HAL open science

A regularization method for ill-posed bilevel optimization problems

Maïtine Bergounioux, Mounir Haddou

To cite this version:

Maïtine Bergounioux, Mounir Haddou. A regularization method for ill-posed bilevel optimization problems. RAIRO - Operations Research, 2006, 40, n°1, p. 19-35. 10.1051/ro:2006009 . hal-00021999

HAL Id: hal-00021999

https://hal.science/hal-00021999

Submitted on 31 Mar 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A regularization method for ill-posed bilevel optimization problems

M. Bergounioux & M. Haddou
MAPMO-UMR 6628
Université d'orléans - BP 6759
45067 Orléans cedex 2
maitine.bergounioux, mounir.haddou@univ-orleans.fr

March 31, 2006

Abstract

We present a regularization method to approach a solution of the pessimistic formulation of ill -posed bilevel problems . This allows to overcome the difficulty arising from the non uniqueness of the lower level problems solutions and responses. We prove existence of approximated solutions, give convergence result using Hoffman-like assumptions. We end with objective value error estimates.

ccsd-00021999, version 1 - 31 Mar 2006

1 Introduction

Bilevel programming problems are of growing interest both from theoretical and practical points of view. These models are used in various applications, such as economic planning, network design, and so on... This large class of important and strategic economic problems can be viewed as static noncooperative asymmetric games. Two players seek to optimize their individual objective functions. The first player (the leader) must take into account the reaction (or any possible reaction when non unique) of the second player (the follower). Such a problem can be ill posed since the lower level (follower's problem) may have many solutions and responses for every (or some) fixed leader's variables. In the so-called "optimistic case ", many optimal reactions of the follower are possible and the follower is assumed to choose in favor of the leader. In this case, the upper level problem can be modelled using a bilevel formulation. These programs are quite difficult nonconvex optimization problems. Several theoretical results and heuristics or approximation techniques can be found in the recent literature [6, 10, 11, 5, 7, 12, 14]. In some of these works, strong assumptions are made to simplify the model. The solution of the lower level is supposed to be unique or if they are many, they provide the same (unique) upper level objective value.

The lower level is replaced by the equivalent first order optimality conditions and can be viewed as an equilibrium problem. Most of time, the complementarity part of these

optimality conditions is smoothed or penalized using different techniques.
In our approach we consider the realistic situation where different reactions of the follower are possible. There are multiple responses and we consider the so-called "pessimistic" formulation of the asymmetric game. This can be interpreted as a kind of non-cooperative asymmetric game

Throughout this paper, we shall consider the general bi-level problem:

$$(\mathcal{P}) \begin{cases} \max f(y,x) \\ y \in K \ , \ x \in \mathcal{S}(y) \ , \end{cases}$$

where K and C are non empty convex, closed, bounded subsets of $\mathbb{R}^n$ and

$$\mathcal{S}(y) = \operatorname{argmin} \{ \, h(y,z) \ | z \in C \ \}, \tag{1.1}$$

f and h are smooth functions from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}$. Moreover, for every $y \in K$, $f(y,\cdot)$ and $h(y,\cdot)$ are convex and f takes only positive values (assumptions will be made more precise later).

Remark 1.1 *Since the upper-level objective function f has to be maximized, one can suppose it (without loss of generality) to be positive. Indeed, f can be replaced by $(\max\{f - f(y_0,x_0),0\})^2$ for some $y_0 \in K$ and $x_0 \in \mathcal{S}(y_0)$.*

As mentionned before, the main difficulty comes from the fact that the cost "function" $f(y,x)$, $x \in \mathcal{S}(y)$ can be a **multivalued application** whenever the lower level set is not unique and distinct solutions yield distinct upper-level objective function values. In addition, it is not clear that $f(y,x) = f(y,\tilde{x})$ for any $x, \tilde{x} \in \mathcal{S}(y)$. Therefore, it is difficult to compute the solutions (if there are any).

The remaining of the paper is organized as follows. We introduce the regularization method and give an existence result in next section. Section 3 is devoted to an asymptotic analysis: we prove that the cluster points of solutions to the penalized problems are solutions of a three-level limit problem corresponding to the " pessimistic" formulation of the considered ill-posed bilevel problems . We give some error estimates in the last section.

2 The penalized problem

We would like to let the upper level objective function single valued. So we are going to use a penalization process that allow to compute approximate solutions more easily. More precisely, $\varepsilon > 0$ being given, we consider the following penalized problem

$$(\mathcal{P}_\varepsilon) \begin{cases} \max f(y,x) \\ y \in K \ , \ x \in \mathcal{S}_\varepsilon(y) \ , \end{cases}$$

where

$$\mathcal{S}_\varepsilon(y) = \operatorname{argmin}\{\, h_\varepsilon(y,z) \;|z \in C\ \}, \tag{2.1}$$

where

$$h_\varepsilon(y) = h(y,z) + \varepsilon f^2(y,z)\ . \tag{2.2}$$

For each nonnegative ε, the bi-level problem $(\mathcal{P}_\varepsilon)$ is well posed . Furthermore, under some general and non restrictive assumptions on f and h we will prove that the upper level function is single valued and continuous with respect to the leader variables y.

This regularization technique makes some selection property on the solutions of the lower level problem which is easy to characterize and have an explicit and simple economic interpretation. In almost all other regularization methods, the lower level is replaced by its optimality conditions. The bi-level problem is then considered as a mathematical program with equilibrium constraints. The "hard" part of these constraints (namely the complementarity conditions) is then smoothed or penalized. In fact these methods make also some selection (the generated sequences converge to the analytic center when using smoothing methods or the least norm center) on the solution set of the lower level but these selections do not have any economic interpretation since they have no link to the objective function of the upper level. Moreover, convergence results need more restrictive assumptions.

For convenience of the reader, we first give or recall some direct and classical results. These results will be useful for forthcoming developpements.

Lemma 2.1 *For any $\varepsilon > 0$, the lower-level problem*

$$\mathcal{Q}_{\varepsilon,y} = \begin{cases} \min h_\varepsilon(y,z) \\ z \in C\ , \end{cases}$$

admits (at least) a solution so that $\mathcal{S}_\varepsilon(y) \neq \emptyset$. Moreover, there exists a constant $\kappa_y \in \mathbb{R}$ such that

$$\forall x \in \mathcal{S}_\varepsilon(y) \qquad f(y,x) = \kappa_y\ .$$

Proof - The existence of a solution to $\mathcal{Q}_{\varepsilon,y}$ is obvious since C is bounded and $f,\ h$ are continuous. Moreover $\mathcal{Q}_{\varepsilon,y}$ may be written as follows

$$\mathcal{Q}^*_{\varepsilon,y} \quad \begin{cases} \min h(y,z) + \varepsilon t^2 \\ f(y,z) - t = 0\ , \\ z \in C\ , \end{cases}$$

Since f takes only positive values, the equality constraint can be obviously replaced by an inequality in this minimization problem

$$\mathcal{Q}^*_{\varepsilon,y} \quad \begin{cases} \min h(y,z) + \varepsilon t^2 \\ f(y,z) - t \leq 0 \ , \\ z \in C \ . \end{cases}$$

$\mathcal{Q}^*_{\varepsilon,y}$ is a convex problem and the cost function is strictly convex with respect to t. This simple observation proves that the optimal value of t is unique and completes the proof. □

Lemma 2.2 *Let be $\varepsilon > 0$ fixed. The multi-application $\mathcal{S}_\varepsilon$ is lower semi-continuous in the following sense : if $y_k \to y$ and $x_k \in \mathcal{S}_\varepsilon(y_k)$ then $x_k \to x \in \mathcal{S}_\varepsilon(y)$ (up to a subsequence).*

Proof - Let be $x_k \in \mathcal{S}_\varepsilon(y_k) \subset C$. As C is bounded, then (x_k) is bounded as well and converges to some x (up to a subsequence). As $x_k \in \mathcal{S}_\varepsilon(y_k)$ we get

$$\forall z \in C \qquad h(y_k, x_k) + \varepsilon f^2(y_k, x_k) \leq h(y_k, z) + \varepsilon f^2(y_k, z) \ .$$

As f and h are continuous with respect to y and x we obtain

$$\forall z \in C \qquad h(y, x) + \varepsilon f^2(y, x) \leq h(y, z) + \varepsilon f^2(y, z) \ ,$$

that is $x \in \mathcal{S}_\varepsilon(y)$. □

Lemma 2.3 *Let be $\varepsilon > 0$ fixed. The cost function*

$$v_\varepsilon : y \mapsto \{ f(y,x) \mid x \in \mathcal{S}_\varepsilon(y) \}$$

is single-valued and continuous.

Proof - We see that the function v_ε is single valued, with Lemma 2.1. Let us prove the continuity: let be (y_k) a sequence that converges to some y. Then $v_\varepsilon(y_k) = f(y_k, x_k)$ where $x_k \in \mathcal{S}_\varepsilon(y_k)$. Lemma 2.2 yields that x_k converges (up to a subsequence) to $x \in \mathcal{S}_\varepsilon(y)$. As f is continuous with respect to y and x we get

$$v_\varepsilon(y_k) = f(y_k, x_k) \to f(y, x) = v_\varepsilon(y) \ .$$

□

We may now give an existence result :

Theorem 2.1 *For any $\varepsilon > 0$, problem $(\mathcal{P}_\varepsilon)$ admits at least an optimal solution y_ε.*

Proof - As v_ε is continuous and K is bounded, the result follows. □

3 Asymptotic results

3.1 A convergence result for the solutions of $(\mathcal{P}_\varepsilon)$

In this subsection, we study the behaviour of solutions of $(\mathcal{P}_\varepsilon)$ as ε goes to 0. First, we introduce some notations:

$$\widetilde{\mathcal{S}}(y) = \operatorname{argmin}\{f^2(y,z) \ |z \in \mathcal{S}(y) \ \}, \tag{3.1}$$

where $\mathcal{S}(y)$ is given by (1.1) and

$$(\widetilde{\mathcal{P}}) \ \begin{cases} \max f(y,x) \\ y \in K \ , \ x \in \widetilde{\mathcal{S}}(y) \ , \end{cases} \tag{3.2}$$

Note that problem $(\widetilde{\mathcal{P}})$ is a three-level problem that can be written in an extended way as follows:

$$(\widetilde{\mathcal{P}}) \ \begin{cases} \max f(y,x) \\ y \in K \\ x \in \operatorname{argmin}\left\{ f^2(y,z) \ |z \in \operatorname{argmin}\{ \, h(y,w) \ |w \in C \ \}\right\} , \end{cases}$$

Lemma 3.1 *$x \in \widetilde{\mathcal{S}}(y)$ is equivalent to*

$$x \in \mathcal{S}(y) \text{ and } \forall z \in C \text{ such that } h(y,z) = h(y,x), \qquad f^2(y,z) \geq f^2(y,x) \ .$$

Proof - Assume that z satisfies $h(y,z) = h(y,x)$ with $x \in \operatorname{argmin} \{ \, h(y,t) \mid t \in C \, \}$. Then $z \in \operatorname{argmin} \{ \, h(y,t) \mid t \in C \, \}$. □

Lemma 3.2 *Let y be fixed. If $x_\varepsilon \in \mathcal{S}_\varepsilon$ converges to some $\bar{x}$, then $\bar{x} \in \widetilde{\mathcal{S}}(y)$*

Proof - Assume $x_\varepsilon \in \mathcal{S}_\varepsilon$ and $x_\varepsilon \to \bar{x}$ as $\varepsilon \to 0$. For every $z \in C$ we get

$$h(y,x_\varepsilon) + \varepsilon f^2(y,x_\varepsilon) \leq h(y,z) + \varepsilon f^2(y,z) \ .$$

When $\varepsilon \to 0$, as the functions are continuous we obtain

$$\forall z \in C \qquad h(y,\bar{x}) \leq h(y,z) \ ,$$

that is $\bar{x} \in \mathcal{S}(y)$.
Let be $\tilde{x} \in C$ such that $h(y,\tilde{x}) = h(y,\bar{x})$. Then

$$\begin{array}{rcll} h(y,x_\varepsilon) + \varepsilon f^2(y,x_\varepsilon) & \leq & h(y,\tilde{x}) + \varepsilon f^2(y,\tilde{x}) & \text{since } \tilde{x} \in C \\ & \leq & h(y,\bar{x}) + \varepsilon f^2(y,\tilde{x}) & \text{since } h(y,\tilde{x}) = h(y,\bar{x}) \\ & \leq & h(y,x_\varepsilon) + \varepsilon f^2(y,\tilde{x}) & \text{since } x_\varepsilon \in C \text{ and } \bar{x} \in \mathcal{S}(y) \ . \end{array}$$

Therefore

$$\forall \tilde{x} \in C \text{ such that } h(y,\tilde{x}) = h(y,\bar{x}),\ f^2(y,x_\varepsilon) \leq f^2(y,\tilde{x}) \ .$$

Passing to the limit with the continuity of f gives

$$\forall \tilde{x} \in C \text{ such that } h(y,\tilde{x}) = h(y,\bar{x}),\ f^2(y,\bar{x}) \leq f^2(y,\tilde{x}) \ .$$

With Lemma 3.1 we conclude that $x \in \widetilde{\mathcal{S}}(y)$. □

To establish the main convergence result of this work, we will use some technical but not so very restrictive assumption.
Let us set

$$\alpha_\varepsilon = h(y_\varepsilon, x_\varepsilon) + o(\varepsilon) \ , \tag{3.3}$$

and

$$\Lambda_\varepsilon = \{\, x \in C \,|\, h(y_\varepsilon, x) \leq \alpha_\varepsilon \,\} \ . \tag{3.4}$$

Assume we can find $\sigma_o > 0$ and $\varepsilon_o > 0$ such that

$$\forall \varepsilon \leq \varepsilon_o \quad \inf_{h(y_\varepsilon,x)=\alpha_\varepsilon} |\nabla_x h(y_\varepsilon, x)| \geq \sigma_o \ , \tag{3.5}$$

This assumption does not seem quite natural at a first glimpse. In fact it a Hoffman-inequality type assumption which is more or less standard in this context. The proof of next theorem, and especially the proof of Lemma 3.3 below will make this hypothesis clear.

Theorem 3.1 *Assume condition (3.5) is verified and let y_ε an optimal solution to $(\mathcal{P}_\varepsilon)$. Then y_ε converges to some $\bar{y}$ (up to a subsequence) and $\bar{y}$ is an optimal solution to $(\widetilde{\mathcal{P}})$.*

Proof - Let y_ε an optimal solution to $(\mathcal{P}_\varepsilon)$. Then $y_\varepsilon \in K$ which is bounded. So (extracting a subsequence) we may assert that y_ε converges to $\bar{y}$. As K is closed then $\bar{y} \in K$. As y_ε is an optimal solution to $(\mathcal{P}_\varepsilon)$ we have

$$\forall \tilde{y} \in K \ , \forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(\tilde{y}) \qquad f(y_\varepsilon, x_\varepsilon) \geq f(\tilde{y}, \tilde{x}_\varepsilon) \tag{3.6}$$

where $x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$. Note that $\tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(\tilde{y})$ implies that $\tilde{x}_\varepsilon \in C$. So $\tilde{x}_\varepsilon$ is bounded and converges to $\tilde{x}$ (up to a subsequence) with $\tilde{x} \in \widetilde{\mathcal{S}}(\tilde{y})$ (Lemma 3.2).
Passing to the limit in (3.6) gives

$$\forall \tilde{y} \in K \ , \exists \tilde{x} \in \widetilde{\mathcal{S}}(\tilde{y}) \text{ such that } f(\bar{y}, \bar{x}) \geq f(\tilde{y}, \tilde{x}) \ ,$$

where $\bar{x}$ is the limit (of a subsequence) of x_ε. Now we need the following result to achieve the proof :

Lemma 3.3 *Assume that (3.5) is satisfied and let* $(y_\varepsilon, x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon))$ *converging to* $(\bar{y}, \bar{x})$. *Then* $\bar{x} \in \widetilde{\mathcal{S}}(\bar{y})$.

Thanks to the definition of $\widetilde{\mathcal{S}}(\tilde{y})$ we note that $f(\tilde{y}, \cdot)$ is constant on $\widetilde{\mathcal{S}}(\tilde{y})$, namely

$$\forall z \in \widetilde{\mathcal{S}}(\tilde{y}) \qquad f(\tilde{y}, z) = f(\tilde{y}, \tilde{x}) \ .$$

Finally

$$\forall \tilde{y} \in K, \ \forall \tilde{x} \in \widetilde{\mathcal{S}}(\tilde{y}) \qquad f(\bar{y}, \bar{x}) \geq f(\tilde{y}, \tilde{x}) \ ,$$

with $\bar{x} \in \widetilde{\mathcal{S}}(\bar{y})$. This means that $\bar{y}$ is an optimal solution to $(\widetilde{\mathcal{P}})$.

• It remains to prove Lemma 3.3.

Let y_ε converging to $\bar{y}$ and $x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$. As $x_\varepsilon \in C$ (bounded) one may extract a subsequence converging to $\bar{x}$. We are going to prove that $\bar{x} \in \widetilde{\mathcal{S}}(\bar{y})$.

We first prove that $\bar{x} \in \mathcal{S}(\bar{y})$. As $x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$ we have

$$\forall z \in C \qquad h(y_\varepsilon, x_\varepsilon) + \varepsilon\, f^2(y_\varepsilon, x_\varepsilon) \leq h(y_\varepsilon, z) + \varepsilon\, f^2(y_\varepsilon, z) \ ; \tag{3.7}$$

as f and g are continuous, passing to the limit gives

$$\forall z \in C \qquad h(\bar{y}, \bar{x}) \leq h(\bar{y}, z) \ ,$$

that is $\bar{x} \in \mathcal{S}(\bar{y})$.

Let $\tilde{x} \in \mathcal{S}(\bar{y})$. Suppose for a while that $\exists \tilde{\varepsilon}$ such that

$$\forall \varepsilon \leq \tilde{\varepsilon} \qquad \tilde{x} \in \Lambda_\varepsilon \ . \tag{3.8}$$

We get

$$h(y_\varepsilon, \tilde{x}) \leq h(y_\varepsilon, x_\varepsilon) + o(\varepsilon) \ ;$$

with relation (3.7) this gives

$$\forall z \in C \qquad h(y_\varepsilon, \tilde{x}) + \varepsilon\, f^2(y_\varepsilon, x_\varepsilon) \leq h(y_\varepsilon, z) + \varepsilon\, f^2(y_\varepsilon, z) + o(\varepsilon) \ . \tag{3.9}$$

As $\tilde{x} \in C$ relation (3.7) yields as well

$$h(y_\varepsilon, x_\varepsilon) + \varepsilon\, f^2(y_\varepsilon, x_\varepsilon) \leq h(y_\varepsilon, \tilde{x}) + \varepsilon\, f^2(y_\varepsilon, \tilde{x}) \ .$$

Adding these two relations gives

$$\forall z \in C \qquad h(y_\varepsilon, x_\varepsilon) + 2\varepsilon\, f^2(y_\varepsilon, x_\varepsilon) \leq h(y_\varepsilon, z) + \varepsilon\, f^2(y_\varepsilon, z) + \varepsilon\, f^2(y_\varepsilon, \tilde{x}) + o(\varepsilon) \ ; \tag{3.10}$$

the choice of $z = x_\varepsilon$ implies

$$\varepsilon\, f^2(y_\varepsilon, x_\varepsilon) \leq \varepsilon\, f^2(y_\varepsilon, \tilde{x}) + o(\varepsilon) \ ,$$

that is

$$f^2(y_\varepsilon, x_\varepsilon) \leq f^2(y_\varepsilon, \tilde{x}) + \frac{o(\varepsilon)}{\varepsilon} \ .$$

Passing to the limit gives finally

$$\forall \tilde{x} \in \mathcal{S}(\bar{y}) \qquad f^2(\bar{y}, \bar{x}) \leq f^2(\bar{y}, \tilde{x}) \ .$$

This means that $\bar{x} \in \widetilde{\mathcal{S}}(\bar{y})$.

Unfortunately, there is no reason for "assumption " (3.8) to be satisfied and we must get rid of it. We are going to adapt the previous proof (we gave the main ideas). If $\tilde{x} \notin \Lambda_\varepsilon$ then we perform a projection: we call $\tilde{x}_\varepsilon$ the projection of $\tilde{x}$ on Λ_ε. We are going to show that $\tilde{x}_\varepsilon$ converges to $\tilde{x}$.

As $\tilde{x} \notin \Lambda_\varepsilon$ we get $\alpha_\varepsilon < h(y_\varepsilon, \tilde{x})$. Let us call $\sigma_{\alpha_\varepsilon}(h)$ the following real number

$$\sigma_{\alpha_\varepsilon}(h) = \inf_{x \in [\alpha_\varepsilon < h(y_\varepsilon, \cdot)]} \frac{h(y_\varepsilon, x) - \alpha_\varepsilon}{d(x, \Lambda_\varepsilon)} \ , \tag{3.11}$$

where $d(x, \Lambda_\varepsilon)$ is the distance between x and Λ_ε and

$$[\alpha_\varepsilon < h(y_\varepsilon, \cdot)] = \{\ x \in \mathbb{R}^n \mid \alpha_\varepsilon < h(y_\varepsilon, x)\} \ .$$

This so called Hoffman constant can be defined following for instance Azé and Corvellec [2]. Therefore

$$h(y_\varepsilon, \tilde{x}) - \alpha_\varepsilon \geq d(\tilde{x}, \Lambda_\varepsilon)\, \sigma_{\alpha_\varepsilon}(h) \ .$$

As $d(\tilde{x}, \Lambda_\varepsilon) = d(\tilde{x}, \tilde{x}_\varepsilon)$ we obtain

$$d(\tilde{x}, \tilde{x}_\varepsilon) \leq \frac{h(y_\varepsilon, \tilde{x}) - \alpha_\varepsilon}{\sigma_{\alpha_\varepsilon}(h)} \ .$$

We have to estimate $\sigma_{\alpha_\varepsilon}(h)$. In particular we look for $\sigma_o > 0$ such that

$$\forall \varepsilon \qquad \sigma_{\alpha_\varepsilon}(h) \geq \sigma_o \ .$$

In [2], it is shown that

$$\sigma_{\alpha_\varepsilon}(h) \geq \inf_{h(y_\varepsilon, x) = \alpha_\varepsilon} |\nabla_x h(y_\varepsilon, x)| \ ,$$

where $|\nabla_x h(y_\varepsilon, x)|$ stands for the strong slope of h at (y_ε, x) with respect to x ([2]); the strong-slope of a function φ at x is defined as

$$|\nabla \varphi(x)| := \begin{cases} 0 & \text{if } x \text{ is a local minimum of } \varphi \ , \\ \limsup\limits_{y \to x} \dfrac{\varphi(x) - \varphi(y)}{d(x, y)} & \text{otherwise} \end{cases}$$

Using (3.5), we have

$$d(\tilde{x}, \tilde{x}_\varepsilon) \leq \frac{h(y_\varepsilon, \tilde{x}) - \alpha_\varepsilon}{\sigma_o} = \frac{h(y_\varepsilon, \tilde{x}) - h(y_\varepsilon, x_\varepsilon) + o(\varepsilon)}{\sigma_o} \to 0 \ .$$

Indeed $y_\varepsilon \to \bar{y}$, $x_\varepsilon \to \bar{x}$, h is continuous and $h(\bar{y}, \bar{x}) = h(\bar{y}, \tilde{x})$.
We may now end the proof. We can use relation (3.10) with $\tilde{x}_\varepsilon$ instead of $\tilde{x}$ so that

$$\forall z \in C \qquad h(y_\varepsilon, x_\varepsilon) + 2\varepsilon\, f^2(y_\varepsilon, x_\varepsilon) \leq h(y_\varepsilon, z) + \varepsilon\, f^2(y_\varepsilon, z) + \varepsilon\, f^2(y_\varepsilon, \tilde{x}_\varepsilon) + o(\varepsilon) \ ;$$

we choose $z = x_\varepsilon$ once again to get

$$f^2(y_\varepsilon, x_\varepsilon) \leq f^2(y_\varepsilon, \tilde{x}_\varepsilon) + \frac{o(\varepsilon)}{\varepsilon} \ .$$

Passing to the limit as $\varepsilon \to 0$ gives (for every $\tilde{x} \in \mathcal{S}(\bar{y})$

$$f^2(\bar{y}, \bar{x}) \leq f^2(\bar{y}, \tilde{x}) \ .$$

This means that $\bar{x} \in \widetilde{\mathcal{S}}(\bar{y})$. □

Remark 3.1 *It is clear that assumption* (3.5) *is satisfied if h is linear ("linear" case). Next problem is to find simple conditions for $(\bar{y}, \bar{x})$ to get* (3.5) *when h is not linear. One hint is to assume that h is $\mathcal{C}^1$ and that $\|\nabla_x h(\bar{y}, \bar{x}\| \neq 0$; then the strong slope $|\nabla_x h(y_\varepsilon, x)|$ coincides with the norm $\|\nabla_x h(y_\varepsilon, x)\|$ of the gradient of h with respect to x. With the convergence of $(y_\varepsilon, x_\varepsilon)$ to $(\bar{y}, \bar{x})$ (up to a subsequence), there exist ε_o and $\eta > 0$ such that*

$$\forall \varepsilon \leq \varepsilon_o \qquad \|\nabla_x h(y_\varepsilon, x_\varepsilon)\| \geq \eta > 0 \ ;$$

next we have to prove that $\|\nabla_x h(y_\varepsilon, x)\| \geq \eta$ for any x such that $h(y_\varepsilon, x) = \alpha_\varepsilon$. A good tool could be an "local inversion theorem" for the multivalued case but it is not obvious. The problem is still open. We have the same challenge in next section.

3.2 Comparison of $(\mathcal{P})$ and $(\widetilde{\mathcal{P}})$

Now, it is clear that a solution of the penalized problem $(\mathcal{P}_\varepsilon)$ is a good approximation of a solution of $(\widetilde{\mathcal{P}})$. Anyway, it is not a solution (a priori) of the problem in consideration $(\mathcal{P})$. So we have to compare $(\mathcal{P})$ and $(\widetilde{\mathcal{P}})$.

The second level of $(\widetilde{\mathcal{P}})$ clearly disappears when the initial problem lower level solutions set corresponds to the same revenue for each value of y (or are unique). In this case $(\mathcal{P})$ and $(\widetilde{\mathcal{P}})$ are equivalent. In other cases, the solution of $(\widetilde{\mathcal{P}})$ corresponds to some "optimal worst" case solution.

This solution is still important for the decision makers of the upper level problem.

Remark 3.2 *Using the same regularization technique, if we replace $\varepsilon f^2(y,z)$ by $-\varepsilon f^2(y,z)$ in the definition of h_ε, we will obtain (at the limit) an optimal solution of $(\mathcal{P})$ which corresponds to an optimal best case solution of our asymmetric game .*

4 Error estimates

The purpose of this section is to study the behavior of $f(y^*,x^*) - f(y_\varepsilon,x_\varepsilon)$ as $\varepsilon \to 0$ and provide (if possible) some error estimates. Since the penalized problems are nonconvex, we can not use any classical perturbation result. We proceed in two steps : we first prove some monotonicity results for the upper level objective function values and then consider classical perturbation analysis results for some auxilliary convex problems.

4.1 Preliminary results

Lemma 4.1 *Let be $\varepsilon > \varepsilon' > 0$ and $y \in K$. Let be $x_\varepsilon \in \mathcal{S}_\varepsilon(y)$ and $\tilde{x} \in \mathcal{S}_{\varepsilon'}(y)$. Then we get*

$$f^2(y,x_\varepsilon) \leq f^2(y,\tilde{x}) \ .$$

Proof - Let us fix $\varepsilon > \varepsilon' > 0$ and choose some $y \in K$. Let be $x_\varepsilon \in \mathcal{S}_\varepsilon(y)$ and $\tilde{x} \in \mathcal{S}_{\varepsilon'}(y)$. Assume that

$$f^2(y,\tilde{x}) < f^2(y,x_\varepsilon) \ . \tag{4.1}$$

As $\tilde{x} \in \mathcal{S}_{\varepsilon'}(y)$ and $x_\varepsilon \in C$, we have

$$h(y,\tilde{x}) + \varepsilon' f^2(y,\tilde{x}) \leq h(y,x_\varepsilon) + \varepsilon' f^2(y,x_\varepsilon) \ ,$$

$$h(y,\tilde{x}) + \varepsilon' f^2(y,\tilde{x}) + (\varepsilon-\varepsilon')f^2(y,\tilde{x}) \leq h(y,x_\varepsilon) + \varepsilon' f^2(y,x_\varepsilon) + (\varepsilon-\varepsilon')f^2(y,\tilde{x})$$

With (4.1) and $\varepsilon > \varepsilon' > 0$, we obtain

$$h(y,\tilde{x}) + \varepsilon f^2(y,\tilde{x}) < h(y,x_\varepsilon) + \varepsilon' f^2(y,x_\varepsilon) + (\varepsilon-\varepsilon')f^2(y,x_\varepsilon) <= h(y,x_\varepsilon) + \varepsilon f^2(y,x_\varepsilon)$$

So

$$h(y,\tilde{x}) + \varepsilon f^2(y,\tilde{x}) < \min \{ \ h(y,x) + \varepsilon f^2(y,x), \ x \in C\}$$

and we get a contradiction. □

Lemma 4.2 *Let be $\varepsilon > \varepsilon' > 0$ and y_ε (respectively $y_{\varepsilon'}$) a solution to $(\mathcal{P}_\varepsilon)$ (respectively $(\mathcal{P}_{\varepsilon'})$). Let be $x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$ and $x_{\varepsilon'} \in \mathcal{S}_{\varepsilon'}(y_{\varepsilon'})$. Then*

$$f^2(y_\varepsilon,x_\varepsilon) \leq f^2(y_{\varepsilon'},x_{\varepsilon'}) \leq f^2(y^*,x^*) \ ,$$

where y^* is a solution to $(\widetilde{\mathcal{P}})$ with $x^* \in \mathcal{S}(y^*)$.

Proof - Using Lemma 4.1 with $y = y_\varepsilon$ and $x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$ gives

$$\forall \tilde{x} \in \mathcal{S}_{\varepsilon'}(y_\varepsilon) \qquad f^2(y_\varepsilon, x_\varepsilon) \leq f^2(y_\varepsilon, \tilde{x}) \ . \tag{4.2}$$

As $y_{\varepsilon'}$ is a solution of $(\mathcal{P}_{\varepsilon'})$ we get

$$\forall y \in K, \ \forall x \in \mathcal{S}_{\varepsilon'}(y) \qquad f(y_{\varepsilon'}, x_{\varepsilon'}) \geq f(y, x) \ .$$

We may choose in particular $y = y_\varepsilon$ and $x = \tilde{x} \in \mathcal{S}_{\varepsilon'}(y_\varepsilon)$ to get

$$\forall \tilde{x} \in \mathcal{S}_{\varepsilon'}(y_\varepsilon) \qquad f(y_{\varepsilon'}, x_{\varepsilon'}) \geq f(y_\varepsilon, \tilde{x}) \ . \tag{4.3}$$

As f is assumed to be nonnegative we finally obtain

$$f(y_\varepsilon, x_\varepsilon) \leq f(y_\varepsilon, \tilde{x}) \leq f(y_{\varepsilon'}, x_{\varepsilon'}) \ .$$

Therefore the family ($f(y_\varepsilon, x_\varepsilon)$ is increasing). The convergence of $f(y_\varepsilon, x_\varepsilon)$ to $f(y^*, x^*)$ (f is continuous) achieves the proof since $f(y^*, x^*)$ is the limit and the upper bound of the family $(f(y_\varepsilon, x_\varepsilon))$. □

Lemma 4.3 *Let be $\varepsilon > 0$ and $\tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$ where y^* is a solution to $(\widetilde{\mathcal{P}})$. Then*

$$\forall x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon) \qquad f(y^*, \tilde{x}_\varepsilon) \leq f(y_\varepsilon, x_\varepsilon) \leq f(y^*, x^*) \ . \tag{4.4}$$

Proof - This is a direct consequence of Lemma 4.2 : the relation $f(y_\varepsilon, x_\varepsilon) \leq f(y^*, x^*)$ is obvious and the relation $f(y^*, \tilde{x}_\varepsilon) \leq f(y_\varepsilon, x_\varepsilon)$ comes from the fact that y_ε is a solution to $(\mathcal{P}_\varepsilon)$. □

Remark 4.1 *The previous lemmas show that it is sufficient to study $f(y^*, x^*) - f(y^*, \tilde{x}_\varepsilon)$ for some $\tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$.*

For a large class of realistic problems, the lower level is linearly constrained. moreover, we use for our analysis some local error bounds and these bounds are very complicated in case of nonlinear constraints. So, we assume from now on that C is polyhedral :

$$C = \{ \ x \in \mathbb{R}^n \ \mid \ Ax = b, \ x \geq 0 \ \} \ ,$$

where A is a $m \times n$ real matrix and $b \in \mathbb{R}^m$.

In the sequel y^* is a solution to $(\widetilde{\mathcal{P}})$ (which existence is given by Theorem 3.1) and $x^* \in \widetilde{\mathcal{S}}(y^*)$ (see (3.1)) so that

$$x^* \in \ \text{argmin} \ \{ \ f^2(y^*, z) \mid z \in \ \text{argmin} \ \{h(y^*, \zeta) \ , \zeta \in C\} \ \} \ .$$

Let us denote

$$\alpha^* = h(y^*, x^*) \text{ and } \beta^* = f(y^*, x^*) \ . \tag{4.5}$$

Note that β^* is the optimal value for $(\widetilde{\mathcal{P}})$ (the upper level) so that we may assume that $\beta^* \neq 0$ (otherwise the problem is trivial). We set

$$C^* = \{ \ x \in C \ \mid \ h(y^*, x) \leq \alpha^* \text{ and } f(y^*, x) \leq \beta^* \ \} \ . \tag{4.6}$$

Let us give an important property of C^* :

Proposition 4.1 *Assume y^* is a solution to $(\widetilde{\mathcal{P}})$ and C^* is defined with (4.6), then*

$$C^* = \{ \ x \in C \ \mid \ h(y^*, x) = \alpha^* \ \textit{and} \ f(y^*, x) = \beta^* \ \}$$

and

$$C^* = \{ \ x \in C \ \mid \ h(y^*, x) + f(y^*, x) \leq \sigma^* \overset{def}{:=} \alpha^* + \beta^* \ \} \ .$$

Proof - Note that it impossible to get $h(y^*, x) \leq \alpha^*$, if $x \in C^*$. Indeed, as $x^* \in \widetilde{\mathcal{S}}(y^*)$ then $x^* \in S(y^*) =$ argmin $\{h(y^*, \zeta) \ , \zeta \in C\}$. Therefore :

$$\forall \zeta \in C \qquad h(y^*, x^*) \leq h(y^*, \zeta) \ . \tag{4.7}$$

Setting $\zeta = x \in C^*$ gives

$$\alpha^* = h(y^*, x^*) \leq h(y^*, x) \leq \alpha^* \ .$$

So

$$\forall x \in C^* \qquad h(y^*, x) = \alpha^* \ .$$

The same remark holds for β^* so that

$$C^* = \{ \ x \in C \ \mid \ h(y^*, x) = \alpha^* \text{ and } f(y^*, x) = \beta^* \ \} \ . \tag{4.8}$$

Let us call

$$C' = \{ \ x \in C \ \mid \ h(y^*, x) + f(y^*, x) \leq \sigma^* \ \} \ .$$

It is obvious that $C^* \subset C'$. Conversely, let be $x \in C'$. Relation (4.7) yields $\alpha^* \leq h(y^*, x)$ so that

$$\alpha^* + f(y^*, x) \leq \alpha^* + \beta^* \ .$$

This gives $f(y^*, x) \leq \beta^*$. Similarly, we get $h(y^*, x) \leq \alpha^*$ and $x \in C^*$. □

The main point is now to estimate (roughly speaking) the distance between the solution x^* and $\mathcal{S}_\varepsilon(y^*)$. As $x^* \in C^*$ and C^* is defined with inequalities, we first assume a Hoffman-type condition.

4.2 Error estimates under an Hoffman hypothesis

Following Azé and Corvellec [2] we know that

$$\inf_{[\sigma^* < f(y^*,\cdot)+h(y^*,\cdot)]} |\nabla_x (f(y^*,\cdot) + h(y^*,\cdot))| \leq \inf_{x\in[\sigma^* < f(y^*,\cdot)+h(y^*,\cdot)]} \frac{f(y^*,x) + h(y^*,x) - \sigma^*}{d(x, [f(y^*,\cdot) + h(y^*,\cdot) \leq \sigma^*]} .$$

The notation $[\sigma^* < f(y^*,\cdot) + h(y^*,\cdot)]$ stands for the set

$$\{x \in \mathbb{R}^n \mid \sigma^* < f(y^*,x) + h(y^*,x) \} .$$

We note that $[f(y^*,\cdot) + h(y^*,\cdot) \leq \sigma^*] = C^*$. In this subsection, we assume the following

$$(\mathcal{H}_1) \qquad \gamma^* := \inf_{[\sigma^* < f(y^*,\cdot)+h(y^*,\cdot)]} |\nabla_x (f(y^*,\cdot) + h(y^*,\cdot))| > 0.$$

Let us call $\gamma = \dfrac{1}{\gamma^*}$: assumption $(\mathcal{H}_1)$ implies that

$$\forall \varepsilon > 0,\ \forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad \exists x^*_\varepsilon \in C^* \text{ s.t. } \|\tilde{x}_\varepsilon - x^*_\varepsilon\| \leq \gamma\, [f(y^*,\tilde{x}_\varepsilon) + h(y^*,\tilde{x}_\varepsilon) - \alpha^* - \beta^*] \ . \tag{4.9}$$

Note also that relation (4.4) of Lemma 4.3 yields that

$$\forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad f(y^*,\tilde{x}_\varepsilon) \leq \beta^*$$

and

$$h(y^*,\tilde{x}_\varepsilon) \leq \alpha^* + \varepsilon\beta^*$$

because of the definition of $\mathcal{S}_\varepsilon(y^*)$.Therefore

$$\forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad f(y^*,\tilde{x}_\varepsilon) + h(y^*,\tilde{x}_\varepsilon) - \alpha^* - \beta^* \leq \varepsilon$$

and

$$\forall \varepsilon > 0,\ \forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad \exists x^*_\varepsilon \in C^* \text{ s.t. } \|\tilde{x}_\varepsilon - x^*_\varepsilon\| \leq \gamma\varepsilon \ . \tag{4.10}$$

The existence of such Lipschitzian error bound for convex or general inequalities is, itself, an interesting domain of research. It is strongly related to metric regularity properties. A large number of conditions and characterizations can be found in [2, 3, 8, 9, 13, 15, 16]. This list of references constitutes a small but significant part of the existent literature.

Remark 4.2 *1. Assumption $(\mathcal{H}_1)$ is fulfilled if the functions f and h are linear with respect to x. Indeed they cannot be identically equal to 0 and the strong slope coincides with the norm of gradient which is a positive constant.*
2. x^*_ε is the projection of $\tilde{x}_\varepsilon$ on C^*.

Lemma 4.4 *Both $\tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$ and x^*_ε given by (4.10) converge to $x^* \in \mathcal{S}(y^*)$ as $\varepsilon \to 0$.*

Proof - We know that $\tilde{x}_\varepsilon \to x^*$ (with the previous results). Let us set $d_\varepsilon = \dfrac{\tilde{x}_\varepsilon - x^*_\varepsilon}{\varepsilon}$. As d_ε is bounded (by γ) it clear that x^*_ε and $\tilde{x}_\varepsilon$ have the same limit point (namely x^*). □

In what follows $\tilde{x}_\varepsilon$ is an element of $\mathcal{S}_\varepsilon(y^*)$ and x^*_ε is the associated element given by (4.10) .

Let us define

$$I(x^*) = \{i \in \{1, \cdots, n\} \ \mid \ x^*_i = 0 \ \} \ , \text{ and } \tilde{C} = \{ \ d \in \mathbb{R}^n \ \mid \ Ad = 0 \ , \ d_{|I(x^*)} \geq 0 \ \} \ .$$

Let d be in $\tilde{C}$.

Then, there exists $\varepsilon_d > 0$ such that $\forall \varepsilon < \varepsilon_o, \ x^*_\varepsilon + \varepsilon d \in C$. Indeed

- $A(x^*_\varepsilon + \varepsilon d) = A(x^*_\varepsilon) + \varepsilon Ad = A(x^*_\varepsilon) = b$.
- If $i \in I(x^*)$, then $(x^*_\varepsilon + \varepsilon d)_i \geq x^*_{\varepsilon,i} \geq 0$.
- If $i \notin I(x^*)$, then $x^*_i > 0$. As $x^*_\varepsilon \to x^*, \exists \varepsilon_i$ such that $x^*_{\varepsilon,i} > 0$ forall $\varepsilon \leq \varepsilon_i$. Then we choose $\eta = \inf\limits_{i \notin I(x^*)} \{\varepsilon_i\}$ so that

$$\forall \varepsilon \leq \eta \qquad x^*_{\varepsilon,i} > 0 \ .$$

  Now choosing $\varepsilon_d \leq \eta$ small enough we get $(x^*_\varepsilon + \varepsilon d)_i \geq 0$ for any $\varepsilon \leq \varepsilon_o$.

As $\tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$ and $x^*_\varepsilon + \varepsilon d \in C$ we have

$$h(y^*, \tilde{x}_\varepsilon) + \varepsilon f^2(y^*, \tilde{x}_\varepsilon) \leq h(y^*, x^*_\varepsilon + \varepsilon d) + \varepsilon f^2(y^*, x^*_\varepsilon + \varepsilon d) \ ,$$

$$h(y^*, \tilde{x}_\varepsilon) - h(y^*, x^*_\varepsilon + \varepsilon d) + \varepsilon \left[f^2(y^*, \tilde{x}_\varepsilon) - f^2(y^*, x^*_\varepsilon + \varepsilon d) \right] \leq 0 \ .$$

As the functions are $\mathcal{C}^1$, we have

$$h(y^*, \tilde{x}_\varepsilon) = h(y^*, x^*_\varepsilon) + \nabla_x h(y^*, x^*_\varepsilon) \cdot (\tilde{x}_\varepsilon - x^*_\varepsilon) + (\tilde{x}_\varepsilon - x^*_\varepsilon) o(\tilde{x}_\varepsilon - x^*_\varepsilon)$$

$$h(y^*, \tilde{x}_\varepsilon) = h(y^*, x^*_\varepsilon) + \varepsilon \nabla_x h(y^*, x^*_\varepsilon) \cdot d_\varepsilon + \varepsilon d_\varepsilon \, o(\varepsilon d_\varepsilon) \ , \tag{4.11}$$

and

$$h(y^*, x^*_\varepsilon + \varepsilon d) = h(y^*, x^*_\varepsilon) + \varepsilon \nabla_x h(y^*, x^*_\varepsilon) \cdot d + \varepsilon d \, o(\varepsilon d) \ , \tag{4.12}$$

where $\nabla_x h$ stands for the derivative of h with respect to x. As $x^*_\varepsilon \in C^*$ and $\tilde{x}_\varepsilon \in C$ then

$$h(y^*, x^*_\varepsilon) = \alpha^* = h(y^*, x^*) \leq h(y^*, \tilde{x}_\varepsilon) \ .$$

With relation (4.11) this gives

$$\nabla_x h(y^*, x_\varepsilon^*) \cdot d_\varepsilon + d_\varepsilon\, o(\varepsilon d_\varepsilon) = \frac{h(y^*, \tilde{x}_\varepsilon) - h(y^*, x_\varepsilon^*)}{\varepsilon} \geq 0 \ .$$

As d_ε is bounded (by γ), there exist cluster points; passing to the limit gives

$$\nabla_x h(y^*, x^*) \cdot \tilde{d} = \lim_{\varepsilon \to 0} \nabla_x h(y^*, x_\varepsilon^*) \cdot d_\varepsilon \geq 0 \ , \tag{4.13}$$

for any cluster point $\tilde{d}$ of the family d_ε.

In addition, we obtain with (4.11) and (4.12)

$$\varepsilon \nabla_x h(y^*, x_\varepsilon^*) \cdot d_\varepsilon + \varepsilon d_\varepsilon\, o(\varepsilon d_\varepsilon) - \varepsilon \nabla_x h(y^*, x_\varepsilon^*) \cdot d - \varepsilon d\, o(\varepsilon d) + \varepsilon \left[f^2(y^*, \tilde{x}_\varepsilon) - f^2(y^*, x_\varepsilon^* + \varepsilon d) \right] \leq 0 \ ,$$

that is

$$\nabla_x h(y^*, x_\varepsilon^*) \cdot (d_\varepsilon - d) + d_\varepsilon\, o(\varepsilon d_\varepsilon) - d\, o(\varepsilon d) + \left[f^2(y^*, \tilde{x}_\varepsilon) - f^2(y^*, x_\varepsilon^* + \varepsilon d) \right] \leq 0 \ .$$

Passing to the limit (with Lemma 4.4) we obtain

$$\nabla_x h(y^*, x^*) \cdot (\tilde{d} - d) \leq 0 \ , \tag{4.14}$$

where $\tilde{d}$ is a cluster point of the sequence d_ε and any $d \in \tilde{C}$. As $d = 0$ belongs to $\tilde{C}$, we get

$$\nabla_x h(y^*, x^*) \cdot \tilde{d} \leq 0 \ .$$

Finally, we obtain with (4.13)

$$\nabla_x h(y^*, x^*) \cdot \tilde{d} = \lim_{\varepsilon \to 0} \nabla_x h(y^*, x_\varepsilon^*) \cdot \frac{\tilde{x}_\varepsilon - x_\varepsilon^*}{\varepsilon} = 0 \ . \tag{4.15}$$

This means that

$$\nabla_x h(y^*, x_\varepsilon^*) \cdot (\tilde{x}_\varepsilon - x_\varepsilon^*) = o(\varepsilon).$$

As

$$h(y^*, \tilde{x}_\varepsilon) = h(y^*, x_\varepsilon^*) - \nabla_x h(y^*, x_\varepsilon^*) \cdot (x_\varepsilon^* - \tilde{x}_\varepsilon) + (x_\varepsilon^* - \tilde{x}_\varepsilon)\, o(x_\varepsilon^* - \tilde{x}_\varepsilon)$$

we get

$$h(y^*, \tilde{x}_\varepsilon) - h(y^*, x_\varepsilon^*) = o(\varepsilon) - \varepsilon d_\varepsilon\, o(\varepsilon d_\varepsilon) = o(\varepsilon) \ .$$

As $x_\varepsilon^* \in C^*$ then $h(y^*, x_\varepsilon^*) = \alpha^*$ and

$$\forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad h(y^*, \tilde{x}_\varepsilon) = h(y^*, x^*) + o(\varepsilon) \ . \tag{4.16}$$

As h and f^2 play similar roles we have the same result for f^2. More precisely

$$\forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad f^2(y^*, \tilde{x}_\varepsilon) - f^2(y^*, x^*) = o(\varepsilon) \ . \tag{4.17}$$

We just proved the following result

Theorem 4.1 *Assume that* $(\mathcal{H}_1)$ *is satisfied ; let* y_ε *be a solution to* $(\mathcal{P}_\varepsilon)$ *and* $\tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$. *Then*

$$h(y^*, \tilde{x}_\varepsilon) - h(y^*, x^*) = o(\varepsilon) \text{ and } f^2(y^*, \tilde{x}_\varepsilon) - f^2(y^*, x^*) = o(\varepsilon) \text{ as } \varepsilon \to 0 \ .$$

Moreover

$$\forall x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon) \qquad f(y^*, x^*) - f(y_\varepsilon, x_\varepsilon) = o(\varepsilon) \text{ as } \varepsilon \to 0 \ .$$

Proof - The first assertion has been proved : relations (4.16) and (4.17) . We use relation (4.4) and the previous result to claim that

$$f^2(y^*, x^*) - f^2(y_\varepsilon, x_\varepsilon) = o(\varepsilon) \ .$$

As $f^2(y^*, x^*) - f^2(y_\varepsilon, x_\varepsilon) = [f(y^*, x^*) + f(y_\varepsilon, x_\varepsilon)]\,[f(y^*, x^*) - f(y_\varepsilon, x_\varepsilon)]$ and $f(y^*, x^*) + f(y_\varepsilon, x_\varepsilon) \to 2f(y^*, x^*) = 2\beta^*$ we get the result since $\beta^* \neq 0$. □

With a bootstrapping technique we obtain the following corollary ;

Corollary 4.1 *Under the assumptions and notations of the previous theorem, we get* $\forall x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$

$$\forall n \in \mathbb{N} \qquad f(y^*, x^*) - f(y_\varepsilon, x_\varepsilon) = o(\varepsilon^n)$$

and $\forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$

$$h(y^*, \tilde{x}_\varepsilon) - h(y^*, x^*) = o(\varepsilon^n) \ .$$

Proof - Using relations (4.16) and (4.17) in assumption $(\mathcal{H}_1)$ we see that relation (4.10) becomes

$$\forall \varepsilon > 0,\ \forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*) \qquad \exists x^*_\varepsilon \in C^* \text{ s.t. } \|\tilde{x}_\varepsilon - x^*_\varepsilon\| \leq \gamma o(\varepsilon) \ . \tag{4.18}$$

Using the same technique leads to relations (4.16) and (4.17) with ε^2 instead of ε and so on. □

Remark 4.3 *These error estimates are still valid (with the same proofs) when the penalized problems are approximatively solved.*

4.3 Error estimates under a "second-order" assumption

If assumption $(\mathcal{H}_1)$ is not ensured, one may, however, give estimates using the following hypothesis

$$(\mathcal{H}_2) \qquad \begin{cases} \exists \varepsilon_o > 0 \ , \exists \delta > 0, \ \text{ such that } \forall x \in C^* + \mathcal{B}(0, \varepsilon_o) \\ \exists \tilde{x} \in C^* \text{ such that } \|x - \tilde{x}\|^2 \leq \delta \left[(h(y^*, x) - \alpha^*)^+ + (f(y^*, x) - \beta^*)^+ \right] \end{cases}$$

Remark 4.4 *$(\mathcal{H}_2)$ means that C^* is H-metrically regular (of the second order). (See the definition of this regularity property for example in [1] Def. 4.3.2). $(\mathcal{H}_2)$ also corresponds to a quadratic growth condition [4] Def.3.1 .*
This assumption is significantly weaker than $(\mathcal{H}_1)$ and covers a large class of problems since it is satisfied when $h(y^*, .) + f(y^*, .)$ is linear or quadratic.

We have a rather similar result which proof is the the same as in the previous subsection (so that we do not detail it) :

Theorem 4.2 *Assume that $(\mathcal{H}_2)$ is satisfied ; let y_ε be a solution to $(\mathcal{P}_\varepsilon)$ and $x_\varepsilon \in \mathcal{S}_\varepsilon(y_\varepsilon)$. Then*

$$f(y^*, x^*) - f(y_\varepsilon, x_\varepsilon) = o(\sqrt{\varepsilon}) \text{ as } \varepsilon \to 0 \text{ , so that}$$

$$\forall \tau > 0 \qquad f(y^*, x^*) - f(y_\varepsilon, x_\varepsilon) = o(\varepsilon^{1-\tau}) \ .$$

Moreover, $\forall \tilde{x}_\varepsilon \in \mathcal{S}_\varepsilon(y^*)$

$$\forall \tau > 0 \qquad h(y^*, \tilde{x}_\varepsilon) - h(y^*, x^*) = o(\varepsilon^{1-\tau}) \ .$$

5 Conclusion

With this new penalization approach, the pessimistic formulation of general bi-level problems becomes, in some way, tractable. Indeed, instead of the complicated limit problem, we only need to solve approximately the penalized one for a small value of the parameter ε. We have given error estimates that prove that this approximation is reasonable even when Hoffman's assumption is not satisfied .

References

[1] A. Auslender and M. Teboulle. Asymptotic cones and functions in optimization and variational inequalities. *Springer Monographs in Mathematics*, Springer-Verlag, New York, 2003.

[2] D. Azé and J.N . Corvellec. Characterizations of error bounds for lower semicontinuous functions on metric spaces. *ESAIM Control Optim. Calc. Var.*, 10 (2004), no. 3, 409–425.

[3] D. Azé and A. Rahmouni. On primal-dual stability in convex optimization. *J. Convex Anal.*, 3 (1996), no. 2, 309–327.

[4] J.F. Bonnans and A. Shapiro. Perturbation analysis of optimization problems. *Springer Series in Operations Research*, Springer-Verlag, New York, 2000.

[5] L. Brotcorne, M. Labbé, P. Marcotte and G. Savard. A bilevel model for toll optimization on a multicommodity transportation network, *Transportation Science*, 35 (2001) 1-14

[6] S. Dempe Foundations of bilevel programming. *Nonconvex optimization and its applications*, Kluwer Acad. Publ., Dordrecht, 2002.

[7] M. Labbé, P. Marcotte and G. Savard. On a class of bilevel programs. *Nonlinear optimization and related topics*, (Erice, 1998), 183–206, Appl. Optim., 36, Kluwer Acad. Publ., Dordrecht, 2000.

[8] A.S. Lewis and J-S. Pang. Error bounds for convex inequality systems. *Generalized convexity, generalized monotonicity: recent results (Luminy, 1996)*, 75–110, Nonconvex Optim. Appl., 27, Kluwer Acad. Publ., Dordrecht, 1998.

[9] W. Li. Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. *SIAM Journal of Optimization*, 7(4):966–978, 1997.

[10] P. Loridan and J. Morgan New results on approximate solutions in two-level optimization *Optimization*, 20 (1989), no. 6, 819–836.

[11] P. Loridan and J. Morgan Regularizations for two-level optimization problems. *Advances in optimization*,239–255, Lecture Notes in Econom. and Math. Systems, 382, Springer, Berlin, 1992.

[12] P. Marcotte et G. Savard. A bilevel programming approach to Price Setting in: Decision and Control in Management Science. *Essays in Honor of Alain Haurie* Georges Zaccour ed., Kluwer Academic Publishers (2002) 97 - 117

[13] K.F. Ng and X.Y. Zheng. Global error bounds with fractional exponents. *Mathematical programming*, Ser B(88):357–370, 2000.

[14] T. Q. Nguyen, M. Bouhtou and J.-L. Lutton. D.C. approach to bilevel bilinear programming problem: application in telecommunication pricing. *Optimization and optimal control (Ulaanbaatar, 2002)*, 211–231, Ser. Comput. Oper. Res., 1, World Sci. Publishing, River Edge, NJ, 2003.

[15] Z. Wu and J. Ye. On error bounds for lower semicontinuous functions. *Mathematical Programming*, Ser. A, 2002.

[16] J. Zhao. The lower semicontinuity of optimal solution sets. *Journal of Mathematical Analysis and Applications*, 207:240–254, 1997.