M2AN. MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS - MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

JOSEPH FRÉDÉRIC BONNANS

Théorie de la pénalisation exacte

M2AN. Mathematical modelling and numerical analysis - Modélisation mathématique et analyse numérique, tome 24, nº 2 (1990), p. 197-210

<http://www.numdam.org/item?id=M2AN_1990__24_2_197_0>

MATHEMATICAL MODELLING AND NUMERICAL ANALYSIS
MODÉLISATION MATHÉMATIQUE ET ANALYSE NUMÉRIQUE

(Vol. 24, n° 2, 1990, p. 197 à 210)

THÉORIE DE LA PÉNALISATION EXACTE (*)

Joseph Frédéric BONNANS [1]

Communiqué par J. CEA

Résumé. — *Depuis les travaux de Pschenichnyi et Han montrant comment leur utilisation permet de globaliser certains algorithmes, les fonctionnelles de pénalisation exacte jouent en optimisation un rôle croissant. L'objet de cet article est de rappeler les principaux résultats en améliorant certains d'entre eux. Nous étudions en particulier la théorie de la normalité et l'utilisation de conditions du deuxième ordre faibles.*

Abstract. — *Since the work of Pschenichnyi and Han concerning their use in the globalization of algorithms, exact penalty functions have in the field of optimization an increasing importance. The purpose of this paper is to review the main results and to improve some of them. We focus on the normality theory and on weak second-order sufficient conditions.*

1. INTRODUCTION

Nous considérons des problèmes d'optimisation du type

$$\text{(P)} \quad \min f(x)\,; \quad g_i(x) \leqq 0\,,\ i \in I\,,\ g_j(x) = 0\,,\ j \in J\,,$$

où I et J sont des ensembles finis, les applications f et g_i $(i \in I \cup J)$ étant de $\mathbb{R}^n$ dans $\mathbb{R}$. Étant donné $K \in I \cup J$, on posera

$$g_K(x) = \{g_i(x)\,; i \in K\}\ .$$

La partie positive z^+ d'un vecteur z étant définie en prenant la partie positive de chaque composante, nous appellerons fonctionnelle pénalisée exacte une fonction du type

$$\theta_r(x) = f(x) + r\left\|(g_I(x)^+, g_J(x))\right\|\ ,$$

(*) Reçu en 1987.

[1] INRIA, Rocquencourt, 78153 Le Chesnay, France.

où la norme ci-dessus est quelconque, le paramètre $r > 0$ étant appelé coefficient de pénalisation. La dénomination de θ_r vient de ce que, dans un assez grand nombre de cas, une solution locale de (P) réalise (pour r assez grand) un minimum local de θ_r. L'intérêt envers les fonctionnelles de pénalisation exacte s'est considérablement développé après que Pschenichnyi (voir [19]) et Han [11] aient montré que son utilisation comme critère pour la recherche linéaire permet de globaliser un certain nombre d'algorithmes (nous renvoyons à Bertsekas [4] et Bonnans et Gabay [7] pour une vue d'ensemble sur ces algorithmes).

L'objet de cet article est de présenter de façon synthétique les principaux résultats connus ainsi que quelques améliorations de ceux-ci. La nécessité des hypothèses sera justifiée autant que possible à l'aide de contre-exemples.

Parmi les aspects importants de la théorie de la pénalisation exacte qui ne sont pas évoqués ici, il faut citer notamment le rapport entre la pénalisabilité et la stabilité du coût optimal lorsque les contraintes sont perturbées ; cet aspect est traité dans Clarke [10] (voir aussi Bertsekas [4], chap. 4). On trouve aussi dans Bertsekas [4] l'étude des rapports entre les points stationnaires de (P) et ceux de $\theta_r(x)$.

Par ailleurs les résultats présentés ici peuvent être partiellement étendus au lagrangien augmenté non différentiable étudié dans Bonnans [5] :

$$\psi_r(\lambda, x) = f(x) + \lambda^t g(x) + r\|(g_I(x)^+, g_J(x))\| = \theta_r(x) + \lambda^t g(x),$$

dont l'utilité apparaît lors de l'étude des algorithmes (et qui peut également se relier à la stabilité du coût optimal).

L'organisation de cet article est comme suit : la Section II introduit certains concepts et notations et présente quelques résultats élémentaires. La Section 3 reprend la théorie de la normalité de Ioffe [14]. Nous introduisons un concept de régularité valable pour des fonctions non régulières et montrons que la normalité implique la régularité. Nous montrons aussi que la normalité implique la pénalisabilité en relaxant les conditions de Ioffe. Nous rappelons que la condition de Mangasarian et Fromovitz entraîne la normalité, et montrons aussi que c'est le cas si les contraintes sont linéaires. Dans la Section 4, nous montrons que l'augmentabilité et certaines conditions faibles du deuxième ordre entraînent la pénalisabilité, avec une estimation précise sur la valeur du coefficient de pénalisation.

2. NOTATIONS ET RÉSULTATS GÉNÉRAUX

Introduisons tout d'abord quelques notations et concepts qui seront utiles dans la suite. Si x est un *point admissible* de (P) (c'est-à-dire qu'il vérifie les contraintes de (P)), et si f et g sont différentiables nous dirons que λ est un *multiplicateur de Lagrange* associé à x si (x, λ) vérifie les relations suivantes, dites *conditions d'optimalité du premier ordre* :

$$\text{(CO1)} \quad \begin{array}{ll} \text{(i)} & \nabla f(x) + \nabla g(x)\,\lambda = 0\,, \\ \text{(ii)} & g_I(x) \leqq 0\,, \quad \lambda_I \geqq 0\,, \quad \lambda_I^t\, g_I(x) = 0\,, \quad g_J(x) = 0\,. \end{array}$$

Nous désignerons le *Lagrangien* associé au problème par

$$L(x, \lambda) = f(x) + \lambda^t g(x)\,,$$

et par $|K|$ le cardinal d'un ensemble K. Soit une norme $\|\cdot\|$; la norme duale notée $\|\cdot\|_*$ est définie par

$$\|z\|_* = \max \{z^t y\,;\ \|y\| \leqq 1\}\,.$$

Une conséquence immédiate de la définition est l'égalité de Cauchy généralisée

$$z^t y \leqq \|z\|_*\, \|y\| \quad \text{pour tout } z, y\,.$$

Nous dirons que le problème (P) est convexe si les fonctions f et g_i, $i \in I$, sont convexes et si g_J est affine.

THÉORÈME 1 : *Soit $\bar{x}$ un point admissible de* (P) *et $r > 0$ tels que la fonction $\theta_r(x)$ admette un minimum local (strict) en $\bar{x}$. Alors $\bar{x}$ est une solution (stricte) de* (P) *et, si f et g sont Gâteaux différentiables en $\bar{x}$, il existe au moins un multiplicateur de Lagrange λ associé à $\bar{x}$. Si de plus les gradients des contraintes actives en $\bar{x}$ sont linéairement indépendants, on a*

$$\|\lambda\|_* \leqq r\,.$$

Démonstration : Pour tout point admissible x, $f(x)$ est égale à $\theta_r(x)$ et un minimum local (strict) de θ_r admissible pour (P) est donc a fortiori minimum local (strict) de (P). Or si f et g sont Gâteaux différentiables en $\bar{x}$, θ_r admet des dérivées directionnelles qui sont donc positives en $\bar{x}$ et (supposant sans perte de généralité que $g_I(\bar{x}) = 0$) vérifient donc :

$$\theta_r'(\bar{x}, d) = \nabla f(\bar{x})^t d + r \left\| \left((\nabla g_I(\bar{x})^t d)^+ , \nabla g_J(x)^t d \right) \right\| \geqq 0\,,$$

d'où en particulier

$$\nabla g_I(\bar{x})^t d \leqq 0 \quad \text{et} \quad \nabla g_J(\bar{x})^t d = 0 \Rightarrow \nabla f(\bar{x})^t d \geqq 0\,,$$

ce qui signifie que $\bar{d} = 0$ est solution du programme linéaire

$$\min_d \nabla f(\bar{x})^t d \; ; \nabla g_I(\bar{x})^t d \leqq 0, \nabla g_J(\bar{x})^t d = 0 \, ,$$

d'où l'existence de multiplicateurs. Par ailleurs, $\bar{x}$ est a fortiori minimum local de $f(x) + r\|g(x)\|$, et donc l'application convexe correspondant aux dérivées directionnelles en $\bar{x}$:

$$d \to \nabla f(\bar{x})^t d + r \|\nabla g(\bar{x})^t d\|$$

admet un minimum en 0. La fonction norme étant continue, une application des règles de calcul sous-différentiel (voir Rockafellar [21]), donne l'existence de μ tel que $\|\mu\|_* \leqq r$ et

$$\nabla f(\bar{x}) + \nabla g(\bar{x}) \mu = 0 \, .$$

Mais si les gradients des contraintes actives sont linéairement indépendants, μ est nécessairement le multiplicateur de Lagrange associé à $\bar{x}$. □

Dans le cas de problèmes d'optimisation convexes, on obtient facilement des caractérisations optimales du coefficient de pénalisation.

THÉORÈME 2 : *On suppose f et g Gâteaux-différentiables et le problème* (P) *convexe. Soit $\bar{x}$ solution de* (P) *et Λ l'ensemble des multiplicateurs de Lagrange* (*indépendant de $\bar{x}$*). *Alors* (P) *est pénalisable de façon exacte si et seulement si $\Lambda \neq \varnothing$, la constante r pouvant être prise égale à*

$$r_0 = \min \left\{ \|\lambda\|_* \, ; \lambda \in \Lambda \right\} . \quad \square$$

Démonstration : D'après le théorème 1, si $\Lambda = \varnothing$, alors (P) ne peut être pénalisable. Supposons au contraire que $\Lambda \neq \varnothing$. On vérifie aisément que Λ est fermé. Le problème

$$\min \|\lambda\|_* , \quad \lambda \in \Lambda \, ,$$

a donc au moins une solution $\bar{\lambda}$. Soit $r \geqq \|\bar{\lambda}\|_*$. Il vient, puisque λ_I est positif :

$$\begin{aligned} \theta_r(x) - L(x, \bar{\lambda}) &= \|(g(x)^+, g_J(x))\| - \bar{\lambda}^t g(x) \, , \\ &\geqq r \|(g_I(x)^+, g_J(x))\| - \bar{\lambda}^t (g_I(x)^+, g_J(x)) \, , \\ &\geqq (r - \|\bar{\lambda}\|_*) \|(g_I(x)^+, g_J(x))\| \geqq 0 \, . \end{aligned}$$

Nous utilisons maintenant l'hypothèse de convexité. L'égalité (CO1(i)) exprime la nullité de $\partial L(x, \bar{\lambda})/\partial x$ en $\bar{x}$. Mais la convexité de (P) et la positivité de $\bar{\lambda}_I$ impliquent la convexité par rapport à x du lagrangien, d'où

$$\theta_r(\bar{x}) = f(\bar{x}) = L(\bar{x}, \bar{\lambda}) \leqq L(x, \bar{\lambda}) , \quad \forall x \in \mathbb{R}^n ,$$

d'où le résultat. □

Exemple 1 : Considérons le problème ($n = 1$, $\alpha \in \mathbb{R}$) :

$$\min x^2 + \alpha x \; ; x^2 \leqq 0 .$$

Ce problème est convexe ; $\bar{x} = 0$ est le seul point admissible, donc est solution unique. Si $\alpha \neq 0$, il n'existe pas de multiplicateur de Lagrange et on peut vérifier directement que le problème n'est pas pénalisable. Si $\alpha = 0$, alors $\Lambda = \mathbb{R}$ et $\theta_0(x) = x^2$ a bien un minimum en $\bar{x}$.

Remarque 1 : Si le problème est convexe mais non différentiable, on peut étendre aisément le résultat du théorème 2, en utilisant les sous-différentiels au lieu des gradients ou en définissant un multiplicateur de Lagrange comme un vecteur $\bar{\lambda}$ tel que $\bar{\lambda}_I \geqq 0$ et que $L(x, \bar{\lambda})$ soit minimum en $\bar{x}$. □

3. PÉNALISABILITÉ, NORMALITÉ, RÉGULARITÉ

Nous reprenons ici et développons la théorie de la normalité de Ioffe [14]. La distance d'un point $x \in \mathbb{R}^n$ à un ensemble X de $\mathbb{R}^n$ étant notée dist et définie par

$$\operatorname{dist}(x, X) = \inf \{ \|x - y\| \; ; y \in X \} ,$$

posons

$$X_{\text{ad}} = \left\{ x \in \mathbb{R}^n \; ; g_I(x) \leqq 0, g_J(x) = 0 \right\} .$$

Nous dirons que les contraintes du problème (P) sont normales au point admissible $\bar{x}$ s'il existe un voisinage de V de $\bar{x}$ et $M > 0$ tels que :

$$\| (g_I(x)^+ , g_J(x)) \| \geqq M \operatorname{dist}(x, X_{\text{ad}}) , \quad \text{pour tout } x \in V .$$

Cette notion est indépendante de la norme ; elle est également invariante par difféomorphisme dans $\mathbb{R}^n$ (changement de variable C^1 d'inverse C^1).

Éclairons la liaison entre la normalité et une notion très générale de régularité. Nous dirons que les contraintes de (P) sont régulières au point admissible $\bar{x}$ si, pour tout d de $\mathbb{R}^n$ admissible au premier ordre en $\bar{x}$, c'est-à-dire vérifiant, pour $\rho > 0$:

$$\| (g_I(\bar{x} + \rho d)^+ , g_J(\bar{x} + \rho d)) \| = o(\rho) ,$$

il existe un chemin $x(\rho)$ $(\rho > 0)$ tel que $x(0) = \bar{x}$, $x'(0) = d$ et $x(\rho) \in X_{\mathrm{ad}}$. Si g est Gâteaux différentiable en $\bar{x}$, l'admissibilité au premier ordre équivaut à

$$\nabla g_i(\bar{x})^t \, d \leqq 0 \text{ si } g_i(\bar{x}) = 0, i \in I \,, \quad \text{et} \quad \nabla g_J(\bar{x})^t \, d = 0 \,,$$

et on retrouve alors la notion habituelle de régularité. On trouvera dans Penot [17] une synthèse récente des conditions suffisantes de régularité dans le cas différentiable.

PROPOSITION 1 : *La normalité des contraintes en $\bar{x}$ entraîne leur régularité.* □

Démonstration : Soit d admissible au premier ordre en $\bar{x}$. La normalité des contraintes entraîne l'existence de $x(\rho)$ $(\rho > 0)$ dans X_{ad} avec

$$\| \bar{x} + \rho d - x(\rho) \| \leqq M^{-1} \| (g_I(x + \rho d)^+ , g_J(x + \rho d)) \| = o(\rho) ,$$

et le chemin $x(\rho)$ vérifie donc la condition de régularité. □

Nous dirons que f est lipschitzienne près de $\bar{x}$ s'il existe un voisinage V de $\bar{x}$ et $L > 0$ tel que

$$| f(y) - f(x) | \leqq L \| y - x \| \quad \text{pour tout } x, y \text{ dans } V \,.$$

Le résultat ci-dessous est similaire à celui de Ioffe [14] à ceci près que nous ne supposons pas les contraintes lipschitziennes.

THÉORÈME 3 : *Soit $\bar{x}$ une solution locale (stricte) de* (P) *vérifiant :*

(i) *Les contraintes de* (P) *sont continues et normales en $\bar{x}$,*

(ii) *le critère est lipschitzien près de $\bar{x}$.*

Alors le problème (P) *est (strictement) pénalisable en $\bar{x}$.* □

Démonstration : Soit $\varepsilon > 0$ tel que

$$f(\bar{x}) \leqq f(y) \text{ si } y \text{ est admissible et si } \| y - \bar{x} \| < \varepsilon \,.$$

D'après (i), pour ε' assez petit, il existe $M > 0$ tel que pour tout x vérifiant $\| x - \bar{x} \| < \varepsilon'$, il existe $y \in X_{\mathrm{ad}}$ vérifiant

$$\| y - x \| \leqq M \| (g_I(x)^+ , g_J(x)) \| \,.$$

La continuité de g entraîne que, si on restreint x à une boule assez petite centrée en $\bar{x}$, on aura $\|y - \bar{x}\| < \varepsilon$. Utilisant l'optimalité locale (stricte) de $\bar{x}$, (ii) et les inégalités ci-dessus, il vient (L étant la constante de Lipschitz) pour tout $L' \geqq L$:

$$\begin{aligned} f(\bar{x}) &\leqq f(y)\,, \\ &\leqq f(x) + L' \|y - x\|\,, \\ &\leqq f(x) + L' M \|(g_I(x)^+, g_I(x))\|\,, \\ &= \theta_{L'M}(x)\,. \end{aligned}$$

Puisque $\theta_{L'M}(\bar{x}) = f(\bar{x})$, $\theta_{L'M}$ admet un minimum local en $\bar{x}$. Si de plus $\bar{x}$ est une solution locale stricte de (P), soit $x \neq \bar{x}$, voisin de $\bar{x}$. Si x est admissible, il est clair que $\theta_{L'M}(x) > \theta_{L'M}(\bar{x})$. Sinon, $y \neq x$ et la seconde inégalité ci-dessus est stricte pour tout $L' > L$; donc $\theta_{L'M}$ admet un minimum local strict en $\bar{x}$ pour $L' > L$. □

Exemple 2 : Prenons $n = 2$, $I = \{1\}$, $J = \emptyset$ et $g_1(x) = (x_1)^4 - (x_2)^2$. Alors

$$X_{\text{ad}} = \{x \in \mathbb{R}^2 \,;\, x_1^2 \leqq |x_2|\}\,.$$

En $\bar{x} = \{0, 0\}$, la contrainte est régulière. Toutefois, avec $d = (1, 0)$, il vient $g_1(x + \rho d)^+ = (\rho)^4$ et dist $(\bar{x} + \rho d, X_{\text{ad}}) \approx \rho^2$. Les contraintes sont donc régulières mais non normales en $\bar{x}$.

Soit le critère $f(x) = |x_2| - (x_1)^2$. Alors $f(x) \geqq 0$ pour tout x de X_{ad} et $\bar{x} = 0$ est donc solution globale. Vérifions que le problème n'est pas pénalisable : on a ici (la norme des contraintes étant la valeur absolue)

$$\theta_r(x) = |x_2| - (x_1)^2 + r((x_1)^4 - (x_2)^2)^+\,,$$

donc, pour $\rho \in \mathbb{R}$, $\theta_r(\rho, 0) = -\rho^2 + r\rho^4$ ne peut admettre de minimum en 0 pour aucune valeur de r. □

Il est utile de pouvoir disposer de critères vérifiables assurant la normalité des contraintes. L'exemple ci-dessus et la proposition 1 montrent que la régularité est une condition nécessaire mais non suffisante de normalité. Nous rappelons un critère de normalité bien connu.

DÉFINITION 1 : *Les contraintes de* (P) *vérifient la condition de Mangasarian et Fromovitz* [16] *en* $\bar{x}$ *si* g *est* C^1 *et*

(MF)

(i) $\nabla g_J(\bar{x})$ *est surjectif,*

(ii) *Il existe* d *dans* Ker $\nabla g_J(\bar{x})$ *tel que* $\nabla g_i(\bar{x})^t d < 0$ *si* $g_i(\bar{x}) = 0, i \in I$.

THÉORÈME 4 (voir Robinson [20], Ioffe [14]) : *Si $\bar{x}$ est un point admissible de* (P) *où les contraintes vérifient la condition de Mangasarian et Fromovitz, alors les contraintes de* (P) *sont normales en $\bar{x}$.* □

La propriété de normalité est vérifiée dans le cas important où les contraintes sont affines (rappelons que l'affinité des contraintes n'implique pas (MF)). Nous en donnons une démonstration élémentaire.

THÉORÈME 5 : *Si l'application g est affine, les contraintes sont normales en tout point admissible.* □

La clé de la démonstration est donnée dans le lemme suivant (analogue à un résultat de Bonnans et Launay [8]). Pour simplifier la démonstration, on peut supposer que $\bar{x} = 0$; comme l'analyse est menée localement, seules les contraintes actives en 0 interviennent et on peut donc supposer que $g(x) = Ax$. Soit A_i la i-ième ligne de A. Posons

$$K = \left\{ i \in I \; ; \exists d^i \text{ avec } A_i\, d^i < 0, A_I\, d^i \leqq 0 \text{ et } A_J\, d^i = 0 \right\} ,$$

$$\bar{K} = I - K\,, \quad L = \bar{K} \cup J\,, \quad d^0 = \sum_{i \in K} d^i\,.$$

LEMME 1 : *Si $\bar{K} \neq \varnothing$, il existe $M' > 0$ tel que pour tout x :*

$$\|A_{\bar{K}}\, x\| \leqq M' [\, \|(A_{\bar{K}}\, x)^+\| + \|A_J\, x\|\,]\,.$$

Démonstration : Si la conclusion est fausse, il existe une suite $\{y^k\}$ vérifiant

$$\|A_{\bar{K}}\, y^k\| > n [\, \|(A_{\bar{K}}\, y^k)^+\| + \|A_J\, y^k\|\,]\,.$$

Changeant éventuellement y^k en $\|A_{\bar{K}}\, y^k\|^{-1}\, y^k$, on peut supposer que $\|A_{\bar{K}}\, y^k\| = 1$; donc $A_L\, y^k$ reste borné. Notons $(A_L)^{-1}$ un pseudo-inverse à droite de A_L (application linéaire continue de Im $(A_{\bar{K}}) \to \mathbb{R}^n$ telle que $A_L (A_L)^{-1}\, w = w$ pour tout w dans Im $(A_{\bar{K}})$). Changeant y^k en $(A_L)^{-1}$ $A_L\, y^k$ si nécessaire, on peut supposer aussi $\{y^k\}$ borné et (extrayant une sous-suite si nécessaire) que $y^k \to \bar{y}$. Il vient alors

$$\|A_{\bar{K}}\, \bar{y}\| = 1\,, \quad A_{\bar{K}}\, \bar{y} \leqq 0\,, \quad A_J\, \bar{y} = 0\,,$$

et il existe donc i_0 dans $\bar{K}$ tel que $A_{i_0}\, \bar{y} < 0$. Pour $\rho > 0$ assez grand on obtient

$$A_{i_0}(\bar{y} + \rho d^0) < 0\,, \quad A_I(\bar{y} + \rho d^0) \leqq 0\,, \quad A_J(\bar{y} + \rho d^0) = 0\,,$$

en contradiction avec la définition de K. □

Démonstration du théorème 5 : Un point x étant donné, il suffit de construire z tel que

$$A_I(x+z) \leqq 0, A_J(x+z) = 0 \quad \text{et} \quad \|z\| \leqq M^*[\|(A_I x)^+\| + \|A_J x\|],$$

M^* étant indépendante de x. D'après le lemme ci-dessus, il suffit de vérifier

$$\|z\| \leqq M[\|(A_K x)^+\| + \|A_L x\|].$$

Soit $(A_L)^{-1}$ un pseudo-inverse à droite de A_L et $y = -(A_L)^{-1} A_L x$. Il vient pour un certain $M_1 > 0$ (indépendant de x) :

$$\|y\| \leqq M_1 \|A_L x\|, \qquad A_L(x+y) = 0.$$

Prenons maintenant z de la forme $y + \rho d^0$ avec $\rho \geqq 0$. Alors $A_L(x+z) = 0$, et $A_K(x+z) \leqq 0$ dès que

$$\rho A_K d^0 \leqq -A_K(x+y).$$

Soit $\beta = -\max\{A_i d^0, i \in K\}$. Alors $\beta > 0$ par définition de K et pour assurer l'admissibilité de z il suffit donc de prendre (M_2 et M_3 dépendant ici de A) :

$$\begin{aligned} \rho &= \beta^{-1} \max\{(A_i(x+y))^+, i \in K\}, \\ &\leqq \beta^{-1}[\max\{(A_i x)^+, i \in \mathrm{K}\} + M_2 \|A_L y\|] \\ &\leqq \beta^{-1} M_3(\|(A_K x)^+\| + \|A_L x\|), \end{aligned}$$

ce qui donne l'inégalité cherchée. □

4. AUGMENTABILITÉ ET CONDITIONS SUFFISANTES DU DEUXIÈME ORDRE

Pour des problèmes convexes, le théorème 2 établit une relation précise entre la norme duale des multiplicateurs de Lagrange et les valeurs des coefficients de pénalisation. Ce type d'estimation est utile pour la mise en œuvre d'algorithmes dans lesquels la valeur du coefficient de pénalisation est mise à jour à chaque itération. Au contraire, on peut noter que les résultats de pénalisabilité fondés sur la théorie de la régularité ne donnent pas une estimation constructive du coefficient de pénalisation, la constante intervenant dans la condition de normalité étant difficile à évaluer. Dans cette section, nous nous intéressons à des estimations de la constante de pénalisation qui soient *constructives* et *optimales* (au sens où les estimations obtenues sont les meilleures possibles pour certaines classes de problèmes). Plus précisément, l'objectif de cette section est d'obtenir des résultats de pénalisabilité dans le cas où le problème est augmentable ou satisfait une condition suffisante du deuxième ordre. Introduisons d'abord quelques définitions. Si λ est un multiplicateur de Lagrange associé à un point admissible, posons

$$I^0(\lambda) = \{i \in I \; ; \lambda = 0\} \; , \; I^+(\lambda) = I - I^0(\lambda) \, , \; K(\lambda) = I^+(\lambda) \cup J(\lambda) \, .$$

Comme l'analyse est menée localement, nous pouvons supposer sans perte de généralité que $g_I(\bar{x}) = 0$. Le *lagrangien augmenté* associé à (P) est (voir par exemple Hestenes [13])

$$L_c(x, \lambda) = f(x) + \lambda^t g\,(x) + c\left[\sum_{i \in I^0(\lambda)} (g_i(x)^+)^2 + \sum_{i \in K(\lambda)} g_i(x)^2\right],$$

où $c > 0$ est la constante de pénalisation.

Cette expression diffère de celle donnée par Rockafellar [22] :

$$\bar{L}_c(x, \lambda) = f(x) + c_i \sum_{i \in I} \left[\left(g_i(x) + \frac{1}{2c}\lambda_i\right)^+\right]^2 - \frac{1}{4c}\sum_{i \in I} \lambda_i^2$$
$$+ \lambda_J^t \, g_J(x) + c \sum_{i \in I} (g_i)^2 \, .$$

Toutefois, si $\bar{x}$ et solution locale de (P) et $\bar{\lambda}$ est un multiplicateur de Lagrange associé à (P), on vérifie facilement que $L_c(x, \bar{\lambda})$ est égal à $\bar{L}_c(x, \bar{\lambda})$ si x est voisin de $\bar{x}$. Puisque nous menons une analyse locale, les résultats que nous établirons en utilisant L_c seront donc aussi valables pour $\bar{L}_c$.

Nous dirons que le problème (P) est augmentable au voisinage d'une solution $\bar{x}$ s'il existe un multiplicateur de Lagrange $\bar{\lambda}$ et $c_0 \geqq 0$ tel que, pour tout $c > c_0$, il existe un voisinage V de $\bar{x}$ tel que

$$f(\bar{x}) = L_c(\bar{x}, \bar{\lambda}) \leqq L_c(x, \bar{\lambda}) \quad \text{pour tout} \quad x \in V \, .$$

Le problème (P) sera dit strictement augmentable si l'inégalité ci-dessus est stricte dès que $x \neq \bar{x}$. On vérifie aisément que l'augmentabilité de (P) implique que $(\bar{x}, \bar{\lambda})$ est un point-selle local de L_c. Le concept d'augmentabilité est étudié de manière approfondie dans Rockafellar [22] et Hestenes [13].

Définissons le cône convexe fermé

$$C = \left\{d \in \mathbb{R}^n \; ; \nabla g_{I^0(\lambda)}^t \, d \leqq 0 \text{ et } \nabla g_{K(\lambda)}^t \, d = 0\right\} \, .$$

La condition suffisante classique du deuxième ordre est :

Les applications f et g sont deux fois différentiables en $\bar{x}$ et il existe un multiplicateur de Lagrange $\bar{\lambda}$ tel que le hessien du lagrangien

$$H_{\bar{\lambda}} = \nabla^2 f(\bar{x}) + \sum \bar{\lambda}_i \nabla^2 g_i(\bar{x}) \tag{CS2}$$

vérifie

$$d^t H_{\bar{\lambda}}\, d > 0\,, \quad \text{pour tout} \quad d \in C, d \neq 0\,.$$

On a alors le :

THÉORÈME 6 (Mac Cormick [15]) : *Soit $\bar{x}$ un point admissible de* (P) *et $\bar{\lambda}$ un multiplicateur associé tel que* (CS2) *est vérifiée. Alors $\bar{x}$ est une solution locale stricte de* (P). □

On sait que les conditions (CS2) impliquent l'augmentabilité (voir Rockafellar [22]). Nous allons maintenant montrer que l'augmentabilité implique la pénalisabilité, avec une estimation précise sur la valeur du coefficient de pénalisation : $r > \|\lambda\|_*$. Ceci impliquera en particulier que si (CS2) est vérifiée, (P) est pénalisable avec $r \geqq \|\bar{\lambda}\|_*$ (résultat de Han et Mangasarian [12]).

THÉORÈME 7 : *Si* (P) *est (strictement) augmentable en $\bar{x}$, un multiplicateur associé à l'augmentabilité étant $\bar{\lambda}$, alors* (P) *est (strictement) pénalisable en $\bar{x}$ et $\theta_r(x)$ admet un minimum (strict) en $\bar{x}$ dès que $r > \|\bar{\lambda}\|_*$.*

Démonstration : Nous allons montrer que, si $\bar{x}$ est un point admissible auquel est associé un multiplicateur $\bar{\lambda}$, et si $r > \|\bar{\lambda}\|_*$, alors la fonctionnelle pénalisée exacte domine le lagrangien augmenté associé au voisinage de $\bar{x}$, ce qui donnera le résultat.

Notons $I^0 = I^0(\bar{\lambda})$, $I^+ = I^+(\bar{\lambda})$, $K = K(\bar{\lambda})$. Il vient, pour x voisin de $\bar{x}$:

$$\begin{aligned} L_c(x, \bar{\lambda}) &= f(x) + (\bar{\lambda}_{I^0} + c g_{I^0}(x))^t\, g_{I^0}(x)^+ + (\lambda_K + c g_K(x))^t\, g_K(x)\,, \\ &\leqq f(x) + (\bar{\lambda}_I + c g_I(x))^t\, g_I(x)^+ + (\lambda_J + c g_J(x))^t\, g_J(x)\,, \\ &\leqq f(x) + \|\bar{\lambda} + c g(x)\|_*\, \|(g_I(x)^+, g_J(x))\|\,. \end{aligned}$$

La première inégalité est justifiée par le fait que $\lambda_{I^+} > 0$, donc $\lambda_{I^+} + c g_{I^+}(x)$ reste positif dans un voisinage V_1 de $\bar{x}$. Soit $r > \|\bar{\lambda}\|_*$. Il existe un voisinage V_2 de $\bar{x}$ tel que $\|\bar{\lambda} + cg(x)\|_* < r$ si $x \in V_2$; il vient donc $L_c(x, \bar{\lambda}) \leqq \theta_r(x)$ si $x \in V_1 \cap V_2$, qui est le résultat cherché. $\square$

Les résultats précédents sont toutefois loin d'être optimaux dans la mesure où on peut exprimer des conditions suffisantes plus faibles qui sont proches des conditions nécessaires. Notons les résultats suivants :

THÉORÈME 8 (voir Ben-Tal [2]) : *Supposons f et g deux fois différentiables en $\bar{x}$. Soit $\bar{x}$ une solution locale de* (P) *à laquelle est associée un multiplicateur de Lagrange $\bar{\lambda}$. Alors pour tout $d \in C$ il existe un multiplicateur de Lagrange λ associé à $\bar{x}$ tel que $d^t H_\lambda d \geqq 0$.*

THÉORÈME 9 (voir Ben-Tal [2], Han et Mangasarian [12]) : *Soit $\bar{x}$ un point stationnaire de* (P) *tel qu'il existe un ensemble de multiplicateurs de Lagrange Λ^* vérifiant :*

(CS2F) *Les applications f et g sont deux fois différentiables en $\bar{x}$ et pour tout $d \in C$, $d \neq 0$, on a* $\sup \{d^t H_\lambda d, \lambda \in \Lambda^*\} > 0$.

Alors $\bar{x}$ est un minimum local strict de (P).

Récemment ont été définis des lagrangiens augmentés « généralisés » de telle manière que (CS2F) implique l'augmentabilité (en un sens généralisé), la pénalisabilité se déduisant alors de l'augmentabilité (Bonnans [6]). Toutefois nous allons voir que la pénalisabilité peut également se déduire d'une manière simple de (CS2F).

THÉORÈME 10 : *Si $\bar{x}$ est une solution locale de* (P) *vérifiant* (CS2F), *et si Λ^* est borné, la fonctionnelle θ_r admet un minimum strict en $\bar{x}$ pour* $r > r^0 = \sup \{\|\lambda\|_*, \lambda \in \Lambda^*\}$.

Remarque 2 : Si (CS2F) est vérifiée, il existe un ensemble compact $\Lambda^\#$ inclus dans Λ^*, tel que $\max \{d^t H_\lambda d, \lambda \in \Lambda^\#\} > 0$ si $d \in C$, $d \neq 0$; on peut même prendre $\Lambda^\#$ fini (Bonnans [6]).

Démonstration du théorème 10 : Supposons les hypothèses du théorème vérifiées. Si la conclusion est fausse, il existe $r > r^0$ et $\{x^k\} \to \bar{x}$ tels que $\theta_r(\bar{x}^k) < \theta_r(\bar{x})$. Posons $\alpha^k = \|x^k - \bar{x}\|$. Extrayant si nécessaire une sous-suite, on peut supposer que $(\alpha^k)^{-1}(x^k - \bar{x}) \to d$ avec $\|d\| = 1$, et il vient alors $x^k = \bar{x} + \alpha^k d + o(\alpha^k)$. Puisque toute fonction convexe continue est localement lipschitzienne, $\theta_r(x)$ est elle-même lipschitzienne : donc $\theta_r(x^k) = \theta_r(\bar{x} + \alpha^k d) + o(\alpha^k)$. Rappelons que θ_r admet des dérivées directionnelles en $\bar{x}$ données par (supposant $g(\bar{x}) = 0$)

$$\theta_r'(\bar{x}, d) = \nabla f(\bar{x})^t d + r \| ((\nabla g_I(\bar{x})^t d)^+, \nabla g_J(\bar{x})^t d) \| ,$$

et il est donc nécessaire que $\theta_r'(\bar{x}, d) \leqq 0$. Soit $\mu \in \mathbb{R}^{|I| + |J|}$ vérifiant $\|\mu\|_* \leqq r$; utilisant l'inégalité de Cauchy généralisée, puis éliminant $\nabla f(\bar{x})$ grâce à (CO1), il vient, λ étant un multiplicateur de Lagrange quelconque de Λ^* :

$$0 \geqq \theta_r'(\bar{x}, d) \geqq \nabla f(\bar{x})^t d + \mu_I^t (\nabla g(\bar{x}) \,.\, d)^+ + \mu_J^t \nabla g_J(\bar{x})^t d ,$$
$$= \sum_{i \in I} [\mu_i (\nabla g_i(\bar{x})^t d)^+ - \lambda_i \nabla g_i(\bar{x})^t d] +$$
$$+ \sum_{i \in J} (\mu_i - \lambda_i) \nabla g_i(\bar{x})^t d .$$

Soit $\varepsilon > 0$: prenons

$$\mu_i = \begin{cases} \lambda_i + \varepsilon & \text{si } \; i \in I \; \text{ ou } \; i \in J \; \text{ et } \; \nabla g_i(\bar{x})^t d \geqq 0 , \\ \lambda_i - \varepsilon & \text{sinon} . \end{cases}$$

Puisque $\|\lambda\|_* < r$, pour ε assez petit on aura bien $\|\mu\|_* \leqq r$. On vérifie alors que chaque terme des deux sommes ci-dessus est positif. La somme étant négative, chaque terme doit être nul, ce qui n'est possible que si

$$\nabla g_{I^0(\lambda)}(\bar{x})^t d \leqq 0 \quad \text{et} \quad \nabla g_{K(\lambda)}(\bar{x})^t d = 0 ,$$

soit $d \in C$. Associons à d un multiplicateur λ appartenant à Λ^* tel que $d^t H_\lambda d > 0$. Alors

$$L(x^k, \lambda) = f(\bar{x}) + \frac{(\alpha^k)^2}{2} d^t H_\lambda d + o((\alpha^k)^2) ,$$

donc $L(x^k, \lambda) > f(\bar{x})$ pour k assez grand. Or, localement, θ_r domine le lagrangien augmenté (voir la preuve du théorème 7), donc en particulier $L(x, \lambda)$, d'où une contradiction. □

BIBLIOGRAPHIE

[1] M. S. Bazaraa, J. J. Goode, *Sufficient conditions for a globally exact penalty function without convexity,* Math. Programming Study 19, 1-15, 1982.

[2] A. Ben-Tal, *Second order and related extremality conditions in nonlinear programming,* J. Optim. Theory Appl. 31, 143-165, 1980.

[3] D. P. Bertsekas, *Necessary and sufficient conditions for a penalty method to be exact,* Math. Programming 9, 87-99, 1975.

[4] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods,* Academic Press, New York, 1982.