambda^k) = \nabla_\lambda L_r(x^k,\lambda^k).$$

Then,

$$u^k = \nabla_\lambda L_r(x^k,\lambda^k).$$

We have by hypothesis

$$L_r(x^k,\lambda^k) - d_r(\lambda^k) \leq \varepsilon_k$$

what implies that

$$\lim_k L_r(x^k,\lambda^k) = \lim_k d_r(\lambda^k) = Sup d_r.$$

According to the Lemma 2.13 and Lemma 2.14, we have

$$\left.\begin{array}{l} \lim_k \nabla d_r(\lambda^k) = 0 \\ \\ \lim_k \frac{r}{2}\left\| \nabla d_r(\lambda^k) - \nabla_\lambda L_r(x^k,\lambda^k) \right\|^2 = 0 \end{array}\right\} \Longrightarrow \lim_k \nabla d_r(\lambda^k) = \lim_k \nabla_\lambda L_r(x^k,\lambda^k) = 0.$$

Then, $\lim_k u^k = 0$.

The sequence $\left\{\lambda^k\right\}_k$ being bounded, then

$$F(x^k,u^k) = L_r(x^k,\lambda^k) - < \lambda^k, u^k > -\tfrac{r}{2}\left\|u^k\right\|^2$$

$$\Longrightarrow \lim_k F(x^k,u^k) = \lim_k (L_r(x^k,\lambda^k) - < \lambda^k, u^k > -\tfrac{r}{2}\left\|u^k\right\|^2) = Sup d_r.$$

We always have $\quad d_r(\lambda) \leq f(x), \ \ \forall\lambda, \ \forall x,$ thus

$$\lim_k F(x^k,u^k) = Sup d_r(\lambda) \leq \alpha.$$

On the other hand,

$$\lim_k F(x^k,u^k) = \lim_k f(x^k) \ \text{ with } \lim_k g_i(x^k) \leq 0 \quad (i = 1,...,m).$$

Then

$$\lim_k f(x^k) = Sup d_r(\lambda) \leq \alpha \ \text{ with } \lim_k g_i(x^k) \leq 0 \quad (i = 1,...,m).$$

It holds $\lim_k f(x^k) = \alpha$. Consequently $\left\{x^k\right\}_k$ is an asymtotically minimizing sequence of $(\mathcal{P})$. □

2.3 Study of the Convergence

We are going to give an algorithm of primal-dual type, where we show that sequences $\left\{\lambda^k\right\}_k$ and $\left\{x^k\right\}_k$ generated by this algorithm converge globally, with at least the Slater condition, to $\overline{\lambda}$ and $\overline{x}$.

The algorithm to be studied depends on the initial choice of $r_0 > 0$, $\lambda^0 \in \mathbb{R}^m$ and the sequence $\{\varepsilon_k\}_k$ with

$$\varepsilon_k \geq 0 \ \text{ and } \ \lim_k \varepsilon_k = 0.$$

Algorithm 3:
Step 0: (initialization) $(k = 0)$
Choose a factor of penalty $r_k > 0$, a precision $\delta > 0$, a multiplier λ^0 and a sequence $\{\varepsilon_k\}_k$ with $\varepsilon_k \geq 0$ and $\lim_k \varepsilon_k = 0$
Step 1: $(k \geq 0)$
Find x^k such that

$$L_{r_k}(x^k,\lambda^k) - d_{r_k}(\lambda^k) \leq \varepsilon_k.$$

Step 2:
Define

$$\lambda_i^{k+1} = \max\left\{\lambda_i^k + r_k g_i(x^k), 0\right\};$$

or

$$\lambda^{k+1} = \lambda^k + r_k \nabla_\lambda L_{r_k}(x^k,\lambda^k).$$

Step 3:
If

$$\left\|\nabla_\lambda L_{r_k}(x^k,\lambda^k)\right\| \leq \delta \tag{2.20}$$

Stop and sets x^k as solution of $(\mathcal{P})$.
Else, $r_{k+1} \geq r_k$ (if need be) return to the step 1.

Lemma 2.11. *([2]) Suppose that the sequence $\left\{\lambda^k\right\}_k$ is bounded (bounded by M), then the expression* (2.20) *implies*

$$f(\overline{x}) \geq f(x^k) - \sigma_k,$$

where

$$\sigma_k = \delta(M + (2\varepsilon_k + \frac{3r_k\,\delta}{2})) + \varepsilon_k.$$

Proof. Let $\overline{x}$ be a solution of $(\mathcal{P})$. From the formula (2.17), we have

$$d_{r_k}(\lambda^{k+1}) \geq d_{r_k}(\lambda^k) + < \lambda^{k+1} - \lambda^k, \nabla_\lambda d_{r_k}(\lambda^k) > -\frac{1}{2r_k}\left\|\lambda^{k+1} - \lambda^k\right\|^2.$$

Thus

$$\begin{aligned} f(\overline{x}) &\geq d_{r_k}(\lambda^{k+1}) \\ &\geq d_{r_k}(\lambda^k) - \left\|\lambda^{k+1} - \lambda^k\right\| \left\|\nabla_\lambda d_{r_k}(\lambda^k)\right\| - \tfrac{1}{2r_k}\left\|\lambda^{k+1} - \lambda^k\right\|^2. \end{aligned}$$

According to the step 2 and the step 3 of the Algorithm 3, we have

$$\begin{aligned} f(\overline{x}) &\geq d_{r_k}(\lambda^k) - r_k\left\|\nabla_\lambda L_r(x^k,\lambda^k)\right\| \left\|\nabla_\lambda d_{r_k}(\lambda^k)\right\| - \tfrac{r_k}{2}\left\|\nabla_\lambda L_r(x^k,\lambda^k)\right\|^2 \\ &\geq d_{r_k}(\lambda^k) - r_k\delta\left\|\nabla_\lambda d_{r_k}(\lambda^k)\right\| - \tfrac{r_k}{2}\delta^2. \end{aligned}$$

From the Lemma 2.14, we have

$$\frac{r_k}{2}\left\|\nabla d_r(\lambda^k)\right\| - \frac{r_k}{2}\left\|\nabla_\lambda L_r(x^k,\lambda^k)\right\| \leq \frac{r_k}{2}\left\|\nabla_\lambda L_r(x^k,\lambda^k) - \nabla d_r(\lambda^k)\right\|^2 \leq \varepsilon_k.$$

What implies that

$$\left\|\nabla d_r(\lambda^k)\right\| \leq \frac{2\varepsilon_k}{r_k} + \delta \Longrightarrow -\left\|\nabla_\lambda d_{r_k}(\lambda^k)\right\| \geq -(\frac{2\varepsilon_k}{r_k} + \delta).$$

It results that

$$\begin{aligned} f(\overline{x}) &\geq d_{r_k}(\lambda^k) - r_k\delta(\tfrac{2\varepsilon_k}{r_k} + \delta) - \tfrac{r_k}{2}\delta^2 \\ &= d_{r_k}(\lambda^k) - \delta(2\varepsilon_k + \tfrac{3r_k}{2}\delta). \end{aligned}$$

On the other hand, according to the step 1 of the same Algorithm, we have

$$d_{r_k}(\lambda^k) \geq L_{r_k}(x^k,\lambda^k) - \varepsilon_k.$$

Then

$$\begin{aligned} d_{r_k}(\lambda^k) &\geq f(x^k) + \tfrac{1}{2r_k}\sum_{i=1}^m(\Psi^+(\lambda_i^k + r_k g_i(x^k))^2 - (\lambda_i^k)^2) - \varepsilon_k \\ &\geq f(x^k) + \tfrac{1}{2r_k}\sum_{i=1}^m((\lambda_i^{k+1})^2 - (\lambda_i^k)^2) - \varepsilon_k \\ &= f(x^k) + \tfrac{1}{2r_k}\sum_{i=1}^m(\lambda_i^{k+1} - \lambda_i^k)(\lambda_i^{k+1} + \lambda_i^k) - \varepsilon_k \end{aligned}$$

namely,

$$\begin{aligned} d_{r_k}(\lambda^k) &\geq f(x^k) + \tfrac{1}{2r_k}\sum_{i=1}^m r_k\tfrac{\partial L(x^k,\lambda^k)}{\partial\lambda_i}(\lambda_i^{k+1} + \lambda_i^k) - \varepsilon_k \\ &= f(x^k) + \tfrac{1}{2}\sum_{i=1}^m\tfrac{\partial L(x^k,\lambda^k)}{\partial\lambda_i}(\lambda_i^{k+1} + \lambda_i^k) - \varepsilon_k \\ &= f(x^k) + \tfrac{1}{2} < \nabla_\lambda L_r(x^k,\lambda^k), \lambda^{k+1} + \lambda^k > -\varepsilon_k. \end{aligned}$$

Thus

$$\begin{aligned} d_{r_k}(\lambda^k) &\geq f(x^k) - \tfrac{1}{2}\left\|\nabla_\lambda L_r(x^k,\lambda^k)\right\| \left\|\lambda^{k+1} + \lambda^k\right\| - \varepsilon_k \\ &\geq f(x^k) - \tfrac{\delta}{2}\left\|\lambda^{k+1} + \lambda^k\right\| - \varepsilon_k. \end{aligned}$$

Finally, we have

$$\begin{aligned} f(\overline{x}) &\geq d_{r_k}(\lambda^k) - \delta(2\varepsilon_k + \tfrac{3r_k}{2}\delta) \\ &\geq f(x^k) - \tfrac{\delta}{2}\left\|\lambda^{k+1} + \lambda^k\right\| - \varepsilon_k - \delta(2\varepsilon_k + \tfrac{3r_k}{2}\delta), \end{aligned}$$

it holds that

$$f(\overline{x}) \geq f(x^k) - \delta(M + (2\varepsilon_k + \frac{3r_k}{2}\delta)) - \varepsilon_k.$$

□

The general result is given by the following theorem :

Theorem 2.4. *Let us suppose that $(\mathcal{P})$ possesses a K-T vector and that*

$$\sum_{k\geq 1} \sqrt{\varepsilon_k} < +\infty \tag{2.21}$$

Then, the following properties are satisfied :
$(\mathbf{a})$ *the sequence $\left\{\lambda^k\right\}_k$ is bounded, and its cluster values are K-T vectors ;*
$(\mathbf{b})$ *the sequence $\left\{x^k\right\}_k$ is an asymtotically minimizing of $(\mathcal{P})$.*

Proof. $(\mathbf{a})$ According to the Lemma 2.14 and by hypothesis (step 1), we have

$$\frac{r_k}{2}\left\|\nabla_\lambda L_{r_k}(x^k,\lambda^k) - \nabla d_{r_k}(x^k)\right\| \leq \varepsilon_k.$$

According ([3], remark 2.2), we have

$$\nabla d_{r_k}(\lambda^k) = \frac{1}{r_k}(z_{\lambda^k} - \lambda^k)$$

where z_{λ^k} realizes the Sup in the definition of d_{r_k}. But

$$\lambda^{k+1} = \lambda^k + r_k \nabla_\lambda L_{r_k}(x^k,\lambda^k).$$

From which it holds

$$\nabla_\lambda L_{r_k}(x^k,\lambda^k) = \frac{1}{r_k}(\lambda^{k+1} - \lambda^k).$$

Then

$$\frac{r_k}{2}\left\|\nabla_\lambda L_{r_k}(x^k,\lambda^k) - \nabla d_{r_k}(x^k)\right\|^2 = \frac{r_k}{2}\left\|\frac{1}{r_k}(\lambda^{k+1} - \lambda^k) - \frac{1}{r_k}(z_{\lambda^k} - \lambda^k)\right\|^2 \leq \varepsilon_k$$

Namely

$$\frac{1}{2r_k}\left\|\lambda^{k+1} - z_{\lambda^k}\right\|^2 \leq \varepsilon_k.$$

Taking the limit on k we find

$$\lim_k(\lambda^{k+1} - z_{\lambda^k}) = 0 \tag{2.22}$$

Consider the application $Prox$ defined by

$$z \longrightarrow Prox(z) = h(z) + \frac{1}{2}\|z-\lambda\|^2$$

where h is a convex function. Let us put

$$Prox(h;\lambda) = \arg\min_z \left\{h(z) + \frac{1}{2}\|z-\lambda\|^2\right\}.$$

We have, according ([5], Theo.31.5, p. 340),

$$\|Prox(h;u) - Prox(h;\lambda)\| \leq \|u-\lambda\|.$$

Let us put

$$h(z) = -r_k d(z)$$

(h is convex), then

$$Prox(h;\lambda) = \underset{z\in\mathbb{R}^m}{\arg\min}\left\{h(z) + \tfrac{1}{2}\|z-\lambda\|^2\right\} = \underset{z\in\mathbb{R}^m}{\arg\min}\left\{-r_k d(z) + \tfrac{1}{2}\|z-\lambda\|^2\right\}$$

$$= -r_k \underset{z\in\mathbb{R}^m}{\arg\min}\left\{d(z) - \tfrac{1}{2r_k}\|z-\lambda\|^2\right\} = -r_k z_\lambda.$$

It holds that

$$\|Prox(h;u) - Prox(h;\lambda)\| = \|-r_k z_u + r_k z_\lambda\|$$

$$\Longrightarrow r_k\|z_u - z_\lambda\| \le \|u-\lambda\| \Longrightarrow \|z_u - z_\lambda\| \le \tfrac{1}{r_k}\|u-\lambda\|.$$

Let $\overline{\lambda}$ be any K-T vector, then

$$\nabla d_{r_k}(\overline{\lambda}) = 0 \Longrightarrow r_k \nabla d_{r_k}(\overline{\lambda}) = 0 \Longrightarrow z_{\overline{\lambda}} = \overline{\lambda} + r_k\nabla d_{r_k}(\overline{\lambda}) = \overline{\lambda}.$$

Thus

$$\|z_{\lambda^{k+1}} - \overline{\lambda}\| = \|z_{\lambda^{k+1}} - z_{\overline{\lambda}}\| \le \frac{1}{r_k}\left\|\lambda^{k+1} - \overline{\lambda}\right\|.$$

Using the previous expressions, we shall have

$$\left\|\lambda^{k+1} - \overline{\lambda}\right\| = \left\|\lambda^{k+1} - z_{\lambda^k} + z_{\lambda^k} - \overline{\lambda}\right\| \le \left\|\lambda^{k+1} - z_{\lambda^k}\right\| + \|z_{\lambda^k} - \overline{\lambda}\|$$

$$\le \sqrt{2r_k\varepsilon_k} + \tfrac{1}{r_k}\left\|\lambda^k - \overline{\lambda}\right\|.$$

In particular

$$\left\|\lambda^{k+1} - \overline{\lambda}\right\| \le \Phi(r_k,\varepsilon_k) < +\infty.$$

Hence, $\left\{\lambda^k\right\}_k$ is a bounded sequence.

Let $\{\lambda^s\}_s$ be a convergent subsequence to $\overline{\lambda}$, according to the expression (2.21), we have

$$\lim_s(\lambda^{s+1} - z_{\lambda^s}) = 0.$$

We know that

$$z_{\lambda^s} = \lambda^s + r_k\nabla d_{r_k}(\lambda^s)$$

$$\Longrightarrow \lim_s(\lambda^{s+1} - \lambda^s - r_k\nabla d_{r_k}(\lambda^s)) = 0 \Longrightarrow \lim_s \nabla d_{r_k}(\lambda^s) = \nabla d_{r_k}(\overline{\lambda}) = 0.$$

As d_{r_k} is concave, then $\overline{\lambda}$ maximizes d_{r_k}, namely, $\overline{\lambda}$ is a K-T vector.

(**b**) According to the Theorem 2.11, $\{x^s\}_s$ is an asymtotically minimizing sequence of $(\mathcal{P})$. □

2.4 Numerical Experiments

In this paragraph, we propose some numerical experiments illustrating the methods of nondifferentiable convex programming problems that we had studied above and in ([3]). We established a comparative study with the results of ([3]).

Let us call back that the previous methods consist in solving a sequence of unconstrained problems. Every problem of which must be solved by the **Algorithm 4** of ([3]) by making the linear search given by the expression (20) in ([3]).

Example 2.1. *Consider the following mathematical programming problem :*

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

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	$r_k h(x^k)$	$s_k = \Vert g_k\Vert \left\Vert x^{k+1} - x^k\right\Vert$	*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

"-Proximal Penalty method : $(\delta = 10^{-6})$

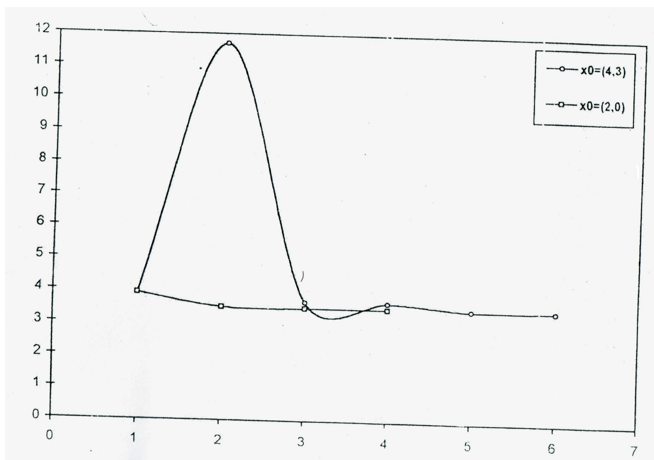

Figure 1: The objective function value at each step

λ_0 *initial*	k	k *total*	x_k	$f(x_k)$	r_k	ε_k	$s = \left\Vert \nabla L_{r_k}(x_k, \lambda_k)\right\Vert$	*time* *s*
5	14	79	$(-0.402, 0.214)$	3.49	14	10^{-14}	$10^{-5}5.0$	0.11
0.5	7	31	$(-0.402, 0.214)$	3.49	7	10^{-7}	$10^{-5}7.0$	0.06
12	15	90	$(-0.402, 0.214)$	3.49	15	10^{-15}	$10^{-5}6.0$	0.11
-1	8	36	$(-0.402, 0.214)$	3.49	8	10^{-8}	$10^{-5}5.0$	0.05
-8	2	13	$(-0.402, 0.214)$	3.49	2	10^{-2}	0.0	0.06

Table 2

augmented Lagrangian method : $(\delta = 10^{-4})$

-

Example 2.2. *Consider the following mathematical programming problem :*

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

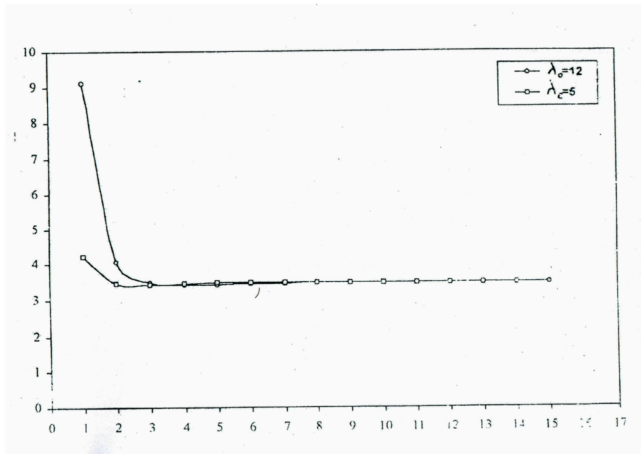

Figure 2: The objective function value at each step

x^0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \|g_k\| \left\|x^{k+1} - x^k\right\|$	*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 3
ε-Proximal Penalty method : $(\delta = 10^{-11})$

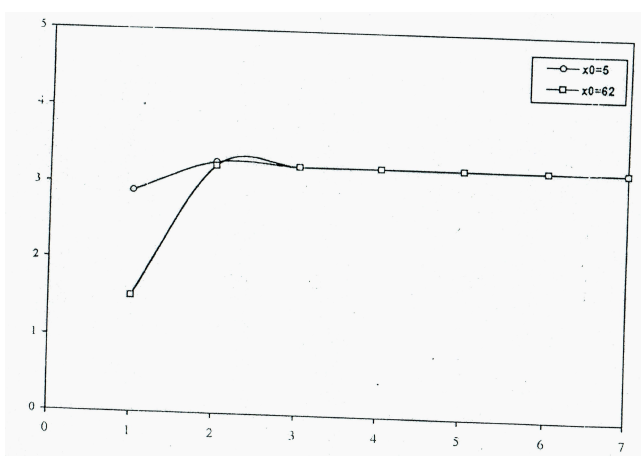

Figure 3: The objective function value at each step

λ_0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$s = \|\nabla L_{r_k}(x_k, \lambda_k)\|$	time s
2	16	47	-1.5	3.25	16	10^{-16}	$10^{-7}7.1$	0.06
20	21	68	-1.5	3.25	21	10^{-21}	$10^{-7}9.1$	0.05
35	22	72	-1.5	3.25	22	10^{-22}	$10^{-7}7.9$	0.05
-1	19	65	-1.5	3.25	19	10^{-19}	$10^{-7}5$	0.06
-5	19	67	-1.5	3.25	19	10^{-19}	$10^{-7}5.3$	0.05

Table 4
augmented Lagrangian method : $(\delta = 10^{-6})$

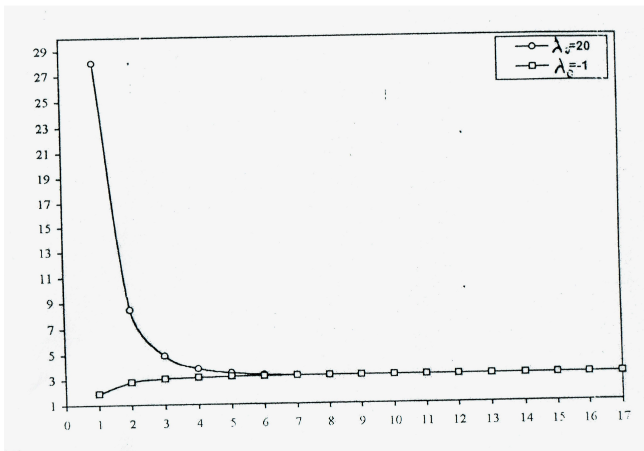

Figure 4: The objective function value at each step

Example 2.3. *Consider the following mathematical programming problem :*

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

where

$$f_1(x) = x_1^2 + x_2^2 - x_2 - x_1 - 1,$$

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

x^0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \|g_k\| \left\|x^{k+1} - x^k\right\|$	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 5
ε- ***Proximal Penalty method*** : $(\delta = 10^{-8})$

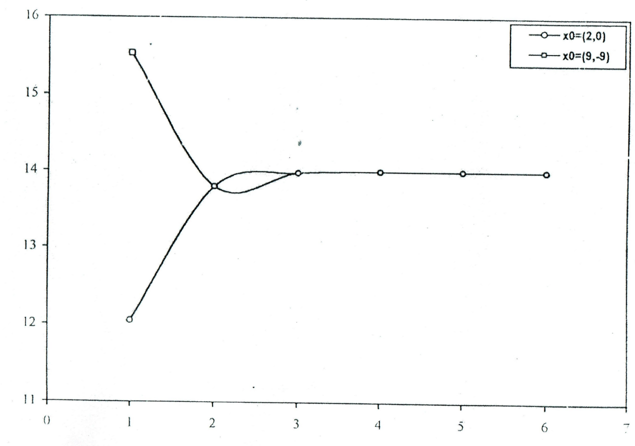

Figure 5: The objective function value at each step

λ_0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$s = \|\nabla L_{r_k}(x_k, \lambda_k)\|$	time s
$(2,6)$	22	280	$(2,-1)$	14	22	10^{-22}	$10^{-5}6.0$	0.16
$(3,0)$	20	220	$(2,-1)$	14	20	10^{-22}	$10^{-5}7.0$	0.11
$(5,3)$	18	185	$(2,-1)$	14	18	10^{18}	$10^{-5}7.0$	0.11
$(-5,-1)$	18	192	$(2,-1)$	14	18	10^{18}	$10^{-5}6.0$	0.11
$(-1,0)$	17	170	$(2,-1)$	14	17	10^{-17}	$10^{-5}8.0$	0.11

Table 6
augmented Lagrangian method : $(\delta = 10^{-4})$

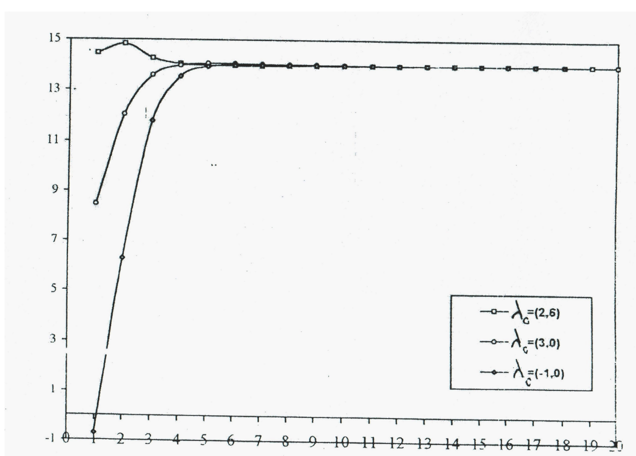

Figure 6: The objective function value at each step

Example 2.4. *Consider the following mathematical programming problem :*

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

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	$r_k h(x^k)$	$s_k = \|g_k\| \left\|x^{k+1} - x^k\right\|$	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 7

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

λ_0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$s = \left\|\nabla L_{r_k}(x_k, \lambda_k)\right\|$	time s
$(3,1)$	11	119	$(-0.5, 0.5)$	0.5	11	10^{-11}	$10^{-5}7$	0.11
$(4,3)$	10	100	$(-0.5, 0.5)$	0.5	10	10^{-10}	$10^{-5}3$	0.11
$(2,5)$	10	134	$(-0.5, 0.5)$	0.5	10	10^{-10}	$10^{-5}8$	0.11
$(-1,0)$	11	126	$(-0.5, 0.5)$	0.5	11	10^{-11}	$10^{-5}3$	0.11
$(-2,-4)$	12	140	$(-0.5, 0.5)$	0.5	12	10^{-12}	$10^{-5}2$	0.11

Table 8

augmented Lagrangian method : $(\delta = 10^{-4})$

Example 2.5. *Consider the following mathematical programming problem :*

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

where

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

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

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

x^0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \|g_k\| \left\|x^{k+1} - x^k\right\|$	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 9

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

λ_0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$s = \left\|\nabla L_{r_k}(x_k, \lambda_k)\right\|$	time s
$(3,2)$	9	57	$(0,-0.5,0)$	2.25	9	10^{-9}	$10^{-6}3.4$	0.06
$(19,2.58)$	9	139	$(0,-0.5,0)$	2.25	9	10^{-9}	$10^{-6}1.8$	0.22
$(4,6)$	9	69	$(0,-0.5,0)$	2.25	9	10^{-9}	$10^{-6}6.9$	0.11
$(-1,-4)$	9	59	$(0,-0.5,0)$	2.25	9	10^{-9}	$10^{-6}4.1$	0.11

Table 10
augmented Lagrangian method *:* $(\delta = 10^{-5})$

Example 2.6. *Consider the following mathematical programming problem :*

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

x^0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$r_k h(x^k)$	$s_k = \|g_k\| \left\|x^{k+1} - x^k\right\|$	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 11
ε-Proximal Penalty method *:* $(\delta = 10^{-11})$

λ_0 initial	k	k total	x_k	$f(x_k)$	r_k	ε_k	$s = \left\|\nabla L_{r_k}(x_k, \lambda_k)\right\|$	time s
3	21	63	-1	4.718	21	10^{-21}	$10^{-7}6.5$	0.06
5	19	57	-1	4.718	19	10^{-19}	$10^{-7}6.9$	0.06
9	23	83	-1	4.718	23	10^{-23}	$10^{-7}5.3$	0.06
1.5	22	77	-1	4.718	22	10^{-22}	$10^{-7}6.3$	0.05
0.6	22	80	-1	4.718	22	10^{-22}	$10^{-7}8.3$	0.06

Table 12
augmented Lagrangian method *:* $(\delta = 10^{-6})$

2.5 Comments and Conclusions

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

1) for the ε-proximal penalty methods, we used the classical penalty functions :

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

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

2) for the augmented Lagrangian method, we use the sequence $(r_k)_k$ such that

$$r_{k+1} = r_k + 1.$$

and for the sequence $(\varepsilon_k)_k$, we make it decrease in the following way :

$$\varepsilon_{k+1} = \frac{\varepsilon_k}{10}.$$

Generally, the obtained solutions are enough precise.

The number of iterations depends, on one hand of the algorithm used to solve the unconstrained subproblems, on the other hand on initial points.

The two previous approaches possess the property of the global convergence.

From a theoretical point of view, both approaches use the proximal regularization. The first one makes the regularity for the subproblems, the other one for the dual function associated with the ordinary Lagrangian. So the idea to return the resolution of primal problem to a sequence of auxiliary problems.

The algorithm that 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.

From point of comparative view, we notice according to the previous numerical experiments that number of necessary iterations to obtain a minimum in the augmented Lagrangian method is higher than counts it of iterations in the ε-proximal penalty method. As well as the run time.

We also notice that the penalty factor is too much large in the ε-proximal penalty method, and enough small in the augmented Lagrangian method.

The stop test in the augmented Lagrangian method is more successful than the stop test in the ε-proximal penalty method.

2.6 General Conclusions

The ε-proximal penalty method is a method of nondifferentiable optimization. It is a member of algorithms whose the generated sequences are asymptotically minimizing. Thus, it is the technique which puts in connection the classical optimization and the asymptotic analysis.

It has advantages for the perturbed problems and in fluid mechanics.

From theoretical point of view, we think that this technique will be widened in problems of positive semidefinite optimization. Thing still is not made and raises open problems in this direction.

The augmented Lagrangien method is a well known technique by its efficiency in the theoretical and practical cases. It applies to differentiable and nondifferentiable optimization problems.

This technique will be widened in positive semidefinite optimization problems with large-sized matrices, thing still is not made and raises open problems still.

3 Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] N. Daili, Some augmented Lagrangian methods applied to nondifferentiable optimization problems, J. Inf. Optim. Sci. 33 (2012), no. 4-5, 487-526. MR3060105.

[2] N. Daili, Applications de Lagrangien Augmenté à des problèmes de programmation convexe non nécessairement différentiables, Pub. 455/P-261, IRMA.- CNRS, Strasbourg I, France, 1986.

[3] N. Daili and Kh. Saadi, ε-Proximal Point Algorithms for Nondifferentiable Convex Optimization and Applications, Adv. Model. Optim. 14 (2012), no. 1, 175-195. MR2910314

[4] A. Gonen and M. Avriel, Duality in Nonlinear Programs using Augmented Lagrangian Function, J. Math. Analys. Appl. 121 (1987), pp. 39-56.

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

[6] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Resea. 1(1976), no. 2, pp. 97-116.

//

Received: December 20, 2015; Accepted: September 27, 2016

UNIVERSITY PRESS

Website: http://www.malayajournal.org/