MULTIPLICATEURS, ANALYSE MARGINALE ET PENALISATION

Jean-Paul Penot, [1]

Mais qu'allait-il donc faire en cette galère? (Molière)

L'origine du texte qui suit réside dans une conversation de quai de gare avec un collègue économiste, bien avant que ne s'ouvre le vif débat sur la place des mathématiques dans les enseignements d'économie de ces derniers mois. "Nous sommes loin du campus, me disait ce collègue. Tu peux donc m'avouer que ces multiplicateurs de Lagrange ne sont qu'une petite astuce". Je lui avais répondu que ce n'était pas mon opinion, bien qu'un multiplicateur puisse apparaître comme un deus-ex-machina artificiel, que l'on s'empresse d'éliminer. Au contraire, je pensais qu'un bon éclairage des multiplicateurs peut les rendre naturels et justifier l'analyse marginale faite par les économistes, à condition de ne pas faire dire aux résultats des mathématiciens plus qu'ils ne peuvent affirmer.

La petite vulgarisation qui suit n'a d'autre but que de faire mieux connaître ce que l'on peut attendre d'une analyse mathématique. Elle ne prétend pas épuiser le sujet; les aspects particuliers de la programmation linéaire ([49], [38]) et de la programmation quadratique ([1], [38]) ne sont pas abordés notamment et le contrôle optimal des systèmes distribués n'est pas examiné (voir [3], [4], [30]). Pour les questions de décomposition et de décentralisation, on pourra consulter par exemple [31]; pour la théorie économique on pourra se référer à [12], [21], [28], [33].

Après quelques rappels sur l'utilisation d'un lagrangien et la notion de multiplicateur global qui ne sont donnés que pour éviter une confusion avec la notion de multiplicateur en un point au sens de Lagrange-Karush-Kuhn-Tucker, nous passons à la partie centrale de cet exposé qui concerne les relations entre les multiplicateurs et la dérivée (généralisée) de la fonction valeur. Une rapide application à la théorie de la commande optimale est amorcée.

Cette introduction à l'optimisation sous contraintes serait incomplète sans une évocation des liens étroits entre multiplicateurs et la pénalisation. L'idée simple de la pénalisation consiste à infliger une pénalité à un agent économique qui ne respecte pas les astreintes attendues (pollution, droit social, travail des enfants...). La difficulté consiste à déterminer la hauteur de la pénalité afin que l'astreinte soit respectée. Cette difficulté n'est pas que mathématique: on la retrouve aussi bien à l'échelle d'une petite communauté qu'au niveau planétaire.

1 Point de vue global

Considérons un problème de minimisation sous contraintes (dit problème de programmation mathématique)

$$(\mathcal{P}) \qquad \text{minimiser } f(x) \text{ sous les contraintes } x \in B,\ g(x) \in C$$

dans lequel B est un sous-ensemble d'un ensemble X, C un cône convexe fermé d'un espace de Banach Z, de dual Y, la fonction objectif (ou fonction coût) est $f : X \to \mathbb{R}$ et $g : X \to Z$ est une application. Dans le cas usuel $X = \mathbb{R}^n$, $Z = \mathbb{R}^p \times \mathbb{R}^q$, $C = \mathbb{R}^p_- \times \{0\}$ et $B = X$ (ou une partie simple de X comme $\mathbb{R}^n_+$ ou un pavé), ce qui signifie que l'ensemble $A := B \cap g^{-1}(C)$

---

[1] Université de Pau, jean-paul.penot@univ-pau.fr

des solutions admissibles ou faisables est défini par p inégalités et q égalités. La difficulté du problème provient du fait que A est défini implicitement, donc peu aisé à déterminer. De plus, la formulation précédente se prête à des modélisations dans lesquelles le cône C peut être d'intérieur vide: c'est le cas si C est le cône positif d'un espace L_p (cas où l'on a un obstacle ou bien une contrainte à respecter dans tout un intervalle de temps). Dans ce qui suit immédiatement nous ne faisons pas d'hypothèses de régularité sur f et g (ce qui permet de supposer que f vaut $+\infty$ sur $X \backslash B$).

Il est classique d'associer à $(\mathcal{P})$ le *lagrangien* $\ell : X \times Y \to \overline{\mathbb{R}}$ donné par

$$\ell(x,y) = f(x) + \langle y, g(x)\rangle \quad \text{pour } (x,y) \in X \times Y_+,$$

$\ell(x,y) = -\infty$ pour $(x,y) \in X \times (Y \backslash Y_+)$, où $Y_+ := C^0$ est le cône polaire de C dans Y donné par

$$C^0 := \{y \in Y : \forall z \in C,\ \langle y, z\rangle \leq 0\},$$

que l'on peut considérer comme le cône positif de Y, ordonné par l'ordre dual de l'ordre sur Z dont le cône positif est $Z_+ := -C$. On dit que $(\overline{x}, \overline{y}) \in X \times Y$ est un *col* (ou un point-selle) du lagrangien ℓ sur $X \times Y$ si $\ell(\overline{x}, \overline{y})$ est fini et si pour tout $(x,y) \in X \times Y$ on a

$$\ell(\overline{x}, y) \leq \ell(\overline{x}, \overline{y}) \leq \ell(x, \overline{y}). \tag{1}$$

Une représentation graphique illustre bien ces terminologies. L'intérêt de ce concept tient essentiellement au résultat élementaire qui suit (qui admet une réciproque sous certaines hypothèses, de convexité, notamment).

Lemme 1.1 *Si $(\overline{x}, \overline{y})$ est un col de ℓ sur $X \times Y$ alors $\overline{x}$ est une solution de $(\mathcal{P})$, $\overline{y} \in Y_+$ et* $\langle \overline{y}, g(\overline{x})\rangle = 0$.

En effet, pour tout $y \in Y_+$ on a

$$\langle y, g(\overline{x})\rangle = \ell(\overline{x}, \overline{y} + y) - \ell(\overline{x}, \overline{y}) \leq 0,$$

de sorte que $g(\overline{x}) \in C^{00} = C$, d'après le théorème du bipolaire. Ainsi $\overline{x} \in A := B \cap g^{-1}(C)$ car $\ell(\overline{x}, \overline{y})$ est fini. De plus, $\overline{y} \in Y_+$ et

$$\langle -\overline{y}, g(\overline{x})\rangle = \ell(\overline{x}, 0) - \ell(\overline{x}, \overline{y}) \leq 0,$$

de sorte que $\langle \overline{y}, g(\overline{x})\rangle = 0$. L'inégalité $\ell(\overline{x}, \overline{y}) \leq \ell(x, \overline{y})$ pour tout $x \in X$ et les relations $\langle \overline{y}, g(\overline{x})\rangle = 0 \geq \langle \overline{y}, g(x)\rangle$ pour $x \in g^{-1}(C)$ assurent alors que $f(\overline{x}) \leq f(x)$ pour tout $x \in A$: $\overline{x}$ est bien une solution optimale.$\square$

On dira que $\overline{y} \in Y_+$ est un *multiplicateur* (global) si $\inf_{x \in A} f(x) = \inf_{x \in X} \ell(x, \overline{y})$. Connaissant un multiplicateur, on est ramené à un problème sans contrainte, ce qui est un avantage précieux, en particulier du point de vue algorithmique. La pénalisation exacte réalise aussi cet objectif, mais en général avec une perte de régularité. Il est facile de caractériser l'ensemble M des multiplicateurs: $\overline{y} \in M$ à l'aide de la *fonction performance* $p : Z \to \overline{\mathbb{R}}$ (opposée de la *fonction marginale* m obtenue par maximisation de $-f$) du problème simplement perturbé

$$(\mathcal{S}_w) \quad \text{déterminer } p(w) := \inf\{f(x) : x \in X,\ g(x) + w \in C\}$$

qui s'introduit naturellement si l'on considère que les contraintes dépendent de paramètres $w \in Z$ que l'on peut augmenter ou modifier (stocks, capacités de production, effectifs du personnel....).

Lemme 1.2 *Si $p(0)$ est fini, l'ensemble M des multiplicateurs coïncide avec le sous-différentiel global $\partial^Z p(0)$ (dit aussi de Fenchel-Moreau) de la fonction performance p en $\overline{w} = 0$:*

$$M = \partial^Z p(0) := \{y \in Y : \forall z \in Z \;\; p(z) \geq p(0) + \langle y, z \rangle\}.$$

Lorsque p est convexe, $\partial^Z p(0)$ peut être considéré comme une sorte de substitut de la dérivée de p en 0. Dans le cas général $\partial^Z p(0)$ reste une mesure du changement de p, mais une mesure bien imprécise, car $\partial^Z p(0)$ est souvent vide, et, même si ce n'est pas le cas, $\partial^Z p(0)$ ne peut donner qu'une estimation par en dessous de la variation de la valeur optimale du problème.

Démonstration. Soit $\overline{y} \in \partial^Z p(0)$. Observons d'abord que $\overline{y} \in C^0$. En effet, pour tout $z \in C$ l'ensemble faisable $F(z) := g^{-1}(-z + C)$ de $(\mathcal{S}_z)$ contient l'ensemble faisable $F(0) := A$ de $(\mathcal{S}_0) = (\mathcal{P})$ puisque $g(x) \in C$ implique $g(x) + z \in C$. Il en résulte que $p(z) \leq p(0)$, donc $\langle \overline{y}, z \rangle \leq 0$. De plus, pour tout $x \in X$, prenant $z := -g(x)$, de sorte que x appartient à l'ensemble $F(z) := g^{-1}(-z + C)$, on obtient $f(x) \geq p(z) \geq p(0) + \langle \overline{y}, z \rangle = p(0) + \langle \overline{y}, -g(x) \rangle$, soit $\ell(x, \overline{y}) \geq p(0) = \inf_{x \in A} f(x)$ donc $\inf_{x \in X} \ell(x, \overline{y}) \geq p(0)$. En fait, l'égalité $\inf_{x \in X} \ell(x, \overline{y}) = p(0)$ a lieu puisque pour tout $x \in F(0)$ on a $\ell(x, \overline{y}) \leq f(x)$ car $g(x) \in C$ et $\overline{y} \in C^0$. Ainsi $\overline{y} \in M$.

Inversement, montrons l'inclusion $M \subset \partial^Z p(0)$. Soit $\overline{y} \in M$. Pour tout $z \in Z$ et pour tout $x \in F(z) := g^{-1}(-z + C)$ nous avons (puisque $\overline{y} \in Y_+$ et $g(x) + z \in C$)

$$p(0) \leq f(x) + \langle \overline{y}, g(x) \rangle \leq f(x) + \langle \overline{y}, -z \rangle.$$

Passant à la borne inférieure sur $x \in F(z)$ nous obtenons $p(0) \leq p(z) + \langle \overline{y}, -z \rangle$: $\overline{y} \in \partial^Z p(0)$.□

On peut déduire de cet énoncé un résultat d'existence.

Corollaire 1.3 *On suppose que la fonction performance p est convexe, ce qui est le cas si f et g sont convexes. Pour que l'ensemble des multiplicateurs M soit non vide il suffit qu'il existe $r > 0$, $m \in \mathbb{R}$ tels que pour tout $z \in Z$ de norme au plus r il existe $x \in X$ tel que $f(x) \leq m$, $g(x) + z \in C$.*

En effet, la fonction p est alors majorée par m sur la boule $B(0, r)$ de centre 0 et de rayon r, ce qui, pour une fonction convexe, est un critère pour que le sous-différentiel $\partial^Z p(0)$ soit non vide. Cette condition suffisante n'est nullement nécessaire, et d'autres critères existent.

Dans le cas convexe, on peut encore identifier l'ensemble M des multiplicateurs avec l'ensemble des solutions d'un problème dual (c.f. [13], [41]). Dans le cas général on ne peut attendre un tel résultat et il convient de prendre un point de vue moins global.

2 Point de vue ponctuel

On s'intéresse désormais à l'interprétation des multiplicateurs de Lagrange-Karush-Kuhn-Tucker en une solution $\overline{x}$ (en bref, un multiplicateur en $\overline{x}$) du problème $(\mathcal{P})$. On supposera désormais que X est un espace de Banach et que f et g sont de classe C^1. Rappelons qu'un élément $\overline{y}$ du dual Y de l'espace Z des contraintes de $(\mathcal{P})$ est un *multiplicateur en* $\overline{x}$ (et nous écrirons $\overline{y} \in M(\overline{x})$) si les relations suivantes sont vérifiées (ici, pour simplifier, nous supposons $B = X$; dans le cas général (2) doit être remplacée par une inéquation variationnelle):

$$\begin{aligned} f'(\overline{x}) + \overline{y} \circ g'(\overline{x}) &= 0, && (2) \\ \langle \overline{y}, g(\overline{x}) \rangle &= 0, \qquad y \in Y_+ = C^0. && (3) \end{aligned}$$

On voit facilement que si $\overline{x}$ est solution de $(\mathcal{P})$, l'ensemble $M(\overline{x})$ des multiplicateurs ponctuels en $\overline{x}$ est plus vaste que l'ensemble M des multiplicateurs globaux (et coincide avec M si $f+\overline{y}\circ g$ est convexe ou pseudo-convexe pour tout $\overline{y}\in M(\overline{x})$). Aussi, nous allons le comparer à un ensemble plus grand que le sous-différentiel global $\partial^Z p(0)$ de p en 0.

Pour cela, il convient d'introduire une notion de sous-différentiel qui généralise la notion de différentielle, et en particulier qui soit locale. Plusieurs choix sont possibles. Celui qui semble le plus approprié est le sous-différentiel contingent ou *sous-différentiel d'Hadamard* ou de Dini. Pour une fonction $f : X \to \overline{\mathbb{R}}$ finie en $\overline{x}\in X$, ce sous-différentiel est l'ensemble, noté $\partial f(\overline{x})$, des formes linéaires continues $\overline{x}^*$ sur X (appelées sous-dérivées) telles que pour tout $u\in X$ on ait

$$\liminf_{(t,v)\to(0_+,u)} \frac{1}{t}\left(f(\overline{x}+tv)-f(\overline{x})\right) \geq \langle \overline{x}^*, u\rangle,$$

soit encore, pour $t\in\mathbb{R}_+$, $v\in X$,

$$f(\overline{x}+tv) \geq f(\overline{x}) + \langle \overline{x}^*, tu\rangle - t\varepsilon(t+\|u-v\|)$$

avec $\varepsilon(r)\to 0$ quand $r\to 0$. Cette notion généralise la notion de dérivée au sens d'Hadamard; plus précisément, elle est un avatar unilatéral de ce concept. En effet, si f est différentiable au sens d'Hadamard, c'est à dire si la limite $f'(\overline{x})(u)$ de $(1/t)(f(\overline{x}+tv)-f(\overline{x}))$ quand $(t,v)\to(0_+,u)$ existe pour tout $u\in X$ et est linéaire continue en u, on a $\partial f(\overline{x}) = \{f'(\overline{x})\}$, et si $\partial(-f)(\overline{x})\cap(-\partial f(\overline{x}))$ est non vide, alors f est différentiable au sens d'Hadamard en $\overline{x}$. Si f est convexe, on voit facilement que $\partial f(\overline{x})$ coincide avec le sous-différentiel global $\partial^X f(\overline{x})$ défini plus haut. Dans les espaces de Hilbert (et aussi dans la classe plus vaste des espaces d'Asplund) le sous-différentiel $\partial f(\overline{x})$ est non vide pour tout $\overline{x}$ d'une partie partout dense du domaine de f pourvu que f soit semicontinue inférieurement. Enfin $\partial f(\overline{x})$ contient le sous-différentiel de Fréchet $\partial^- f(\overline{x})$ de f en $\overline{x}$ défini comme l'ensemble des $\overline{x}^*\in X^*$ tels

$$f(\overline{x}+x) \geq f(\overline{x}) + \langle \overline{x}^*, x\rangle - \varepsilon(\|x\|)\,\|x\|$$

avec $\varepsilon(r)\to 0$ quand $r\to 0$. Ces deux sous-différentiels jouissent de propriétés analogues. Lorsque X est de dimension finie, $\partial f(\overline{x})$ coincide avec $\partial^- f(\overline{x})$, de même que la dérivabilité au sens d'Hadamard coincide avec la différentiabilité au sens de Fréchet.

Proposition 2.1 *Pour toute solution $\overline{x}$ de $(\mathcal{P})$, toute sous-dérivée de p en 0 est un multiplicateur de $(\mathcal{P})$ en $\overline{x}$: $\partial p(0)\subset M(\overline{x})$.*

Démonstration. Soit $\overline{y}\in\partial p(0)$. Posons encore $F(z) := g^{-1}(C-z)$ pour $z\in Z$.Nous avons observé plus haut que $p(z)\leq p(0)$ pour tout $z\in C$. Il en résulte que

$$\langle \overline{y}, z\rangle \leq \liminf_{t\to 0_+} \frac{1}{t}\left(p(tz)-p(0)\right) \leq 0,$$

de sorte que $\overline{y}\in C^0 := Y_+$. De plus, comme pour $t\in[0,1]$ on a $g(\overline{x})-tg(\overline{x})\in C$ soit $\overline{x}\in F(-tg(\overline{x}))$ donc $p(-tg(\overline{x}))\leq f(\overline{x}) = p(0)$, nous obtenons $\langle \overline{y}, g(\overline{x})\rangle = 0$ car

$$\langle \overline{y}, -g(\overline{x})\rangle \leq \liminf_{t\to 0_+} \frac{1}{t}\left(p(-tg(\overline{x}))-p(0)\right) \leq 0.$$

Notons aussi que pour tout $x \in X$ nous avons $x \in F(-g(x))$ donc $p(-g(x)) \leq f(x)$. Comme $p(w+z) \leq p(w)$ pour $w \in Z$, $z \in C$, puisque $F(w) \subset F(w+z)$, en prenant $w = -g(x)$, $z = g(\overline{x})$ nous obtenons

$$p(g(\overline{x}) - g(x)) \leq p(-g(x)) \leq f(x),$$

avec égalité pour $x = \overline{x}$. Comme, pour $v \in X$, $\lim_{t \to 0_+} t^{-1}(g(\overline{x}) - g(\overline{x} + tv)) = -g'(\overline{x})v$, il en résulte que

$$\begin{aligned} \langle \overline{y}, -g'(\overline{x})v \rangle &\leq \liminf_{t \to 0_+} \frac{1}{t}\left(p(g(\overline{x}) - g(\overline{x} + tv)) - p(0)\right) \\ &\leq \liminf_{t \to 0_+} \frac{1}{t}\left(f(\overline{x} + tv) - f(\overline{x})\right) = f'(\overline{x})v. \end{aligned}$$

Ainsi $f'(\overline{x})v + \langle \overline{y}, g'(\overline{x})v \rangle \geq 0$, et en changeant v en $-v$ nous obtenons $f'(\overline{x})v + \langle \overline{y}, g'(\overline{x})v \rangle = 0$ pour tout $v \in X$. Nous avons bien obtenu que $\overline{y} \in M(\overline{x})$.$\square$

Corollaire 2.2 *Si p est différentiable en 0, $p'(0)$ est un multiplicateur en toute solution $\overline{x}$ de $(\mathcal{P})$.*

Exemple (problème de partage d'une ressource comme une surface cultivable ou une production de gaz naturel à transformer en divers composants chimiques). Une ressource de quantité $\overline{w}$ doit être partagée en n productions; chaque quantité x_i affectée à la production i apporte un revenu de $f_i(x_i)$. On suppose f_i différentiable. Le multiplicateur $\overline{y}$ en une solution $\overline{x}$ du problème

$$\text{minimiser } -(f_1(x_1) + ... + f_n(x_n)) \quad \text{sous la contrainte } \overline{w} - (x_1 + ... + x_n) = 0$$

vérifie $\overline{y} = f_i'(\overline{x}_i)$ pour $i = 1, ..., n$. Si le revenu total $m(w) := -p(w) = \max\{f_1(x_1) + ... + f_n(x_n) : x_1 + ... + x_n = w\}$ est une fonction dérivable de w en $\overline{w}$, alors on a $m'(\overline{w}) = \overline{y}$. La connaissance du multiplicateur $\overline{y}$ donne donc une information sur la variation du revenu lorsque la ressource change (extraction intensifiée ou diminuée, mise en jachère...).$\square$

Exemple (détermination de la forme d'un baril de surface donnée S et de volume maximum). Le problème

$$\text{minimiser } -\pi r^2 h \quad \text{sous la contrainte } S - 2\pi r(r+h) = 0$$

a pour solution $\overline{x} = (r, 2r)$ avec $r = (S/6\pi)^{1/2}$ et sa valeur est $p(S) = -2\pi(S/6\pi)^{3/2}$. La dérivée $p'(S) = -(1/2)(S/6\pi)^{1/2}$ de p coincide avec le multiplicateur $\overline{y}$.$\square$

Corollaire 2.3 *Si p est sous-différentiable en 0 (i.e. si $\partial p(0)$ n'est pas vide) et si pour une solution $\overline{x}$ de $(\mathcal{P})$ on a $g'(\overline{x})(X) = Z$, alors $\partial p(0) = M(\overline{x})$ et cet ensemble est un singleton.*

En effet, la surjectivité de $g'(\overline{x})$ assure que $M(\overline{x})$ est un singleton, de sorte que si $\partial p(0)$ n'est pas vide on a $\partial p(0) = M(\overline{x})$. On ne saurait conclure (comme on le fait parfois abusivement) que p est automatiquement (sous-) différentiable en 0.

Exemple. Soient $X = \mathbb{R}^2$, $Z = \mathbb{R}$, $\overline{x} = (0,0)$ et soient f et g données par $f(x_1, x_2) := x_1^2 \exp(-x_1^2) - \frac{2}{\pi} x_2 \arctan x_1$, $g(x_1, x_2) = x_2$. Alors $p(w) = -|w|$, $\partial p(0) = \emptyset$, $M(0) = \{0\}$.$\square$

On peut déduire du corollaire précédent un résultat d'unicité pour la solution de $(\mathcal{P})$ (c.f. [18] pour un problème de calcul des variations).

L'analyse précédente peut être étendue au cas de perturbations plus générales que celle de $(\mathcal{S}_w)$. Présentons une formulation qui permet une application aux problèmes de commande optimale. Soit W un espace métrique et soient W, Z des espaces normés. Soient $f : W \times X \to \mathbb{R}$, $g : W \times X \to Z$ des applications continues, différentiables par rapport à leur seconde variable et dont les dérivées $D_X f$, $D_X g$ par rapport à X sont continues sur $W \times X$. Considérons le problème

$$(\mathcal{P}_w) \quad \text{minimiser } f(w,x) \quad \text{sous la contrainte } g(w,x) = 0$$

dont nous notons $p(w)$ la valeur. Introduisons le lagrangien $\ell : W \times X \times Y \to \mathbb{R}$, où Y est le dual de Z, par

$$\ell(w,x,y) := f(w,x) + \langle y, g(w,x) \rangle.$$

Considérons $\overline{w} \in W$ et une solution $\overline{x}$ de $(\mathcal{P}_{\overline{w}})$. Supposons qu'il existe $c > 0$ et un voisinage V de $\overline{w}$ tels que pour tout $w \in V$ l'ensemble $S(w)$ des solutions de $(\mathcal{P}_w)$ soit non vide et que l'on ait

$$d(\overline{w}, S(w)) \leq c d(w, \overline{w}) \quad \text{pour tout } w \in V. \tag{4}$$

Pour établir le résultat suivant, connu sous le nom "envelop theorem", les économistes font implicitement l'hypothèse qu'il existe une solution $x(w)$ qui dépend de façon différentiable du paramètre. Dans ce cas, le résultat découle d'une simple application de la règle de composition des différentielles. Mais cette hypothèse est bien restrictive et elle ne peut être employée dans le cas où W n'a pas de structure d'espace normé comme dans l'application au principe de Pontriaguine qui suivra (Théorème 3.1). En fait, sous des hypothèses de différentiabilité d'ordre deux, la condition (4) peut elle-même être remplacée par une hypothèse plus faible de comportement höldérien, satisfaite si une condition suffisante du second ordre est vérifiée ainsi qu'une hypothèse assurant que les minimiseurs locaux proches de $\overline{x}$ sont des solutions globales (voir [25], [6], [39]). De plus, on peut remplacer l'ensemble des solutions par un ensemble de solutions approchées.

Proposition 2.4 *Soit $\overline{y}$ un multiplicateur en $\overline{x}$, i.e. un élément $\overline{y}$ de Y vérifiant $D_X f(\overline{w}, \overline{x}) + \overline{y} \circ D_X g(\overline{w}, \overline{x}) = 0$. Alors la variation de p est donnée par*

$$p(w) - p(\overline{w}) = \ell(w, \overline{x}, \overline{y}) - \ell(\overline{w}, \overline{x}, \overline{y}) + o(w), \tag{5}$$

où la fonction $o(w)$ vérifie $o(w)/d(w, \overline{w}) \to 0$ quand $w \to \overline{w}$ dans $W \backslash \{\overline{w}\}$.

L'intérêt de cette estimation réside dans le fait que l'on peut évaluer la variation de p avec la seule aide d'un multiplicateur en $\overline{x}$, sans résoudre le problème perturbé.

Démonstration. (Egorov et [32]) Prenons $b > c$ et, pour $w \in W$, choisissons $x_w \in S(w)$ tel que $\|x_w - \overline{x}\| \leq b d(w, \overline{w})$. Comme $p(w) = f(w, x_w)$, nous avons, pour un $\theta_w \in [0,1]$

$$\begin{aligned} p(w) - p(\overline{w}) &= f(w, x_w) - f(w, \overline{x}) + f(w, \overline{x}) - f(\overline{w}, \overline{x}) \\ &= D_X f(w, \overline{x} + \theta_w (x_w - \overline{x})).(x_w - \overline{x}) + f(w, \overline{x}) - f(\overline{w}, \overline{x}) \\ &= D_X f(\overline{w}, \overline{x}).(x_w - \overline{x}) + \alpha(w) b d(w, \overline{w}) + f(w, \overline{x}) - f(\overline{w}, \overline{x}) \end{aligned}$$

avec $|\alpha(w)| \leq \|D_X f(w, \overline{x} + \theta_w (x_w - \overline{x})) - D_X f(\overline{w}, \overline{x})\| \to 0$ quand $w \to \overline{w}$. De même, pour une fonction β satisfaisant $|\beta(w)| \leq \sup_{\theta \in [0,1]} \|\langle \overline{y}, D_X g(w, \overline{x} + \theta(x_w - \overline{x})) - D_X g(\overline{w}, \overline{x}) \rangle\| \to 0$ quand $w \to \overline{w}$ nous avons

$$\begin{aligned} 0 &= \langle \overline{y}, g(w, x_w) - g(\overline{w}, \overline{x}) \rangle = \langle \overline{y}, g(w, x_w) - g(w, \overline{x}) + g(w, \overline{x}) - g(\overline{w}, \overline{x}) \rangle \\ &= \langle \overline{y}, D_X g(\overline{w}, \overline{x}).(x_w - \overline{x}) + \beta(w) b d(w, \overline{w}) + g(w, \overline{x}) - g(\overline{w}, \overline{x}) \rangle. \end{aligned}$$

En ajoutant membre à membre les relations précédentes, et en tenant compte du fait que $\overline{y}$ est un multiplicateur, nous obtenons

$$p(w) - p(\overline{w}) = \ell(w, \overline{x}, \overline{y}) - \ell(\overline{w}, \overline{x}, \overline{y}) + (\alpha(w) + \beta(w))bd(w, \overline{w})).$$

Remarque. Une simple adaptation de la démonstration précédente donne une estimation dans le cas où la contrainte d'égalité du problème $(\mathcal{P}_w)$ est remplacée par une contrainte d'inégalité de la forme $g(w,x) \leq 0$, i.e. $g(w,x) \in C$, cône négatif de Z, avec $\overline{y} \in M(\overline{w}, \overline{x}) := \{y \in C^0 : \langle y, g(\overline{w}, \overline{x})\rangle = 0,\ D_X f(\overline{w}, \overline{x}) + \overline{y} \circ D_X g(\overline{w}, \overline{x}) = 0\}$. Observant que l'on a $\langle \overline{y}, g(w, x_w) - g(\overline{w}, \overline{x})\rangle \leq 0$, on obtient

$$p(w) - p(\overline{w}) \geq \ell(w, \overline{x}, \overline{y}) - \ell(\overline{w}, \overline{x}, \overline{y}) + o(w).$$

Corollaire 2.5 *Si W est un espace normé et si f et g admettent des dérivées partielles par rapport à W en $(\overline{w}, \overline{x})$ alors, sous les hypothèses qui précèdent, p est Fréchet-différentiable en $\overline{w}$ et pour tout multiplicateur $\overline{y}$ en $(\overline{w}, \overline{x})$ on a $Dp(\overline{w}) = D_W \ell(\overline{w}, \overline{x}, \overline{y})$.*

Dans le cas où la contrainte d'égalité est remplacée par une contrainte d'inégalité, la remarque précédente conduit à l'inclusion

$$\{D_W \ell(\overline{w}, \overline{x}, \overline{y}) : \overline{y} \in M(\overline{w}, \overline{x})\} \subset \partial^- p(\overline{w}), \tag{6}$$

tandis que l'inclusion inverse a lieu sous une simple hypothèse de qualification comme la surjectivité de la dérivée de g en $(\overline{w}, \overline{x})$. Cette dernière condition est satisfaite lorsque l'on considère le problème simplement perturbé $(\mathcal{S}_w)$, i.e. quand f ne dépend pas de w et $g(w,x) = \overline{g}(x) + w$ pour une fonction $\overline{g}$ de classe C^1, avec $W = Z$. Notons que dans ce cas les relations

$$\partial p(\overline{w}) = \partial^- p(\overline{w}) = \{D_W \ell(\overline{w}, \overline{x}, \overline{y}) : \overline{y} \in M(\overline{w}, \overline{x})\} \tag{7}$$

se réduisent aux égalités $\partial p(\overline{w}) = \partial^- p(\overline{w}) = M(\overline{w}, \overline{x})$. Il est alors intéressant de comparer ces simples égalités au résultat principal du chapitre 6 de [10] formulé à l'aide du sous-différentiel de Clarke $\partial^\uparrow p(\overline{w})$ et du sous-différentiel asymptotique associé $\partial^\uparrow_\infty p(\overline{w})$: avec les notations précédentes, la formule de [10] Théorème 6.5.2 s'écrit

$$\partial^\uparrow p(\overline{w}) = \overline{co}\left(\left(\bigcup_{\overline{x} \in S(\overline{w})} M(\overline{w}, \overline{x})\right) \cap \partial^\uparrow p(\overline{w}) + \left(\bigcup_{\overline{x} \in S(\overline{w})} M_0(\overline{w}, \overline{x})\right) \cap \partial^\uparrow p(\overline{w})\right), \tag{8}$$

où $M_0(\overline{w}, \overline{x})$ est l'ensemble des multiplicateurs singuliers en $(\overline{w}, \overline{x})$:

$$M_0(\overline{w}, \overline{x}) := \{y \in Y : y \in C^0,\ \langle y, g(\overline{w}, \overline{x})\rangle = 0,\ y \circ D_X g(\overline{w}, \overline{x}) = 0\}.$$

On observe qu'outre sa complexité, cette formule présente l'inconvénient de faire apparaître $\partial^\uparrow p(\overline{w})$ dans les deux membres. L'exemple suivant éclaire aussi les avantages respectifs de $\partial^\uparrow p$ et de $\partial p = \partial^- p$ (c'est un exemple en dimension 1).

Exemple. Soit f_0 une fonction de classe C^1 sur $\mathbb{R}$ et soit $f_1 : \mathbb{R} \to \mathbb{R}$ lipschitzienne, différentiable en 0, avec $f_1'(0) = 0$, $\partial^\uparrow f_1(0) = [-n, n]$ pour $n \in \mathbb{N}$ fixé. Prenons $W = X = Y = Z = \mathbb{R}$, $\overline{w} = 0$, $C = \{0\}$ et posons $f(w,x) = f_0(x) + f_1(w)$, $g(w,x) = w - x$. Alors $S(w) = \{w\}$, $p(w) = f_0(w) + f_1(w)$, de sorte que $\partial p(0) = \partial^- p(0) = \{f_0'(0)\} = \{M(0,0)\}$ tandis que $\partial^\uparrow p(0) = \{f_0'(0)\} + [-n, n]$, un ensemble qui peut être beaucoup plus grand.$\square$

3 Application au contrôle optimal

Illustrons ce qui précède par un problème de commande optimale (problème de Bolza); nous ne considérons que le cas le plus simple, sans contrainte sur l'état. Etant donné un espace métrique U, un espace de Banach E et un intervalle compact $T := [t_0, t_1]$ de $\mathbb{R}$, désignons par W l'ensemble des fonctions réglées normalisées de T dans U (fonctions qui ont une limite à droite et à gauche en tout point de $]t_0, t_1[$, qui sont continues à droite sur $[t_0, t_1[$ et à gauche en t_1) et désignons par X l'espace des fonctions continues, dérivables à droite sur $[t_0, t_1[$ qui sont primitives d'applications réglées normalisées de T dans E. Munissons X de la norme donnée par

$$\|x\| = \|x(t_0)\| + \int_T \left\| \dot{x}(t) \right\| dt,$$

$B := X_0 := \{x \in X : x(t_0) = 0\}$ de la norme induite et W de la distance donnée par

$$d(w_1, w_2) = \int_T d(w_1(t), w_2(t)) dt.$$

D'autres choix d'espaces fonctionnels sont possibles (on pourrait remplacer X par $W^{1,p}(T, E)$, notamment, mais dans l'analyse qui suit, il faudrait prendre des précautions et écarter des ensembles de mesure nulle; voir [34], [48] pour un développement rigoureux). Considérons pour $e_0 \in E$ et $w \in W$ donnés l'équation différentielle

$$\begin{aligned} \dot{x}(t) &= G(w(t), x(t)), && (9) \\ x(t_0) &= e_0, && (10) \end{aligned}$$

où $G : U \times E \to E$ une application continue, dont la dérivée partielle par rapport à la seconde variable $D_2 G$ existe et est continue. Etant donnée une fonction continue $F : T \times E \to \mathbb{R}$ dont la dérivée partielle $D_2 F$ existe et est continue, le problème de Bolza consiste en

$$(\mathcal{B}) \quad \text{minimiser } f(w, x) := \int_T F(w(t), x(t)) dt$$

sur l'ensemble des couples $(w, x) \in W \times X$ vérifiant (9)-(10). Le principe de Pontriaguine apporte une condition nécessaire pour qu'un couple $(\overline{w}, \overline{x})$ soit solution de $(\mathcal{B})$. Nous supposons qu'il existe $k > 0$ tel que pour tout $(t, u, e) \in T \times U \times E$ on ait

$$\|G(u, e) - G(\overline{w}(t), \overline{x}(t))\| \leq k d(u, \overline{w}(t)) + k d(e, \overline{x}(t)) \tag{11}$$

Introduisons la fonction de Pontriaguine, ou Hamiltonien, $H : U \times E \times E^* \to \mathbb{R}$, donnée par

$$H(u, e, e^*) = F(u, e) + \langle e^*, G(u, e) \rangle$$

et considérons la solution $\overline{y} : T \to E^*$ de l'équation adjointe

$$\begin{aligned} \dot{\overline{y}}(t) &= -\overline{y}(t) \circ D_2 G(\overline{w}(t), \overline{x}(t)) - D_2 F(\overline{w}(t), \overline{x}(t)) && (12) \\ \overline{y}(t_1) &= 0. && (13) \end{aligned}$$

Théorème 3.1 *(Principe de Pontriaguine) Pour que* $(\overline{w}, \overline{x})$ *soit solution de* $(\mathcal{B})$ *il faut que*

$$H(\overline{w}(t), \overline{x}(t), \overline{y}(t)) = \min_{u \in U} H(u, \overline{x}(t), \overline{y}(t)) \quad \text{pour tout } t \in T. \tag{14}$$

Démonstration.(Egorov, [23], [32], [44]) Comme on ne perd pas de généralité en supposant $e_0 = 0$ (quitte à faire une translation) les équations (9)-(10) peuvent se traduire par la contrainte

$$g(w,x) := G(w(\cdot), x(\cdot)) - \overset{\cdot}{x}(\cdot) = 0, \qquad x \in X_0,$$

g étant considéré comme une application de $W \times X$ dans l'espace $Z := R(T,E)$ des applications réglées normalisées de T dans E. Identifions $R(T,E^*)$ à un sous-espace du dual Y de $R(T,E)$ de sorte que la restriction du lagrangien à $W \times X_0 \times R(T,E^*)$ s'écrit

$$\ell(w,x,y) = \int_T F(w(t),x(t))dt + \int_T \langle y(t), G(w(t),x(t)) - \overset{\cdot}{x}(t)\rangle dt.$$

La propriété pour $y \in R(T,E^*)$ d'être un multiplicateur en $(\overline{w}, \overline{x})$ se traduit par

$$\int_T [D_2F(\overline{w}(t),\overline{x}(t))v(t) + \langle y(t), D_2G(\overline{w}(t),\overline{x}(t))v(t) - \overset{\cdot}{v}(t)\rangle]dt = 0$$

pour tout $v \in X$ tel que $v(t_0) = 0$. Le lemme fondamental du calcul des variations (dû à Du Bois Reymond, que l'on peut considérer comme un précurseur de la théorie des distributions) assure que ceci est le cas si y est une primitive de fonction réglée vérifiant

$$\begin{aligned} \overset{\cdot}{y}(t) &= -y(t) \circ D_2G(\overline{w}(t),\overline{x}(t)) - D_2F(\overline{w}(t),\overline{x}(t)), \\ y(t_1) &= 0. \end{aligned}$$

En particulier, c'est le cas pour $y = \overline{y}$ (une simple intégration par parties suffit dans ce cas). L'hypothèse (11) et le lemme de Gronwall entraînent que pour $w \in W$ la solution x_w de (9)-(10) vérifie la majoration

$$\|x_w(t) - \overline{x}(t)\| \leq ke^{k(t_1-t_0)} \int_T d\,(w(s), \overline{w}(s))\, ds$$

de sorte que (4) est satisfaite. Nous pouvons donc appliquer la proposition précédente:

$$p(w) - p(\overline{w}) = \ell(w, \overline{x}, \overline{y}) - \ell(\overline{w}, \overline{x}, \overline{y}) + o(w).$$

Comme $\int_T \langle \overline{y}(t), \overset{\cdot}{\overline{x}}(t)\rangle dt$ ne dépend pas de w, nous obtenons

$$\ell(w, \overline{x}, \overline{y}) - \ell(\overline{w}, \overline{x}, \overline{y}) = \int_T H(w(t), \overline{x}(t), \overline{y}(t))dt - \int_T H(\overline{w}(t), \overline{x}(t), \overline{y}(t))dt. \tag{15}$$

Supposons qu'il existe $r \in [t_0, t_1[$, $v \in U$ et $\alpha > 0$ tels que

$$H(\overline{w}(r), \overline{x}(r), \overline{y}(r)) > H(v, \overline{x}(r), \overline{y}(r)) + \alpha.$$

Nous pouvons alors trouver $s > 0$ tel que pour tout $t \in [r, r+s] \subset T$

$$H(\overline{w}(t), \overline{x}(t), \overline{y}(t)) > H(v, \overline{x}(t), \overline{y}(t)) + \alpha$$

car le premier membre converge vers $H(\overline{w}(r), \overline{x}(r), \overline{y}(r))$ quand $t \to r$ à droite. Pour $s \in]0, t-r[$, soit $w_s \in W$ l'application qui coincide avec $\overline{w}$ sur $T \backslash [r, r+s[$ et vaut v sur $[r, r+s[$. D'après (15) nous avons

$$\begin{aligned} p(w_s) - p(\overline{w}) &= \int_T H(w_s(t), \overline{x}(t), \overline{y}(t))dt - \int_T H(\overline{w}(t), \overline{x}(t), \overline{y}(t))dt + o(w_s) \\ &\leq -\alpha s + o(d(w_s, \overline{w})). \end{aligned}$$

Comme $\overline{w}(t) \to \overline{w}(r)$ quand $t \to r$ à droite, nous avons que $d(w_s, \overline{w})$ est équivalent à $sd(v, \overline{w}(r))$. L'inégalité précédente contredit alors l'optimalité de $(\overline{w}, \overline{x})$. Ainsi, pour tout $v \in U$, on a

$$H(\overline{w}(t), \overline{x}(t), y(t)) \leq H(v, \overline{x}(t), y(t)).$$

En faisant un passage à la limite, nous obtenons cette inégalité pour $t = t_1$.$\square$

4 Multiplicateurs et pénalisation

Dans cette section finale, nous abordons les relations entre les multiplicateurs et la pénalisation. L'idée générale est commune : remplacer un problème avec contrainte

$$(\mathcal{M}) \qquad \text{Minimiser } f(x) \text{ sous la contrainte } x \in A \subset X$$

par un problème sans contrainte comportant un terme complémentaire. Lorsque f est (localement) lipschitzienne de taux ℓ on voit facilement qu'un minimiseur (local) de f sur A est un minimiseur (local) de $f + \ell d(\cdot, A)$ sur X ([10]). Ce fait n'est pas d'un grand intérêt pratique lorsque A est déterminé de manière implicite par des égalités ou des inégalités comme dans le problème $(\mathcal{P})$ où $B = X,\ A = g^{-1}(C)$ car, si A est difficile à déterminer, $d(\cdot, A)$ le sera aussi. En revanche, si la contrainte est *métriquement régulière* en ce sens qu'il existe une constante $c > 0$ telle que

$$d(\cdot, A) \leq cd(g(\cdot), C)$$

(au moins localement), alors on pourra s'intéresser à la minimisation de la fonction pénalisée f_r donnée pour $r > 0$ (ici $r \geq c\ell$) par

$$f_r := f + rd(g(\cdot), C).$$

Ici, nous ne considérons que le cas où r est fixé (pénalisation "exacte") tandis que dans les algorithmes on est amené à augmenter r au cours des itérations.

Il est clair que tout minimiseur $\overline{x}$ de f_r appartenant à A est solution de $(\mathcal{M})$. On a plus:

Proposition 4.1 *Si $\overline{x}$ est un minimiseur de f_r appartenant à $A = g^{-1}(C)$ et si f et g sont différentiables en $\overline{x}$ alors $\overline{x}$ est un minimiseur de f sur A et l'ensemble $M(\overline{x})$ des multiplicateurs en $\overline{x}$ est non vide. De plus, $\overline{r} := \inf\{\|y\| : y \in M(\overline{x})\} \leq r$.*

Démonstration. Pour tout $x \in X$ et pour tout $z \in C$ on a

$$f(\overline{x}) = f_r(\overline{x}) \leq f(x) + r\,\|g(x) - z\|\,,$$

de sorte que, prenant $x = \overline{x} + tu$ avec $t > 0$, $u \in X$, $z = g(\overline{x}) + tw$ avec $w \in \mathbb{R}_+(C - g(\overline{x}))$ on obtient

$$f'(\overline{x})(u) + r\,\|g'(\overline{x})u - w\| \geq 0.$$

Les règles de calcul de l'analyse convexe fournissent alors $\overline{y} \in rB_Y$ tel que $f'(\overline{x}) + \overline{y} \circ g'(\overline{x}) = 0$ et $0 \in -\overline{y} + N(C, g(\overline{x}))$ où $N(C, g(\overline{x})) = \{y \in C^0 : \langle y, g(\overline{x})\rangle = 0\}$ est le cône normal à C en $g(\overline{x})$, cône polaire du cône tangent radial $\mathbb{R}_+(C - g(\overline{x}))$: on a bien $\overline{y} \in M(\overline{x})$ et $\|\overline{y}\| \leq r$.□

Si la condition de Robinson

$$g'(\overline{x})(X) - \mathbb{R}_+(C - g(\overline{x})) = 0 \tag{16}$$

est satisfaite, ce résultat peut aussi être déduit d'une forme du lemme de Farkas puisque l'on a $f'(\overline{x})(u) \geq 0$ pour tout $u \in X$ tel que $g'(\overline{x})u \in \mathbb{R}_+(C - g(\overline{x}))$. Mais notons qu'une "condition de qualification" comme (16), ou celles de [5] Th. 1., [26] Th. 1 n'est pas indispensable.

Proposition 4.2 *Si l'ensemble M des multiplicateurs n'est pas vide on a* $\inf f_r(X) \geq \inf f(A)$ *pour* $r > \widehat{r} := \inf\{\|y\| : y \in M\}$. *Si $\overline{x}$ est un minimiseur de f_r avec $r > \widehat{r}$, alors $\overline{x} \in A$ et $\overline{x}$ est solution de ($\mathcal{P}$).*

Démonstration. Observons que pour $x \in X$, $\overline{y} \in Y_+$ et pour $r \geq \|\overline{y}\|$ on a $r\,\|g(x) - z\| \geq \langle \overline{y}, g(x) - z\rangle \geq \langle \overline{y}, g(x)\rangle$ pour tout $z \in C$, de sorte que $rd(g(x), C) \geq \langle \overline{y}, g(x)\rangle$ et $f_r(x) \geq \ell(x, \overline{y})$. Ainsi, si $\overline{y} \in M$, on obtient $\inf f_r(X) \geq \inf \ell(X, \overline{y}) = \inf f(A)$. D'où la première assertion. Pour obtenir la seconde, on observe que si $d(g(\overline{x}), C) > 0$ pour $r > \widehat{r}$, on peut trouver $\overline{y} \in M$ et $z \in C$ tels que $r > \|\overline{y}\|$ et $rd(g(\overline{x}), C) > \|\overline{y}\|\,\|g(\overline{x}) - z\| \geq \langle \overline{y}, g(\overline{x})\rangle$, de sorte que $\inf f_r(X) = f_r(\overline{x}) > \ell(\overline{x}, \overline{y}) \geq \inf \ell(X, \overline{y}) = \inf f(A) = \inf f_r(A)$, une contradiction. Ainsi $\overline{x} \in A$.□

On trouvera dans [5], [8], [20], [22], [26] par exemple des résultats supplémentaires donnant notamment des conditions suffisantes du second ordre pour qu'un minimiseur local de f sur A soit aussi un minimiseur local de f_r. Notons aussi que les méthodes de pénalisation constituent un moyen classique pour obtenir l'existence de multiplicateurs (Beltrami, Barbu, Hestenes...) et que la théorie des lagrangiens augmentés peut être vue comme une synthèse de ces deux approches, attirante aussi bien du point de vue théorique que du point de vue numérique.

Les références [2], [6], [29], [39] contiennent de vastes bibliographies sur l'utilisation des multiplicateurs pour les questions de sensibilité, la liste suivante étant très limitative.

References

[1] J.-P. Aubin, P. Nepomiastchy et A.M. Charles, *Méthodes explicites de l'optimisation*, Dunod, (1982).

[2] B. Bank, et al., *Non-linear parametric optimization*, Akademie Verlag, Berlin (1982).

[3] M. Bergounioux, *Optimisation et contrôle des systèmes linéaires*, Dunod, Paris (2001).

[4] M. Bergounioux and F. Mignot, Control of variational inequalities and Lagrange multipliers, ESAIM: Control, Optim. and Calculus of Variations.5 (2000), 45-70.

[5] J.F. Bonnans, Théorie de la pénalisation exacte, Math. Mod. Num. Anal. 24 (2) (1990), 197-210.

[6] J.F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer, New York (2000).

[7] J. F. Bonnans, A.D. Ioffe and A. Shapiro, Expansion of exact and approximate solutions in nonlinear programming, in "*Advances in Optimization, Proceedings, Lambrecht, FRG, 1991*", W. Oettli and D. Pallaschke (eds.), Lecture Notes in Economics and Mathematical Systems, No 382, Springer Verlag, Berlin (1992), 103-117.

[8] J.V. Burke, An exact penalization viewpoint of constrained optimization, SIAM J. Control and Op. 29 (1991), 968-998.

[9] J.-M. Bonnisseau and C. Le Van, Subdifferentiability of the value function in economic problems, Southeast Asian Bull. Math. 19, No.2 (1995), 107-116.

[10] F.H. Clarke, *Optimization and nonsmooth analysis*, Wiley, New York (1983).

[11] F.H. Clarke and P. Loewen, The value function in optimal control: sensitivity, controllability and time optimality, SIAM J. Control Optim. 24 (1986), 243-263.

[12] A. Deaton and J. Muellbauer, *Economics and consumer theory*, Cambridge Univ. Press, Cambridge (1980).

[13] I. Ekeland et R. Témam, *Analyse convexe et problèmes variationnels*, Dunod, Paris (1974).

[14] A.V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming,* Academic Press, New York (1983).

[15] A.V. Fiacco, ed.,"*Mathematical programming with data perturbations*", Lectures notes in pure and applied math. No 195, Dekker, New York (1998)

[16] A.V. Fiacco and W.P. Hutzler, Basic results in the development of sensitivity and stability analysis in nonlinear programming, Comput. and Oper. Research 9 (1) : 1-28 (1982).

[17] J. Gauvin, *Théorie de la programmation mathématique non convexe/ Theory of nonconvex programming*, Les Publications du CRM, Montréal, (1994).

[18] D. Henri, Nonuniqueness of solutions in the calculus of variations: a geometric approach, SIAM J. Control and Opt. 18 (6) : 627-639 (1980).

[19] J.-B. Hiriart-Urruty, Gradients généralisés de fonctions marginales, SIAM J. Control Opt. 16 : 310-316 (1978).

[20] J.-B. Hiriart-Urruty, Optimisation et analyse convexe, Presses Universitaires de France (1998)

[21] M.D. Intriligator, *Mathematical optimization and economic theory*, Prentice-Hall, Englewood Cliffs, N.J. (1971).

[22] A.D. Ioffe, Necessary and sufficient conditions for a local minimum I: A reduction theorem and first order conditions, SIAM J. Control Optim. 17 (1979), 245-250.

[23] A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems*, Nauka, Moscou (1974), North-Holland, Amsterdam (1979).

[24] R. Janin, Directional derivative of the marginal function in nonlinear programming, Math. Prog. Study 21 : 110-126 (1984).

[25] R. Janin, J.-C. Mado and J. Narayaninsamy, Second order multipliers and marginal function in non linear programs, Optimization 22 : 163-176 (1991).

[26] J.-B. Lasserre, Exact penalty functions and Lagrange multipliers, R.A.I.R.O. Automatique, Systems Anal. and Control 14 (1980), 117-125.

[27] F. Lempio and H. Maurer, Differential stability in infinite dimensional programming, Applied Mathematics and Optimization, 6 : 139-152 (1980).

[28] D. Léonard and Ngo Van Long, Optimal control theory and static optimization in economics, Cambridge Univ. Press, Cambridge (1992).

[29] E.S. Levitin, *Perturbation theory in mathematical programming and its applications,* J. Wiley, Chichester (1994).

[30] J.-L. Lions, *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, Dunod, Paris (1971), traduction anglaise, Springer-Verlag (1971).

[31] J.-L. Lions et G.I. Marchouk, *Sur les méthodes numériques en sciences physiques et économiques*, Dunod, Paris (1974).

[32] D. G. Luenberger, *Optimization by vector space methods*, Wiley, New York (1969).

[33] A. MasColell, M.D. Whinston and J.R. Green, *Microeconomic theory*, Oxford Univ. Press, Oxford (1995).

[34] S. Mirica, A proof of Pontryagin's minimum principle using dynamic programming, J. Math. Anal. Appl. 170 (2) (1992), 501-512.

[35] R. Pallu de la Barrière, *Cours d'automatique théorique*, Dunod, Paris, 1966.

[36] J.-P. Penot, Differentiability of relations and differential stability of perturbed optimization problems, SIAM J. Control and Optim. 22 (4) (1984), 529-551.

[37] J.-P. Penot, Multipliers and generalized derivatives of performance functions, J. Optim. Th. Appl. 93 (3), (1997), 609-618.

[38] J.-P. Penot, Central and peripheral results in the study of marginal and performance functions, dans "*Mathematical programming with data perturbations*", A.V. Fiacco ed., Lectures notes in pure and applied math. No 195, Dekker, New York (1998), 305-337.

[39] J.-P. Penot, Points de vue sur l'analyse de sensibilité en programmation mathématique, "*Actes des sixièmes journées du groupe MODE*", Editions Atlantiques et Université de Poitiers (1999), 176-203.

[40] S.M. Robinson, Stability theory for systems of inequalities. Part I :linear systems, SIAM J. Numer. Anal. 12 (5) (1975), 754-769.

[41] R.T. Rockafellar, *Conjugate duality and optimization*, SIAM, Philadelphia (1974).

[42] R.T. Rockafellar, Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming, Math. Prog. Study 17 : 28-66 (1982).

[43] R.T. Rockafellar, Directional differentiability of the optimal value function in a nonlinear programming problem, Math. Prog. Study 21 : 213-226 (1984).

[44] I.L. Rozonoer, The maximum principle of L.S. Pontryagin in optimal-system theory, part III, Automat. Remote Control, 20 (1959), 1517-1532.

[45] P.A. Samuelson, *Foundations of economic analysis*, Harvard Univ. Press, Cambridge, Mass. (1947).

[46] L. Thibault, On subdifferentials of optimal value functions, SIAM J. Control Opt. 29 (1991), 1019-1036.

[47] R.B. Vinter, *Optimal Control*, Birkhäuser, Boston (2000).

[48] R.B. Vinter, P. Wolenski, Hamilton-Jacobi theory for optimal control problems with data measurable in time, SIAM J. Control Optim. 28 (6) (1990), 1404-1419.

[49] A.C. Williams, Marginal values in linear programming, SIAM J. Appl. Math. 11 : 82-94 (1963).