MASTER EN SCIENCES MATHÉMATIQUES

La méthode de Ritz Galerkin en contrôle optimal : convergence et calcul de l'erreur

Toint, Philippe

Award date:
1974

Awarding institution:
Universite de Namur

La méthode de Ritz Galerkin en contrôle optimal

Convergence et calcul de l'erreur

Philippe Toint

1973-1974

La méthode de Ritz-Galerkin en contrôle optimal :

Convergence et calcul de l'erreur.

Philippe TOINT 1973-74

Introduction

On sait que tout problème de physique mathématique conduit naturellement à la résolution d'une ou plusieurs équations fonctionnelles :

$$Ax = f \qquad (I)$$

où A opère d'un espace X dans un espace Y, f est donné dans Y et x est donné dans X (exemples : équations différentielles ordinaires, équations intégrales, équations aux dérivées partielles)

En général, la solution de l'équation (I) est impossible à déterminer explicitement ou encore sa forme explicite est si compliquée qu'elle est inutilisable pratiquement, et on s'intéresse par conséquent à la résolution approchée de l'équation. L'idée est alors de remplacer les espaces X et Y par des espaces "plus simples" X_n et Y_n et d'associer à (I) une famille de problèmes approchés

$$A_n x_n = f_n$$

où A_n approxime A, f_n approxime f dans Y_n et x_n approxime x

dans X_n (du moins on le souhaite). Ce procédé porte généralement le nom de Ritz-Galerkin.

Les problèmes qui se posent sont les suivants :

1) étude de l'équation exacte
2) étude des équations approchées
3) étude de la stabilité et de la convergence
4) étude de l'erreur (entre x et x_n) pour un n fixé.

Au cours de ce mémoire, nous montrons, via la théorie des multiplicateurs de Lagrange, comment on peut appliquer les idées précédentes à des problèmes de contrôle optimal.

Rappelons, en termes volontairement vagues, que la théorie mathématique du contrôle optimal se fait à partir des données suivantes : 1) un contrôle u "à notre disposition" dans un ensemble U_{adm} (l'ensemble des "contrôles admissibles"), 2) l'état $x(u)$ du système à commander qui est donné, pour u choisi, par la résolution d'une équation

$$B\,x(u) = \text{fonction donnée de } u$$

où B est l'opérateur (supposé connu) qui "représente" le système à contrôler (B est le modèle du système), 3) la fonction économique $J[u]$. On cherche

$$\text{Inf}\,\{J(u) \mid u \in U_{adm}\}$$

Les questions étudiées dans ce travail sont les suivantes.

Le procédé de Ritz-Galerkin est-il convergent ? Comment calcule-t-on les contrôles approchés, les états approchés et les valeurs de la fonction économique J ?

Le plan que nous adopterons est le suivant.

Dans un premier chapitre, nous rappellerons différents points de la théorie générale de l'optimisation pour pouvoir aborder, avec les outils nécessaires, le second chapitre qui posera, de manière précise, le problème de contrôle optimal. Un dernier chapitre sera alors consacré à l'application de la méthode de Ritz-Galerkin à quelques modèles plus précis.

Chapitre I

Questions générales d'optimisation

I Introduction.

Ce chapitre a pour but essentiel de rappeler quelques outils fréquemment utilisés dans la théorie de l'optimisation, et qui nous seront utiles dans la suite. Il n'est évidemment pas exhaustif, et beaucoup des cas qu'il traite peuvent l'être aussi dans un cadre beaucoup plus général

Nous parlerons d'abord des problèmes de dualité, qui seront introduits par le biais des formes bilinéaires. Ensuite, nous aborderons la théorie de l'optimisation avec contraintes, globale et locale, vue sous l'angle des multiplicateurs de Lagrange.

II La dualité : espaces duaux

§1 Définissons d'abord le concept même de dualité dans le cadre des espaces vectoriels.

Nous dirons que

> Deux espaces vectoriels F et G sont mis en dualité par la forme B ssi
> 1) $B(x,y)$ est bilinéaire sur $F \times G$
> 2) $\forall_F x \neq 0 \quad \exists_G y \quad B(x,y) \neq 0$
> 3) $\forall_G y \neq 0 \quad \exists_F x \quad B(x,y) \neq 0$

Exemples fondamentaux :

a) Considérons d'abord $B(x,y) = \langle x, y \rangle = y(x)$ où G est l'espace vectoriel des formes linéaires sur F. Nous retrouvons alors la notion de dual algébrique (F^*) La propriété 2) est vraie puisque y est une forme linéaire et, pour vérifier 3), on s'appuie sur le fait que $\forall x \neq 0$ dans E, on peut trouver une forme linéaire y sur E telle que $y(x) \neq 0$.

b) Choisissons maintenant pour G l'espace des formes linéaires continues sur F, soit F'. On a donc F' est un sous espace de F^*. On prend aussi $B(x,y) = \langle x, y \rangle$ qui vérifie de la même manière 2). Pour vérifier 3) il faut utiliser un corollaire du théorème de Hahn - Banach pour affirmer la non trivialité de F' lorsque F est de dimension infinie
On retrouve ici le dual topologique F'.
Remarque: il est clair que F' dépend de la topologie de F.

§2 Topologies sur le dual topologique

Soit X un espace linéaire normé et $Y = \mathbb{R}$ (ou $\mathbb{C}$) (muni de la topologie usuelle)

$\mathcal{L}(X,Y)$ = { formes linéaires continues de X dans Y } peut donc être considéré comme le dual topologique X' de X.

> La topologie de la convergence uniforme sur X' est appelée la topologie forte de X'

> La topologie de la convergence simple dans Y est appelée la topologie faible* de X'

Rappelons aussi que X' muni de la norme $\|f\| = \sup_{\|x\| \leq 1} |f(x)|$ est un espace de Banach.

Considérons maintenant X'' le bidual de X, i.e. le dual topologique de X' muni de la topologie forte

> Une suite $(x'_n)_{\mathbb{N}} \in X'^{\mathbb{N}}$ converge faiblement vers $x' \in X'$ ssi $\forall x'' \in X''$ $\langle x'_n, x'' \rangle \longrightarrow \langle x', x'' \rangle$
> La topologie associée est la topologie faible de X'

§3 Plongement de X dans X''

Pour chaque $x_0 \in X$, on peut définir une forme linéaire continue $f_0(x')$ sur X' par $f_0(x') = \langle x_0, x' \rangle$. Nous avons donc construit

$$J : X \longrightarrow X'' \; ; \; x_0 \longmapsto Jx_0 = f_0$$

On peut alors démontrer que J est un isomorphisme de X sur $JX \subseteq X''$

On dit qu'un espace de Banach est réflexif ssi X peut être identifié à X'' (algébriquement et topologiquement) en utilisant J

Il est immédiat, par le théorème de Riesz, que tout espace de Hilbert est réflexif.

De plus, comme remarqué dans Yosida, X'' est un espace de Banach (cfr supra) et donc tout espace reflexif normé linéaire est au moins un espace de Banach.

§4 Remarque pour les espaces de Hilbert.

Par le théorème de Riesz, on peut identifier tout espace de Hilbert avec son dual. La forme de dualité définie sur $X \times X'$ devient donc définie sur $X \times X$ par (x, y) où $(\cdot,\cdot)$ est le produit scalaire.

III La dualité : opérateurs duaux.

§1 Opérateurs duaux

Nous allons étendre la notion de "matrice transposée" à celle d'"opérateur dual"

Théorème I.1

Soient X, Y deux espaces linéaires normés et X' et Y' leurs duaux topologiques forts.

Soit $T : \mathrm{dom}(T) \subseteq X \longrightarrow Y$ linéaire

Considérons les paires $(x', y') \in X' \times Y'$ qui satisfont

$$\langle Tx, y' \rangle = \langle x, x' \rangle \qquad \forall x \in \mathrm{dom}(T)$$

Alors x' est déterminé uniquement par y'
ssi $\overline{dom(T)} = X$

En effet, par la linéarité du problème, il suffit de considérer l'équation

$$\langle x, x' \rangle = 0 \quad \forall x \in dom(T) \Rightarrow x' = 0$$

La condition suffisante découle alors de la continuité de x'.
Supposons que $\overline{dom(T)} \neq X$. Alors, par le théorème de Hahn Banach, il existe un $x_0' \neq 0$ tel que $\langle x, x_0' \rangle = 0$ pour tout $x \in dom(T)$.
Ce qui prouve alors la condition nécessaire. ■

Nous pouvons donc définir l'opérateur T' dual de T

Si $\overline{dom(T)} = X$, il existe $T' : dom(T') \subseteq Y' \to X'$
et qui associe à y' l'élément $Ty' = x'$ suivant la condition du théorème 1 (de manière unique)

Il découle immédiatement de la bilinéarité du crochet de dualité que T' est un opérateur linéaire tel que

$$\langle Tx, y' \rangle = \langle x, T'y' \rangle$$

pour tout $x \in dom(T)$ et tout $y \in dom(T')$, où $dom(T')$ est l'ensemble des y' tel qu'il existe x' pour satisfaire la condition du th. 1.

Théorème I.2

Si $dom(T) = X$ et $T \in \mathcal{L}(X, Y)$ (i.e. T lin (c°)
alors $T' \in \mathcal{L}(Y', X')$

En effet, nous définissons l'opérateur dual T' comme plus haut

Soit B un ensemble borné de X. Puisque T est continu, T.B $= \{ Tx \mid x \in B \}$ est un ensemble borné de Y. Mais $\langle Tx, y' \rangle = \langle x, x' \rangle$. Donc, si $y' \to 0$ dans Y', $x' \to 0$ dans X'. Ce qui implique la continuité de T' de Y' dans X' ■

Exemple: Soient $X = Y = \mathbb{R}^n$ normé par la norme ℓ^2.
Si $T \in \mathcal{L}(X,X)$, soit $Tx = y$ où $x = (x_1 \cdots x_n)$ et $y = (y_1 \cdots y_n)$
Alors $y_i = \sum_{j=1}^{n} t_{ij} x_j$
Donc

$$\langle Tx, z \rangle = \langle y, z \rangle = \sum y_i z_i = \sum_i \left(\sum_j t_{ij} x_j \right) z_i = \sum_j x_j \left(\sum_i t_{ij} z_i \right)$$

et donc $\langle Tx, z \rangle = \langle x, T'z \rangle$ et T' est la matrice T transposée.
L'opérateur dual généralise donc bien la transposée.

De plus, dans le cas continu, on peut énoncer le

Théorème I.3

> Soient X et Y deux espaces linéaires normés.
> Si $T \in \mathcal{L}(X,Y)$ on sait que $T' \in \mathcal{L}(Y', X')$
> De plus
> $$\|T\| = \|T'\|$$

Démonstration : Par la relation de définition $\langle Tx, y' \rangle = \langle x, x' \rangle$, on obtient

$$\|T'y'\| = \|x'\| = \sup_{\|x\| \leq 1} |\langle x, x' \rangle| = \sup_{\|x\| \leq 1} |\langle Tx, y' \rangle|$$

$$\leq \|y'\| \sup_{\|x\| \leq 1} \|Tx\| \leq \|y'\| . \|T\|$$

et donc $\|T'\| \leq \|T\|$.
D'autre part, $\forall x_0 \in X \ \exists f_0 \in Y' \ \|f_0\| = 1$ et $f_0(Tx_0) = \langle Tx_0, f_0 \rangle$

$= \|Tx_0\|$. Donc $f_0 = Tf_0$ satisfait $\langle x_0, f_0 \rangle = \|Tx_0\|$ et donc

$$\|Tx_0\| = \langle x_0, Tf_0 \rangle \leq \|T'\| \|f_0\| . \|x_0\| = \|T'\| . \|x_0\|$$

ce qui implique que $\|T\| \leq \|T'\|$ ■

On peut de plus démontrer que

> 1) Si T et S appartiennent à $\mathcal{L}(X,Y)$, alors $(\alpha T + \beta S)' = \alpha T' + \beta S'$
> 2) Soient T et S linéaires tels que $\mathrm{dom}(T)$, $\mathrm{dom}(S)$, $R(T)$ et $R(S)$ sont tous contenus dans X.
> Si $S \in \mathcal{L}(X,X)$ et $\overline{\mathrm{dom}(T)} = X$, alors $(ST)' = T'S'$
> De plus, si $\overline{\mathrm{dom}(TS)} = X$, $(TS)'$ est une extension de $S'T'$

On trouvera une preuve de ce théorème dans Ref (Y-1) p 195

---

§2 Opérateurs adjoints.

Plaçons nous maintenant dans le cas hilbertien. Soient X et Y deux espaces de Hilbert. Soit T un opérateur linéaire défini sur $\mathrm{dom}(T) \subseteq X$ et à valeurs dans Y. Supposons de plus que $\mathrm{dom}(T)$ est dense dans X et définissons alors T' le dual de T comme précédemment. Donc $\langle Tx, y' \rangle = \langle x, T'y' \rangle$ pour $x \in \mathrm{dom}(T)$ et $y' \in \mathrm{dom}(T')$. Appelons J_X la bijection (isomorphisme) de X dans son dual X', qui est la conséquence du théorème de Riesz.

Alors $\langle Tx, y' \rangle = y'(Tx) = (Tx, J_Y y')$

$\langle x, T'y' \rangle = (T'y')(x) = (x, J_X T'y')$

Comme $\langle Tx, y' \rangle = \langle x, T'y' \rangle$, on obtient $(Tx, J_Y y') = (x, J_X T'y')$

c'est à dire $(Tx, y) = (x, J_X T' J_Y^{-1} y)$

Dans le cas où $X = Y$, on obtient

$$T^* = J_X T' J_X^{-1}$$

qui est appelé l'opérateur adjoint de T

Exactement comme dans le cas des opérateurs duaux, on démontre

Théorème I 4

T^* existe ssi $\overline{dom(T)} = X$
Il est défini de la manière suivante :
soit $y \in X$ et $\left[y \in dom(T^*) \Leftrightarrow \exists y^* \in X \ (Tx, y) = (x, y^*) \ \forall x \in dom(T) \right]$
alors $T^* y = y^*$

Théorème I 5

Si $dom(T) = X$
alors $T \in \mathcal{L}(X,X) \Rightarrow T^* \in \mathcal{L}(X,X)$
et $\|T\| = \|T^*\|$

De nouveau, de plus amples renseignements peuvent être trouvés dans Ref (Y-1) p 196 sq.

IV Optimisation avec contraintes : version globale.

Après avoir examiné et s'être remis en mémoire la théorie élémentaire de la dualité, nous allons maintenant aborder un sujet qui a plus directement trait à l'optimisation au sens usuel du terme : la théorie des multiplicateurs de Lagrange.

Dans cette première partie, nous nous placerons dans l'espace des contraintes où nous interpréterons le multiplicateur de Lagrange comme un hyperplan. Dans la seconde partie, nous considérerons la théorie locale dans l'espace primal X.

§1 Multiplicateurs de Lagrange.

Nous allons considérer, dans les quelques sections suivantes, le problème suivant :

$$\begin{cases} \text{minimiser } f(x) \\ \text{sous la contrainte } x \in \Omega,\ G(x) \leqslant \Theta \end{cases} \qquad (1)$$

où Ω est un sous ensemble convexe d'un espace linéaire X, f est une forme convexe de Ω dans $\mathbb{R}$, et où G est une application convexe de Ω dans un espace linéaire normé Z possédant un cône positif P et Θ comme origine.

Nous examinerons le problème (1) et développerons la théorie des multiplicateurs de Lagrange en considérant la classe de problèmes

$$\begin{cases} \text{minimiser } f(x) \\ \text{sous la contrainte } x \in \Omega,\ G(x) \leqslant z \end{cases}$$

où z est un vecteur arbitraire de Z. La solution de ce problème dépend évidemment de z (problème perturbé)

Définissons

$$\Gamma = \{ z \mid \exists_\Omega x \quad G(x) \leq z \}$$

Nous avons immédiatement que Γ est convexe. En effet $z_1, z_2 \in \Gamma$ implique l'existence de $x_1, x_2 \in \Omega$ tels que $G(x_1) \leq z_1$ et $G(x_2) \leq z_2$. Donc, pour $\alpha \in]0,1[$, $G(\alpha x_1 + (1-\alpha) x_2) \leq \alpha z_1 + (1-\alpha) z_2$ par la convexité de G, ce qui implique que $\alpha z_1 + (1-\alpha) z_2 \in \Gamma$

Sur Γ définissons maintenant la fonctionnelle primale ω

$$\omega(z) = \inf \{ f(x) \mid x \in \Omega \wedge G(x) \leq z \}$$

Cette fonctionnelle peut prendre des valeurs infinies.
Le problème (1) peut être vu sous l'angle suivant : déterminer $\omega(0)$.

On a maintenant le

Théorème I_6

ω est convexe

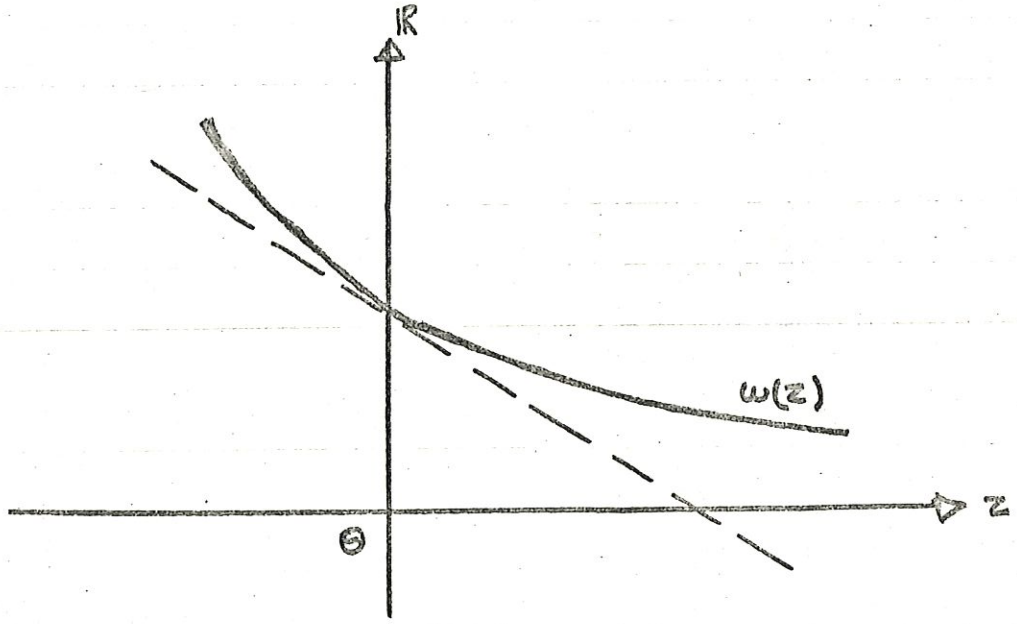
En effet, pour $\alpha \in]0,1[$

$$\begin{aligned}
\omega(\alpha z_1 + (1-\alpha) z_2) &= \inf \{ f(x) \mid x \in \Omega \wedge G(x) \leq \alpha z_1 + (1-\alpha) z_2 \} \\
&\leq \inf \{ f(x) \mid x = \alpha x_1 + (1-\alpha) x_2 , x_1 \in \Omega, x_2 \in \Omega \\
&\qquad , G(x_1) \leq z_1 , G(x_2) \leq z_2 \} \\
&\leq \alpha \inf \{ f(x_1) \mid x_1 \in \Omega, G(x_1) \leq z_1 \} \\
&\qquad + (1-\alpha) \inf \{ f(x_2) \mid x_2 \in \Omega , G(x_2) \leq z_2 \} \\
&\leq \alpha \omega(z_1) + (1-\alpha) \omega(z_2) \qquad \blacksquare
\end{aligned}$$

Il est de plus immédiat que

ω est décroissante

Typiquement, on peut représenter un w unidimensionel comme suit



Conceptuellement, le théorème des multiplicateurs de Lagrange résulte du fait que si w est convexe, il existe un hyperplan tangent à w(z) en Θ et qui reste "en-dessous" de w(z) partout où w est définie. Si l'on considère l'hyperplan comme une nouvelle horizontale, on voit que w est minimisé en Θ, ce qui est équivalent à dire qu'en ajoutant à w(z) une fonctionnelle linéaire appropriée $\langle z, z_0' \rangle$, le résultat $w(z) + \langle z, z_0' \rangle$ est minimum au point z = Θ. La fonctionnelle z_0' est le multiplicateur de Lagrange du problème.

Théorème I 7

Soient X et Y respectivement un espace linéaire et un espace linéaire normé muni d'un cône positif P.
Si $\Omega \subseteq X$ est convexe et si $\overset{0}{P} \neq \emptyset$,
Soit f convexe de Ω dans IR et G convexe de X dans Z.
Soit $\exists x_1 \in \Omega \; G(x_1) < \Theta$
Soit $\mu_0 = \inf f(x)$ (pour $x \in \Omega$ et $G(x) \leq \Theta$) fini (2.)

Alors, il existe $z_0' \geq \Theta$ dans Z' tel que

$$\mu_0 = \inf_{x \in \Omega} \{ f(x) + \langle G(x), z_0' \rangle \} \quad (3)$$

De plus, si l'infimum est atteint dans (3) en $x_0 \in \Omega$, $G(x_0) \leq \Theta$, c'est aussi le cas dans (3) et

$$\langle G(x_0), z_0' \rangle = 0 \quad (4)$$

Remarque : Le cône positif de Z' est défini comme suit

$$P^* = \{ x' \in X' \mid \langle x, x' \rangle \geq 0 \ \forall x \in P \}$$

Démonstration : Soit $W = \mathbb{R} \times Z$, où nous définissons

$$A = \{ (r,z) \mid \exists x \in \Omega \ r \geq f(x), z \geq G(x) \}$$

$$B = \{ (r,z) \mid r \leq \mu_0, z \leq \Theta \}$$

Comme f et G sont convexes, A et B le sont aussi (D'ailleurs A représente la région convexe située au-dessus du graphe de la fonctionnelle primale w). La définition de μ_0 implique que A ne contient aucun point intérieur à B.

Soit $N = -P$. Par hypothèse, il contient un point dans son intérieur. Donc l'intérieur de B n'est pas vide. On déduit alors du théorème de l'hyperplan séparant qu'il existe un élément non nul $w_0' = (r_0, z_0')$ de W' tel que

$$\forall (r_1,z_1) \in A \ \forall (r_2,z_2) \in B \quad r_0 r_1 + \langle z_1, z_0' \rangle \geq r_0 r_2 + \langle z_2, z_0' \rangle$$

De la nature de B on déduit immédiatement que $w_0' \geq \Theta$ i.e. $r_0 \geq 0$ et $z_0' \geq \Theta$. Montrons maintenant que $r_0 > 0$.

Le point $(\mu_0, \Theta) \in B$. Donc $r_0 r + \langle z, z_0' \rangle \geq r_0 \mu_0$ pour tout $(r,z) \in A$. Si $r_0 = 0$, on aurait en particulier $\langle G(x_i), z_0' \rangle \geq 0$ avec, de plus, $z_0' \neq 0$. Cependant, puisque $G(x_i) \in \mathring{N}$ et $z_0' \geq 0$, on a $\langle G(x_i), z_0' \rangle < 0$, ce qui est contradictoire.

Donc $r_0 > 0$ et, sous perte de généralité, nous le prendrons égal à -1.
Puisque le point $(\mu_0, 0)$ est arbitrairement près de A et de B, on obtient (avec $r_0 = 1$)

$$\mu_0 = \inf_{(r,z)\in A} [r + \langle z, z_0' \rangle] \leq \inf_{x\in\Omega} [f(x) + \langle G(x), z_0' \rangle]$$

$$\leq \inf_{\substack{x\in\Omega \\ G(x)\leq 0}} f(x) = \mu_0$$

Ce qui prouve la première partie du théorème.
De plus, s'il existe un x_0 tel que $G(x_0) \leq 0$, $\mu_0 = f(x_0)$, alors

$$\mu_0 \leq f(x_0) + \langle G(x_0), z_0' \rangle \leq f(x_0) = \mu_0$$

Ce qui implique alors que $\langle G(x_0), z_0' \rangle = 0$ ∎

Donnons maintenant une version algébrique de ce théorème.

Théorème I.F.1

> Sous les hypothèses du th I.F., supposons que x_0 fournit le minimum avec contrainte. Alors, $\exists z_0' \geq 0$ tel que
> $$L(x, z') = f(x) + \langle G(x), z' \rangle$$
> possède un point de selle en x_0, z_0' i.e
> $$\forall x, \forall z' \geq 0 \quad L(x_0, z') \leq L(x_0, z_0') \leq L(x, z_0')$$

En effet, choisissons z_0' comme dans le th I.F. (3) implique que $L(x_0, z_0') \leq L(x, z_0')$

Mais (4) donne

$$L(x_0, z') - L(x_0, z_0') = \langle G(x_0), z' \rangle - \langle G(x_0), z_0' \rangle = \langle G(x_0), z' \rangle \leq 0$$

■

---

§2 Sans les conditions de convexité.

Il est clair que sans les conditions de convexité et l'existence de points intérieurs, on ne peut généralement assurer l'existence d'un hyperplan séparant dans $\mathbb{R} \times Z$. Cependant, si cet hyperplan existe tout de même, la technique des multiplicateurs de Lagrange est toujours d'application.

Théorème I.8

> Soit $f : \Omega \longrightarrow \mathbb{R}$ où $\Omega \subseteq X$
> Soit $G : \Omega \longrightarrow Z$ normé, de cône positif P
> Si $\exists\, z_0' \in Z'$, $z_0' \geq \Theta$ $\exists\, x_0 \in \Omega$ tel que
> $\forall_{x \in \Omega}$ $\quad f(x_0) + \langle G(x_0), z_0' \rangle \leq f(x) + \langle G(x), z_0' \rangle$
> Alors x_0 est la solution de
> $$\begin{cases} \text{minimiser } f(x) \\ \text{sous la contrainte } G(x) \leq G(x_0),\ x \in \Omega \end{cases}$$

Démonstration: Si $\exists\, x_1 \in \Omega$ tel que $f(x_1) < f(x_0)$ et $G(x_1) \leq G(x_0)$, et, comme $z_0' \geq \Theta$, on déduit

$$\langle G(x_1), z_0' \rangle \leq \langle G(x_0), z_0' \rangle$$

et donc

$$f(x_1) + \langle G(x_1), z_0' \rangle \leq f(x_0) + \langle G(x_0), z_0' \rangle$$

ce qui contredit l'hypothèse ■

Théorème I.9

Soient H, X, Z, Ω, P, f et G comme supra.
Soit P fermé, $L(x,z') = f(x) + \langle G(x), z' \rangle$
Si $\exists z_0' \in Z', z_0' \geq \Theta$, $\exists x_0 \in \Omega$ tel que
$$L(x_0, z') \leq L(x_0, z_0') \leq L(x, z_0')$$
pour $\forall x \in \Omega$ et $\forall z' \in Z', z' \geq \Theta$
Alors x_0 est la solution de
$$\begin{cases} \text{minimiser } f(x) \\ \text{sous la contrainte } G(x) \leq \Theta, x \in \Omega \end{cases}$$

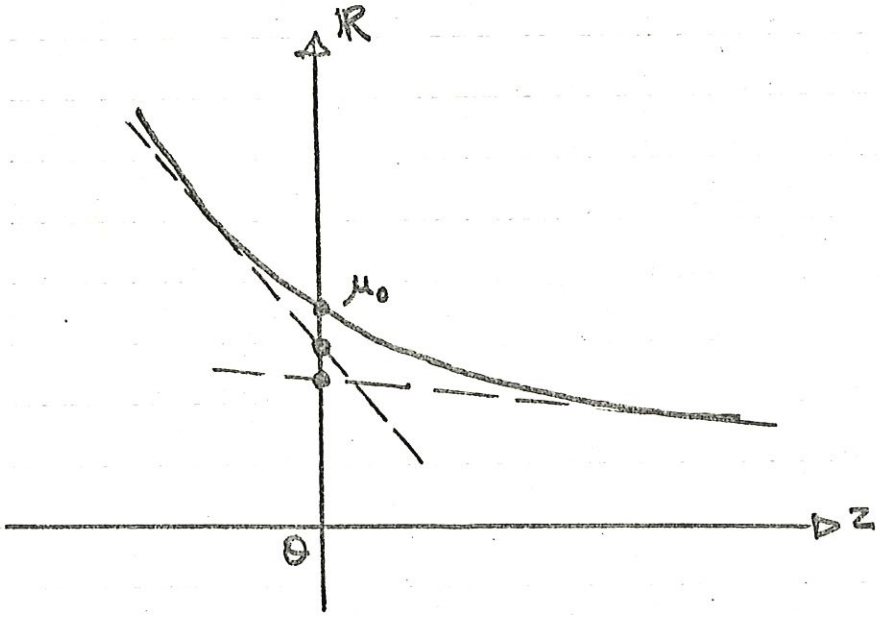
Démonstration : En particularisant la condition de point de selle, on obtient
$$\langle G(x_0), z' \rangle \leq \langle G(x_0), z_0' \rangle \quad \forall z' \in Z', z' \geq \Theta$$
Donc, pour $z_1' \geq \Theta$
$$\langle G(x_0), z_1' + z_0' \rangle \leq \langle G(x_0), z_0' \rangle$$
a qui implique $\langle G(x_0), z_1' \rangle \leq 0$
On peut en en déduire que $G(x_0) \leq \Theta$.
La condition de point de selle implique alors $\langle G(x_0), z_0' \rangle = 0$
Supposons maintenant que $x_1 \in \Omega$ et $G(x_1) \leq \Theta$. Alors,
$$f(x_0) = f(x_0) + \langle G(x_0), z_0' \rangle \leq f(x_1) + \langle G(x_1), z_0' \rangle \leq f(x_1)$$
Donc x_0 est bien la solution cherché ∎

---

§3 Dualité

Replaçons nous dans les hypothèses de convexité et considérons toujours
$$\begin{cases} \text{minimiser } f(x) \\ \text{sous la contrainte } G(x) \leq \Theta, x \in \Omega \end{cases}$$

Définissons comme supra $w(z) = \inf\{f(x) \mid G(x) \leq z, x \in \Omega\}$ et posons $\mu_0 = w(\Theta)$. Le principe de dualité est basé sur le fait que μ_0 est l'intersection maximum de l'axe vertical avec les hyperplans qui sont en dessous de $w(z)$. Ce maximum est atteint pour l'hyperplan représentant le multiplicateur de Lagrange du problème.



Pour exprimer ce fait analytiquement, introduisons la fonctionnelle duale φ (correspondant à w) définie sur le cône positif du Z', soit P^*,

$$\varphi(z') = \inf_{x \in \Omega} \{f(x) + \langle G(x), z' \rangle\}$$

En général, φ n'est pas finie partout dans P^* mais la région où elle est finie est convexe. De plus, on voit aisément que φ elle-même est concave.

En effet $\varphi(\alpha z_1' + (1-\alpha) z_2') = \inf_{\Omega} \{f(x) + \langle G(x), \alpha z_1' + (1-\alpha) z_2' \rangle\}$

$$\geq \inf_{\Omega} \{\alpha f(x) + \langle G(x), \alpha z_1' \rangle + (1-\alpha) f(x) + \langle G(x), (1-\alpha) z_2' \rangle\}$$

$$\geq \alpha \inf_{\Omega} \{f(x) + \langle G(x), z_1' \rangle\} + (1-\alpha) \inf_{\Omega} \{f(x) + \langle G(x), z_2' \rangle\} \quad \square$$

Voyons de plus qu'on a le

Théorème I.10

$$\varphi(z') = \inf_{z \in \Gamma} \{ w(z) + \langle z, z' \rangle \}$$

où Γ est défini comme supra $\Gamma := \{ z \mid \exists x \in \Omega \;\; \theta(x) \leq z \}$.

Démonstration Pour tout $z' \geq 0$ et tout $z \in \Gamma$, on a

$$\varphi(z') = \inf_{x \in \Omega} \{ f(x) + \langle \theta(x), z' \rangle \} \leq \inf_{x \in \Omega} \{ f(x) + \langle z, z' \rangle \mid \theta(x) \leq z, x \in \Omega \}$$

$$= w(z) + \langle z, z' \rangle$$

D'autre part, pour tout $x_1 \in \Omega$, on obtient, en posant $z_1 = \theta(x_1)$

$$f(x_1) + \langle \theta(x_1), z' \rangle \geq \inf \{ f(x) + \langle z_1, z' \rangle \mid \theta(x) \leq z_1, x \in \Omega \}$$

$$= w(z) + \langle z_1, z' \rangle$$

et donc $\varphi(z') \geq \inf_{z \in \Gamma} \{ w(z) + \langle z, z' \rangle \}$ ∎

L'élément $(-1, z')$ de $\mathbb{R} \times Z'$ détermine une famille d'hyperplans dans $\mathbb{R} \times Z$, chaque hyperplan étant formé des points (r, z) satisfaisant $r + \langle z, z' \rangle = k$ où k est une constante. Le th. 10 dit que pour $k = \varphi(z')$, cet hyperplan supporte l'ensemble $[w, \Gamma]$, la région située au dessus du graphe de w. De plus, en $z = 0$, on obtient $r = \varphi(z')$; donc $\varphi(z')$ est égal à l'intersection de cet hyperplan avec l'axe vertical, ce qui nous ramène à la discussion géométrique faite plus haut.

Nous sommes à même, maintenant, de prouver le théorème de dualité de Lagrange.

Théorème I.11

Soit $f : \Omega (\subset X) \longrightarrow \mathbb{R}$ convexe, avec Ω convexe
Soit $G : X \longrightarrow Z$ normé, avec G convexe
Si $\exists x_1$, $G(x_1) < \Theta$ et $\mu_0 = \inf \{ f(x) \mid G(x) \leq \Theta, x \in \Omega \}$ est fini.
Alors

$$\inf_{\substack{G(x) \leq \Theta \\ x \in \Omega}} f(x) = \max_{z' \geq \Theta} \varphi(z') \qquad (5)$$

et le maximum à droite est atteint en $z_0' \geq \Theta$.
Si, de plus, l'infimum à gauche est atteint en $x_0 \in \Omega$, alors

$$\langle G(x_0), z_0' \rangle = 0$$

et x_0 minimise $f(x) + \langle G(x), z_0' \rangle$ sur Ω

Démonstration : Pour tout $z' \geq \Theta$, on a

$$\inf_{x \in \Omega} \{ f(x) + \langle G(x), z' \rangle \} \leq \inf_{\substack{x \in \Omega \\ G(x) \leq \Theta}} \{ f(x) + \langle G(x), z' \rangle \} \leq \inf_{\substack{x \in \Omega \\ G(x) \leq \Theta}} f(x) = \mu_0$$

Donc, le membre de droite de (5) est $\leq \mu_0$.
Cependant, le théorème I.7 établit l'existence d'un z_0' qui donne l'égalité.
Ce même théorème I.7 fournit d'ailleurs le reste de la preuve ∎

V Optimisation avec contraintes : version locale

Examinons maintenant le point de vue local de l'optimisation avec contraintes.

Introduisons d'abord la différentielle de Fréchet.

> Soient X et Y deux espaces linéaires normés.
> Soit D ouvert $\subseteq X$
> Soit $T : D \to Y$
> Si $\forall x \in D \ \forall h \in X \ \exists \delta T(x,h) \in Y$ linéaire et continu en h tel que
>
> $$\lim_{\|h\| \to 0} \frac{\|T(x+h) - T(x) - \delta T(x,h)\|}{\|h\|} = 0$$
>
> alors, T est Fréchet différentiable en x et $\delta T(x,h)$ est la différentielle de Fréchet de T en x avec h comme incrément.

Cette différentielle peut encore s'écrire $\delta T(x,h) = T'(x)h$ où $T'(x)$ est la dérivée de T au sens de Fréchet.

> Si T est une transformation continûment différentiable au sens de Fréchet de l'ouvert $D \subseteq X$ Banach vers Y Banach, si $x_0 \in D$ est tel que $T'(x_0)$ est surjectif sur Y, alors le point x_0 est appelé point régulier de T

Nous énonçons ensuite le théorème de l'inverse généralisé, sans en donner de preuve :

Théorème

Soit x_0 un point régulier de la transformation T du Banach X dans le Banach Y. Alors, il existe un voisinage $N(y_0)$ du point $y_0 = T(x_0)$ et une constante K telle que l'équation $T(x) = y$ admette un solution pour tout $y \in N(y_0)$, satisfaisant

$$\|x - x_0\| \leq K \|y - y_0\|$$

Une démonstration détaillée se trouve dans Ref (L-1) p240 sq.

§1 Contraintes sous forme d'égalités.

Notre but, dans ce paragraphe, est de développer des conditions nécessaires pour un extremum de $f(x)$ sous les contraintes $H(x) = 0$ où $f : X \to \mathbb{R}$ avec X Banach et $H : X \to Z$ avec Z Banach.

Commençons par le

Théorème I.12

Si f admet un minimum local (sous la contrainte $H(x) = 0$) au point x_0 et si f et H sont continuement différentiables au sens de Fréchet dans un ouvert contenant x_0 et si, enfin, x_0 est un point régulier de H, alors $f'(x_0)h = 0$ pour tout h tel que $H'(x_0)h = 0$.

Démonstration : Considérons la transformation $T : X \to \mathbb{R} \times Z$ définie comme suit : $T(x) = (f(x), H(x))$. Si il existe un h tel que $H'(x_0)h = 0$ et $f'(x_0)h \neq 0$, alors $T'(x_0) = (f'(x_0), H'(x_0)) : X \to \mathbb{R} \times Z$

est surjective car $H'(x_0)$ l'est sur Z par hypothèse. Par le théorème de l'inverse généralisé, on pourrait déduire que $\forall \varepsilon > 0 \; \exists x \; \exists \delta > 0$ tel que $\|x - x_0\| < \varepsilon$ et $T(x) = (f(x_0) - \delta, \Theta)$; ce qui contredit l'hypothèse que x_0 minimise $f(x)$ localement ∎

Nous pouvons maintenant parler à nouveau des multiplicateurs de Lagrange.

Théorème I.13

> Si la fonctionnelle f continûment différentiable au sens de Fréchet admet un extremum local, sous la contrainte $H(x) = \Theta$, au point x_0, alors il existe $z_0' \in Z'$ tel que
> $$L(x) = f(x) + \langle H(x), z_0' \rangle$$
> est stationnaire en x_0 ; i.e. $f'(x_0) + z_0'[H'(x_0)] = \Theta$

Démonstration : D'après le th I.12 il est clair que $f'(x_0)$ est orthogonal au noyau de $H'(x_0)$. Comme l'image de $H'(x_0)$ est fermée, on peut déduire $f'(x_0) \in R[(H'(x_0))']$.

Donc, il existe $z_0' \in Z'$ tel que $f'(x_0) = -(H'(x_0))' z_0'$, ce qui est équivalent à $f'(x_0) + \langle H'(x_0), z_0' \rangle = 0$ ∎

Supposons maintenant que x_0 n'est plus un point régulier.

Théorème I.14

> Dans les hypothèses du th 13., sauf que $R(H'(x_0))$ est seulement fermé dans Z, il existe un élément non nul de $R \times Z'$, soit (r_0, z_0') tel que $r_0 f(x) + z_0'[H(x)]$ est stationnaire en x_0.

Démonstration : Si x_0 est régulier, on choisit $r_0 = 1$ et on applique le th 13. Si x_0 n'est pas régulier, soit $M = R(H'(x_0))$. Il existe alors un point $z \in Z$ tel que $\inf_{m \in M} \|z - m\| > 0$ et, donc, on peut déduire qu'il existe $z_0' \in M^+$, $z_0' \neq \Theta$. Or $M^+ = R(H'(x_0))^+ = \mathcal{N}(H'(x_0))$ (Rappel. $M^+ = \{z' \mid \forall x \in M \ \langle x, x' \rangle = 0\}$). Il suffit alors de choisir $(0, z_0')$ ∎

§2 Contraintes sous forme d'inégalités.

Nous nous intéressons à présent au problème suivant

$$\begin{cases} \text{minimiser } f(x) \\ \text{sous la contrainte } G(x) \leq \Theta \end{cases}$$

où f est définie sur X espace linéaire et $G : X \to Z$, espace linéaire normé de cône positif P.

Définissons aussi les points réguliers. Dans ce but, nous introduisons la différentielle de Gateaux

> Soit $x \in \mathrm{dom}(T) \subseteq X$, et $h \in X$ quelconque.
> Si $\delta_G T(x,h) = \lim_{\substack{\alpha \to 0 \\ \alpha > 0}} \frac{1}{\alpha}[T(x + \alpha h) - T(x)]$ existe, elle est appelée la différentielle de Gateaux de T en x d'incrément h.

Remarque : T est Fréchet-différentiable $\Rightarrow$ T est Gateaux-différentiable

> Soit $\mathring{P} \neq \phi$, G Gateaux différentiable telle que $\delta_G G(x,h)$ est linéaire en h. $x_0 \in X$ est un point régulier de $G(x) \leq \Theta$ si $G(x_0) \leq \Theta$ et $\exists h \mid G(x_0) + \delta_G G(x_0, h) < 0$

Il est à noter que traduire $H(x)=0$ par $P=\{\Theta\}$ et $-H(x)\leq\Theta$ devient impossible car il faut $\mathring{P}\neq\phi$. Les contraintes exprimés sous forme d'égalité ne peuvent donc être reprises dans ce cadre.

Prouvons maintenant le théorème important

Théorème I.15 (Kuhn-Tucker généralisé)

> Dans les hypothèses précédentes, supposons que f soit une fonction sur $\mathbb{R}$ G-différentiable, et que G $X\to Z$ soit G-différentiable, et que $\delta_G f$ et $\delta_G G$ soient linéaires en h. Supposons de plus que x_0 minimise f sous la contrainte $G(x)\leq\Theta$ et que x_0 est un point régulier de $G(x)\leq\Theta$. Alors, $\exists\, z_0'\in Z', z_0'\geq\Theta$ tel que
> $$L = f(x) + \langle G(x), z_0'\rangle$$
> soit stationnaire en x_0. De plus, $\langle G(x_0), z_0'\rangle = 0$.

Démonstration Dans $W=\mathbb{R}\times Z$, définissons

$A=\{(r,z) \mid r\geq\delta_G f(x_0,h),\ z\geq G(x_0)+\delta_G G(x_0,h) \text{ pour un } h\in X\}$

$B=\{(r,z) \mid r\leq 0,\ z\leq\Theta\}$

A et B sont évidemment convexes. Puisque $\mathring{P}\neq 0$, on a $\mathring{B}\neq\phi$.

A ne contient pas de point intérieur à B. En effet, si $(r,z)\in A$ avec $r<0$ et $z<\Theta$, alors $\exists h\in X$ tel que

$$\delta_G f(x_0,h)<0 \qquad G(x_0)+\delta_G G(x_0,h)<\Theta$$

Le point $G(x_0)+\delta_G G(x_0,h)$ est alors le centre d'un sphère ouverte contenue dans le cône négatif N de Z. Soit ρ le rayon de cette sphère. Alors, pour $\alpha\in]0,1[$, le point $\alpha[G(x_0)+\delta_G G(x_0,h)]$ est le centre d'une sphère ouverte de rayon $\alpha\rho$ contenue dans N; donc le point $(1-\alpha)G(x_0)+\alpha[G(x_0)+\delta_G G(x_0,h)]$ $= G(x_0)+\alpha\delta_G G(x_0,h)$ s'y trouve aussi. Puisque pour h fixé

$$\|G(x_0+\alpha h)-G(x_0)-\alpha\delta G(x_0,h)\| = o(\alpha)$$

il est donc clair que $G(x_0+\alpha h) < \Theta$ pour α suffisamment petit.
On peut reproduire un argument similaire pour prouver que $f(x_0+\alpha h) < f(x_0)$ pour α petit, ce qui contredit l'optimalité de x_0. Donc $A \cap \mathring{B} = \phi$.

Employons à nouveau le théorème de l'hyperplan séparant A et B. On déduit $\exists\, r_0, z_0', \delta$ tels que

$$\forall (r,z) \in A \qquad r_0 r + \langle z, z_0' \rangle \geq \delta$$
$$\forall (r,z) \in B \qquad r_0 r + \langle z, z_0' \rangle \leq \delta$$

Comme $(0,\Theta) \in A \cap B$, on a $\delta = 0$.
De la nature de B, on déduit que $r_0 \geq 0$ et $z_0' \geq \Theta$. De plus l'hyperplan ne peut être vertical puisque $\exists h \; G(x_0) + \delta_G G(x_0,h) < \Theta$. Donc, posons $r_0 = 1$.

On déduit aussi de la propriété de séparation

$$\forall h \in X \qquad \delta f(x_0,h) + \langle G(x_0) + \delta_G G(x_0,h), z_0' \rangle \geq 0$$

En choisissant $h = \Theta$, on obtient $\langle G(x_0), z_0' \rangle \geq 0$. Mais comme $G(x_0) \leq \Theta$ et $z_0' \geq \Theta$, on a aussi $\langle G(x_0), z_0' \rangle \leq 0$. Donc $\langle G(x_0), z_0' \rangle = 0$.
Reste donc

$$\forall h \in X \qquad \delta_G f(x_0,h) + \langle \delta_G G(x_0,h), z_0' \rangle \geq 0$$

Comme les G-différentielles sont linéaires en h, il est nécessaire que

$$\delta_G f(x_0,h) + \langle \delta_G G(x_0,h), z_0' \rangle = 0 \qquad \blacksquare$$

VI Le principe du maximum de Pontryagine.

Dans ce paragraphe, nous allons aborder un théorème qui est à la base de bien des applications, spécialement en contrôle optimal, comme nous le verrons plus tard.

Nous développerons la théorie d'un point de vue abstrait, en explicitant les opérateurs duaux.

Posons X et U des espaces linéaires normés et $g[x,u]$ une fonction (de coût) sur $X \times U \to \mathbb{R}$. Considérons de plus une contrainte de la forme $A[x,u] = \Theta$ où $A : X \times U \to X$. Cet opérateur A peut représenter un ensemble d'équations différentielles, intégrales, aux différences, aux dérivées partielles, etc.

Supposons que $A[x,u] = \Theta$ définit une fonction $x(u)$ implicite unique. Supposons, de plus, que A et g sont différentiables au sens de Fréchet par rapport à x et que $A_x[x,u]$ et $g_x[x,u]$ sont continues sur $X \times U$. Supposons enfin que

$$\| x(u) - x(v) \| \leq K \| u - v \|$$

Notre problème est de minimiser $J = g[x,u]$ sous la contrainte $A[x,u] = \Theta$ avec $u \in \Omega \subseteq U$. Il est clair que $J = g[x,u] = g[x(u),u] = J[u]$.

Introduisons maintenant le Lagrangien

$$\begin{aligned} L[x,u,\lambda'] &= \lambda' A[x,u] + g[x,u] \\ &= \langle A[x,u], \lambda' \rangle + g[x,u] \end{aligned}$$

pour $x \in X$, $u \in U$, $\lambda' \in X'$. Nous pouvons alors démontrer le

Théorème I.6

> Pour tout $u \in \Omega$, soit λ' une solution de l'équation
> $$\lambda' A_x[x,u] + g_x[x,u] = \Theta$$
> Alors, pour $v \in \Omega$,
> $$J(u) - J(v) = L[x(u),u,\lambda'] - L[x(v),v,\lambda'] + o(\|u-v\|)$$

Démonstration : Par définition

$$\begin{aligned} J[u] - J[v] &= g[x(u),u] - g[x(v),v] \\ &= g[x(u),u] - g[x(u),v] + g[x(u),v] - g[x(v),v] \\ &= g[x(u),u] - g[x(u),v] + g_x[x(v),u] \cdot [x(u) - x(v)] \\ &\quad + (g_x[x(v),v] - g_x[x(v),u])[x(u) - x(v)] + o(\|x(u) - x(v)\|) \end{aligned}$$

$$J[u]-J[v] = g[x(u),u]-g[x(u),v]+g_x[x(u),u][x(u)-x(v)] + o(\|u-v\|)$$

par la continuité de g_x et la condition $\|x(u)-x(v)\| \leq K\|u-v\|$.

De même

$$\|A[x(u),u]-A[x(u),v]-A_x[x(u),u][x(u)-x(v)]\| = o(\|v-u\|)$$

Donc

$$\begin{aligned} L[x(u),u,\lambda']-L[x(u),v,\lambda']+o(\|u-v\|) &= \lambda' A[x(u),u]+g[x(u),u]-\lambda' A[x(u),v] \\ &\quad -g[x(u),v]+o(\|v-u\|) \\ &= \lambda' A_x[x(u),u][x(u)-x(v)]+g[x(u),u]-g[x(u),v]+o(\|u-v\|) \\ &= g[x(u),u]-g[x(u),v]+g_x[x(u),u][x(u)-x(v)]+o(\|v-u\|) \\ &= J[u]-J[v] \quad \blacksquare \end{aligned}$$

Appliquons maintenant ce résultat au système d'équations différentielles ordinaires de la forme

$$\dot{x}(t) = f(x(t),u(t)) \qquad x(t_0)=x_0$$

et à la fonction de coût

$$J = \int_{t_0}^{t_1} \ell(x(t),u(t))\,dt$$

où nous supposons que les fonctions f et ℓ sont continûment différentiables par rapport x, que f satisfait

$$\|f(x,u)-f(y,v)\| \leq M[\|x-y\|+\|u-v\|]$$

sachant que f, x et u prennent leurs valeurs dans l'espace $\mathbb{R}^n$ et $\mathbb{R}^m$ respectivement. En fait nous restreignons $u(t) \in \Omega \subseteq \mathbb{R}^m$ et $u(t)$ est différentiable par morceaux sur $[t_0,t_1]$.

Dans ces conditions, nous pouvons aborder le principe de Pontryagine.

Théorème I.17

Soient x_0 et u_0 optimaux pour le problème

$$\min J = \min \int_{t_0}^{t_1} \ell(x,u)\, dt$$

sous la contrainte $\dot{x}(t) = f(x(t),u(t))$ avec $x(t_0)$ donné, $u(t) \in \Omega$.
Soit λ la solution de l'équation

$$-\dot{\lambda}(t) = f_x^t \lambda(t) + \ell_x^t \qquad \lambda(t_1) = 0$$

où les dérivées partielles sont évaluées le long de la trajectoire optimale, et définissons le Hamiltonien

$$H[x,u,\lambda,t] = \lambda^t(t) f(x,u) + \ell(x,u).$$

Alors, pour tout $t \in [t_0,t_1]$

$$H(x_0(t), u_0(t), \lambda(t), t) \leq H(x_0(t), u(t), \lambda(t), t)$$

pour tout u à valeurs dans Ω.

Démonstration : Pour une question de facilité, nous supposerons $m = 1$, i.e. les contrôles sont des fonctions scalaires.
Prenons $X = C^n[t_0,t_1]$ et $U = \{u^n$ continus par morceaux sur $[t_0,t_1]\}$ muni de la norme L^1

Posons $A[x,u] = x(t) - x(t_0) - \int_{t_0}^{t} f(x(\tau),u(\tau))\, d\tau$

et $g[x,u] = \int_{t_0}^{t_1} \ell(x,u)\, dt$

On peut voir que A et g sont continûment différentiables au sens de Fréchet par rapport à x.

Si $x, x+\delta x$ correspond à $u, u+\delta u$ dans $\{(x,u) \mid A[x,u] = 0\}$, alors

$$\|\delta x(t)\|_{\mathbb{R}^n} \leq \int_{t_0}^{t} |H| \{\|\delta x(\tau)\|_{\mathbb{R}^n} + |\delta u(\tau)|\}\, d\tau$$

a qui fournit [par le théorème de Gronwall],

$$\|\delta x(t)\|_{\mathbb{R}^n} \leq M e^{M(t_1-t_0)} \int_{t_0}^{t_1} |\delta u(\tau)|\, d\tau$$

et donc $\|\delta x\| \leq K \|\delta u\|$ et la transformation A satisfait la condition du théorème précédent.

De plus on a bien que $X \subseteq L^2$ qui est Hilbert. Donc $\lambda' \mapsto \lambda$ l'élément de X associé à λ' dans X' par le théorème de Riesz. La forme de dualité se réduit alors au produit scalaire.
On a alors que $-\dot\lambda(t) = f_x^t \lambda + \ell_x^t$ est bien équivalent, après intégration, à $\lambda^* A_x[x(u), u] + g_x[x(u), u] = 0$, pour $t \in [t_0, t_1]$
De même $\int_{t_0}^{t_1} H[x, u, \lambda]\, dt$ est identique au Lagrangien $L[x, u, \lambda']$ à part un terme $\int_{t_0}^{t_1} \dot x(t)\, \lambda(t)\, dt$ qui n'est pas important puisqu'il ne dépend pas de u explicitement.
Le théorème précédent nous donne

$$J[u_0] - J[u] = \int_{t_0}^{t_1} \big[H[x_0, u_0, \lambda] - H[x_0, u, \lambda]\big] dt + o(\|u - u_0\|)$$

Voyons maintenant que cette équation implique la minimisation du Hamiltonien. Supposons l'inverse : il existe un $\bar t \in [t_0, t_1]$ et $\bar u(\bar t) \in \Omega$ tel que

$$H[x(\bar t), u_0(\bar t), \lambda(\bar t)] > H[x(\bar t), \bar u(\bar t), \lambda(\bar t)]$$

Comme u est continu par morceaux et que x est continu, ainsi que λ, f et ℓ, il existe un intervalle $[t', t'']$ qui contient $\bar t$ et il existe un $\epsilon > 0$ tel que

$$H[x(t), u_0(t), \lambda(t)] - H[x(t), \bar u(t), \lambda(t)] > \epsilon$$

pour tout $t \in [t', t'']$.
Choisissons $u(t)$ la fonction continue par morceaux égale à $u_0(t)$ en dehors de $[t', t'']$ et à $\bar u(t)$ sur $[t', t'']$. Nous obtenons alors

$$J[u_0] - J[u] > \epsilon (t'' - t') + o(\|u - u_0\|).$$

Or $t''-t'$ est tel que $\|u-u_0\| = o(t''-t')$. Donc, en choisissant $[t', t'']$ suffisamment petit, on peut rendre $J[u_0] - J[u]$ positif, ce qui contredit l'optimalité de u_0. ∎

Remarque

Il est clair que l'équation $-\dot{\lambda}(t) = f_x^t \lambda + \ell_x^t$ équivaut à

$$-\frac{\partial H}{\partial x} = \dot{\lambda}$$

ce qui fournit parfois un moyen de calculer λ effectivement.

Chapitre II

Les systèmes dynamiques contrôlés.

I Introduction

Comme annoncé, ce chapitre sera consacré principalement à la description des systèmes dynamiques contrôlés, ou systèmes dynamique à contrôle. D'abord d'un point de vue intuitif et pratique, ensuite d'un point de vue plus formel, nous tenterons de définir cette classe de systèmes si fréquents dans les applications, en même temps que théoriquement intéressants pour eux-mêmes.

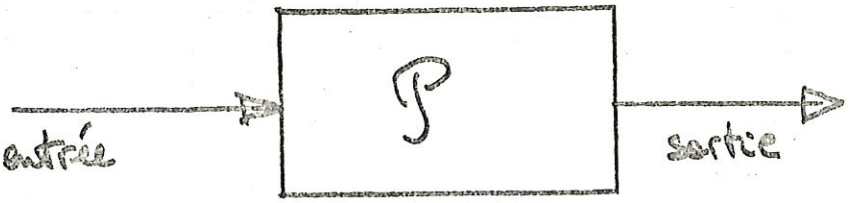
Les outils développés dans les pages précédentes ne nous serviront pas directement dans cette description, mais plutôt dans le paragraphe consacré aux méthodes de résolution et surtout dans le chapitre plus technique, traitant de la méthode de Ritz Galerkin.

Mais abordons d'abord la description.

II Les systèmes dynamiques contrôlés : vision intuitive

La connaissance du monde physique est basée sur l'expérience et l'abstraction. Tous les scientifiques connaissent la signification pratique des mots : induction et déduction. D'autre part, autant que la synthèse, l'analyse est un élément indispensable de cette connaissance. Le scientifique, en face d'un phénomène, tente d'abord de le cerner pour mieux le comprendre. C'est ce que nous allons faire.

Commençons par nous donner un système physique P, une de ces "boîtes noires" proverbiales, auquel nous pouvons fournir des "signaux" d'entrée et duquel nous pouvons observer le comportement par le biais de "signaux" de sortie.



Notre objectif final est la détermination d'un signal d'entrée, ou "input", qui produira, à travers le système P, un signal de sortie, ou "output", possédant certaines caractéristiques que nous nous donnons d'avance, tentant ainsi de minimiser "le coût" de notre opération, dans la mesure où le signal de sortie est conforme à nos désirs.

Dans les pages qui vont suivre, nous nous intéresserons à une classe de systèmes pour lesquels on peut construire des modèles mathématiques adéquats, appelés systèmes dynamiques.

Effectuons d'abord une petite expérience mentale. Partant d'un "moment initial", nous appliquons un signal à l'entrée de notre système $\mathcal{P}$, et cela jusqu'à un instant futur. Pendant ce laps de temps, que nous appellerons intervalle d'observation, nous observons la sortie du système. Une question naturelle se pose alors : " Partant du même instant initial, nous appliquons à $\mathcal{P}'$, système en tous points identique à $\mathcal{P}$, un signal d'entrée semblable à celui appliqué à $\mathcal{P}$. Observerons nous, pour $\mathcal{P}'$, les mêmes signaux de sortie que pour $\mathcal{P}$? "

L'expérience nous enseigne qu'il faut, en plus, tenir compte de l'état du système ($\mathcal{P}$ ou $\mathcal{P}'$) au moment initial. L'output du système dépend donc de son état initial.

Nous venons d'introduire, de façon naturelle, le concept supplémentaire d' "état du système". En fait, dans la description du système dynamique, nous utiliserons les concepts d'input, d'output et d'état. Formalisons un peu : appelons $u(t)$ le signal d'input en fonction du temps ; de même appelons $y(t)$ le signal d'output en fonction du temps et $x(t)$ l'état du système en fonction du temps.

Au temps initial t_0, nous commençons donc d'introduire $u_{(t_0,t)}(t) = u(t)$ pour tout t dans $[t_0, t_1]$ où t_1 est l'instant final, et nous observons $y_{(t_0,t)}(t)$ sur le même intervalle.

De plus, nous allons faire quelques hypothèses sur le système lui-même.

1. Nous supposerons que le système est entièrement déterministe.
2. Nous supposerons qu'il n'est pas anticipatif, i.e. que l'input, l'output et l'état à l'instant t ne dépendent pas de leurs valeurs respectives aux instants postérieurs.

Précisons un peu plus encore notre modèle : donnons nous des relations qui unissent l'output à l'input, via l'état et l'état initial.

$$y(t) = g[x(t_0), u_{(t_0,t)}, (t_0,t)]$$
$$x(t) = f[x(t_0), u_{(t_0,t)}, (t_0,t)]$$

où la première équation est appelée équation d'output et la seconde équation d'état.

A présent, passons à une définition plus formelle de ces concepts.

III Les systèmes dynamiques contrôlés : définition formelle.

Soit $T \subseteq \mathbb{R}$, (Σ, d) un espace métrique, de même de $(\Omega, \hat{d})$, soit U un ensemble de fonctions continues par morceaux $T \longrightarrow \Omega$.
Soit $x(t) : T \longrightarrow \Sigma$, soit $g : \Sigma \times \Omega \times T \longrightarrow \mathbb{R}^p$.
Si $t_0 \in T$ et $t_1 \in T$ et $t_0 \leq t_1$, soit $]t_0, t] = \{t \in T \mid t_0 < t \leq t_1\}$
Nous utiliserons la notation $u_{]t_0,t]}$ pour la restriction de $u \in U$ à l'intervalle $]t_0, t]$.

Si $t \in T$, $u \in U$, $x(t) \in \Sigma$ alors $g[x(t), u(t), t]$ est un élément de $\mathbb{R}^p$ bien défini que nous notons $y(t)$, i.e.

$$y(t) = g[x(t), u(t), t]$$

Nous posons aussi

$$y_{]t_0,t]} = \hat{g}[x(t_0); u_{]t_0,t]}]$$

Après ces quelques précisions, nous sommes maintenant en mesure d'envisager les axiomes qui définissent notre système avec exactitude.

Axiome 1

Pour $x(t_0) \in \Sigma$, pour tout $t_0 \in T$, tout $t \in T$ avec $t \geq t_0$, et pour tout $u_{]t_0,t]}$ avec $u \in U$, la connaissance de $x(t_0)$ et $u_{]t_0,t]}$ détermine $y_{]t_0,t]}$ de manière unique.

En particulier, si u et v appartiennent à U et satisfont

$$u_{]t_0,t]} = v_{]t_0,t]}$$

alors

$$\hat{g}\left[x(t_0); u_{]t_0,t]}\right] = \hat{g}\left[x(t_0); v_{]t_0,t]}\right]$$

Axiome 2

Si $t_0 < \hat{t} < t$ sont des éléments de T, si $x(t_0) \in \Sigma$, si $\Sigma[x(t_0), u, \hat{t}]$ représente l'ensemble des $x(\hat{t})$ de Σ qui satisfont l'équation

$$y_{]\hat{t},t]} = \hat{g}\left[x(t_0); u_{]t_0,\hat{t}]}\right] = \hat{g}\left[x(\hat{t}), u_{]\hat{t},t]}\right]$$

et si u^* est un élément donné de U, alors l'intersection des ensembles $\Sigma[x(t_0), u, \hat{t}]$ où $u \in U$ avec $u_{]t_0,\hat{t}]} = u^*_{]t_0,\hat{t}]}$ n'est pas vide, i.e.

$$\bigcap_{\substack{u \in U \\ u_{]t_0,\hat{t}]} = u^*_{]t_0,\hat{t}]}}} \Sigma[x(t_0), u, \hat{t}] \neq \phi$$

En particulier, aucun $\Sigma[x(t_0), u, \hat{t}]$ n'est vide. L'axiome assure l'existence d'au moins un élément de Σ pour toute paire $(u_{]\hat{t},t]}, y_{]\hat{t},t]})$

On peut alors montrer que ces deux axiomes impliquent l'existence d'une fonction $\phi[t, u_{]t_0,t]}, x(t_0)]$ telle que

$$x(t) = \phi[t, u_{]t_0,t]}, x(t_0)]$$

(cfr Référence (Z-1))

Axiome 3

Les fonctions g, $\hat{g}$ et ϕ sont continues en toutes leurs variables. Pour ce qui est de la dépendance de g et de ϕ en fonction de $u_{]t_0,t]}$, cela signifie que si $u, v \in U$ et vérifient

$$d_u(u_{[t_0,t]}, v_{[t_0,t]}) = \sup_{\tau \in [t_0,t] \cap T} \{ d(u(\tau), v(\tau)) \} \text{ petit}$$

alors $d\{\phi[t; u_{]t_0,t]}, x(t_0)], \phi[t; v_{]t_0,t]}, x(t_0)]\}$

et $\sup_{\tau \in [t_0,t] \cap T} \{ \| \hat{g}[x(t_0); u_{]t_0,t]}] - \hat{g}[x(t_0); v_{]t_0,t]}] \| \}$

sont aussi petits, et où $\hat{g}[x(t_0); u_{]t_0,t]}] = g[x(t_0), u(t_0+), t_0]$ et $\hat{g}[x(t_0); v_{]t_0,t]}] = g[x(t_0), v(t_0+), t_0]$ par définition.

Axiome 4

La fonction ϕ satisfait les conditions suivantes :

1.) Pour tout $t, t_0 \in T$, $u \in U$ et $x(t_0) \in \Sigma$, $\phi[t_0, u_{]t_0,t]}, x(t_0)] = x(t_0)$
ou sens $\lim_{t \to t_0} \phi[t, u_{]t_0,t]}, x(t_0)] = x(t_0)$

2) Pour tout $t_0 < \hat{t} \leq t \in T$, $u \in U$ et $x(t_0) \in \Sigma$,

$$\phi[t; u]_{t_0,t}], x(t_0)] = \phi[t, u]_{\hat{t},t}], \phi[\hat{t}; u]_{t_0,\hat{t}}], x(t_0)]]$$

où $t_0 < \hat{t} < t$. Cette propriété est appelée la propriété de transition ou de semi-groupe.

3) Pour tout τ, t_0, t avec $\tau \in [t_0, t] \cap T$, et pour tout $x(t_0) \in \Sigma$, si $u, v \in U$ avec $u]_{t_0,t}] = v]_{t_0,t}]$, alors

$$\phi[\tau, u]_{t_0,t}], x(t_0)] = \phi[\tau, v]_{t_0,t}], x(t_0)]$$

On peut maintenant poser la définition :

Un système dynamique contrôlé, ou système dynamique S est la donnée de T, Σ, Ω et U, d'une variable $x(t)$ et d'une fonction g tels que les axiomes 1 à 4 soient vérifiés. Dans ce cas, T est appelé le domaine du système ; Σ l'espace d'état du système, U l'espace d'input du système et $x(t)$ la variable d'état ; $u]_{t_0,t}]$ est l'input pendant l'intervalle d'observation $]t_0, t]$ et $y]_{t_0,t}] = g[x(t_0); u]_{t_0,t}]$ est l'output du système. La fonction $\phi[t, u]_{t_0,t}], x(t_0)]$ est la fonction de transition du système. L'ensemble $\{x(\tau) \mid x(\tau) = \phi[\tau, u]_{t_0,t}], x(t_0)]$ pour $\tau \in [t_0, t] \cap T\}$ est la trajectoire du système pendant $[t_0, t] \cap T$, au départ de l'état initial $x(t_0)$ et générée par le contrôle (input) $u]_{t_0,t}]$. Enfin les équations

$$y(t) = g[x(t), u(t), t]$$
$$x(t) = \phi[t, u]_{t_0,t}], x(t_0)]$$

sont respectivement les équations de sortie et d'état du système.

Donnons maintenant une brève interprétation des axiomes.

Axiome 1. Cet axiome dit principalement que si on connait $x(t_0)$ et t_0, et si l'on applique un contrôle donné pendant $]t_0,t]$ avec $t>t_0$, l'output du système est déterminé de manière unique. Pour prévoir la sortie pendant $]t_0,t]$, il n'est pas nécessaire de connaitre l'input avant t_0. $x(t_0)$ et $u|]t_0,t]$ suffisent. On note aussi que les valeurs futures du contrôle n'affectent pas $y|]t_0,t]$, ce qui implique que le système est non anticipatif.

Axiome 2. L'axiome 2 assure l'existence d'"assez" d'états possibles du système pour pouvoir tout couple input, output. Il implique aussi que la connaissance de $x(t_0)$ et de $u|]t_0,t]$ suffit à déterminer, non seulement $y(\hat{t})$ pour $\hat{t}\in]t_0,t]$, mais aussi $x(\hat{t})$, $t_0<t_1\leq t$. L'état concentre, d'une certaine façon, toute l'information du passé requise pour prédire les états futurs de l'output et de $x(t)$.

Axiome 3 C'est un axiome de continuité qui concerne la "stabilité" du système pour des modifications de $x(t_0)$ ou $u|]t_0,t]$ en plus de la continuité de la trajectoire.

Axiome 4 Il faut que 1) $x(t_0)$ soit le point de départ de la trajectoire, 2) si un contrôle mène $x(t)$ de $x(t_0)$ à $x(\hat{t})$ en passant par $x(\tilde{t})$, ce même contrôle mène $x(t)$ de $x(t_0)$ en $x(\tilde{t})$, 3) le système soit non anticipatif.

Nous voyons donc qu'il s'agit bien de la traduction formelle des idées exprimées dans le chapitre précédent, qui correspondent, en gros, à une description "réaliste" de systèmes physiques.

Nous pouvons, après ces généralités, aborder maintenant la description des systèmes qui seront étudiés plus en détail dans le reste de ce travail.

IV Les systèmes dynamiques contrôlés: cas particuliers.

1§ La fonctionnelle de coût

Maintenant que nous disposons d'un système dont nous pouvons influencer l'output par un contrôle, revenons à notre expérience du début. Comment produire, à partir d'un contrôle bien choisi, un output possédant cette qualité supplémentaire d'avoir des caractéristiques que nous déterminons d'avance ?

En d'autres termes, si nous assignons un coût déterminé à une paire input, output, quel input choisir pour minimiser ce coût. Représentons le coût sous la forme suivante :

$$J[u,y] = f(y(t)) + \int_{t_0}^{t} g(y,u,t)\,dt$$

Le terme $f(y(t))$ s'introduit de manière naturelle. Il est en effet possible que le coût associé à la trajectoire de l'output soit influencé par la valeur de cet output à la fin du processus.

Désormais, notre problème sera de minimiser $J[u,y]$ en travaillant sur un système dynamique contrôlé $\mathcal{S}$ donné.

$$y(t) = g[x(t), u(t), t]$$

$$x(t) = \phi[t, u]_{t_0,t}, x(t_0)]$$

$$\min_{u \in U} J[u,y] = f[y(t)] + \int_{t_0}^{t} g_0[y(t), u(t), t]\,dt$$

2§ Système différentiel non linéaire.

Nous considérons ici le problème suivant:

$$[t_0, t] \cap T = [0, T] \subseteq \mathbb{R}$$

$$G = \{ W_2^\alpha[0,T], \mathbb{R}^n \} \qquad \Sigma = \mathbb{R}^n$$

$$U = \{ W_2^\alpha[0,T], \mathbb{R}^r \} \qquad \Omega = \mathbb{R}^r$$

où G représente l'ensemble de $x(t)$ et où $\{ W_2^\alpha[0,T], \mathbb{R}^k \}$ est l'espace de Sobolev de toutes les fonctions définies sur $[0,T]$ et à valeurs dans $\mathbb{R}^k$ telle qu'elles soient de carré intégrable sur $[0,T]$ ainsi que toutes leurs dérivées au sens des distributions, jusqu'à l'ordre α. Cet espace est muni de la norme

$$\|z\|_{2,\alpha}^2 = \sum_{i=1}^{k} \int_0^T \sum_{j=0}^{\alpha} |z_i^{(j)}(t)|^2 \, dt$$

Les équations de ce système sont

$$\begin{aligned} \dot{x}(t) &= f(x,u,t) \\ y(t) &= x(t) \\ \text{avec } x(0) &= x_0 \end{aligned}$$

et le coût associé $\quad J[u] = \int_0^T g(x,u,t) \, dt$

où $\quad f : G \times U \times [0,T] \longrightarrow \mathbb{R}^n$
$\quad g : G \times U \times [0,T] \longrightarrow \mathbb{R}$

Il est clair que notre système est maintenant bien défini.
Nous étudierons son comportement dans les pages qui suivent, lorsqu'il sera question de la méthode de Ritz-Galerkin.

3§ Le régulateur d'état différentiel linéaire

Choisissons cette fois

$$[t_0, t] \cap T = [0, T] \subset \mathbb{R} \qquad \Sigma = \mathbb{R}^n$$

$$G = \Phi^n = \{ \text{fonctions } [0,T] \to \mathbb{R}^n, \text{ continues par morceaux et à dérivée bornée} \}$$

$$U = \Phi^r \qquad \Omega = \mathbb{R}^r$$

Les équations seront

$$\dot{x}(t) = A(t)x(t) + B(t)u(t)$$
$$y(t) = x(t)$$
avec $x(0) = x_0$

où $A(t)$ et $B(t)$ sont deux matrices respectivement $n \times n$ et $n \times r$, continues par morceaux sur $[0, T]$.

Le coût est donné par la fonction suivante, où F est $n \times n$ définie positive,

$$J[u, x] = \frac{1}{2} \int_0^T \{ \langle x(t), Q(t)x(t) \rangle_n + \langle u(t), R(t)u(t) \rangle_r \} dt + \frac{1}{2} \langle x(T), Fx(T) \rangle_n$$

où $Q(t)$ est une matrice $n \times n$ symétrique, définie positive et $R(t)$ $r \times r$, symétrique et définie positive, toutes deux continues sur $[0, T]$.
La notation $\langle \cdot, \cdot \rangle_p$ représente le produit scalaire habituel de $\mathbb{R}^p$.
Le système sera aussi étudié dans les pages suivantes.

§4 Le régulateur d'état parabolique linéaire

Introduisons un dernier problème :

$$[t_0,t] \cap T = [0,T] \subset \mathbb{R} \qquad \Sigma = \mathbb{R}^n \qquad \Omega = \mathbb{R}^r$$

$$G = \left\{ V_2^1[(0,1)\times(0,T)], \mathbb{R}^n \right\}$$

$$U = \left\{ V_2^0[(0,1)\times(0,T]], \mathbb{R}^r \right\}$$

où $\left\{ V_p^q[(0,1)\times(0,T)], \mathbb{R}^k \right\}$ est l'ensemble des fonctions à valeurs dans $\mathbb{R}^k$ telles que

$$\frac{\partial^{i+j} f}{\partial x^i \partial t^j} \in L^{\frac{p}{p-1}}[(0,1)\times(0,T)]$$

pour $i = 1-q$ et $j = 1-q-1$, muni de la norme

$$\|f(x,t)\|_{V_q^p} = \int_0^T \int_0^1 \left[\sum_{j=0}^{q} \sum_{k=0}^{q-1} \left| \frac{\partial^{j+k}}{\partial x^j \partial t^k} f(x,t) \right|^2 \right]^{p/2} dx\,dt$$

Les équations, condition initiale et conditions frontières sont

$$\frac{\partial v}{\partial t} = A(x,t) \frac{\partial^2 v}{\partial x^2} + B(x,t)\, u(x,t)$$

$$\lim_{t \to 0} v(x,t) = v_0(x)$$

$$\alpha\, v(0,t) + \frac{\partial v(0,t)}{\partial x} = c\, f_1(t)$$

$$\beta v(1,t) + \frac{\partial v(1,t)}{\partial x} = d\, f_2(t)$$

$$y(x,t) = v(x,t)$$

où f_1 et $f_2 \in \{W_2^0[0,T], \mathbb{R}^n\}$, A et B sont deux matrices respectivement $n \times n$ et $n \times r$, toutes deux par morceaux $C^\gamma(0,1) \times C^{\gamma-1}(0,T)$ pour $\gamma \geq 1$, et où $v_0(x) \in C^\gamma(0,1)$

La fonctionnelle de coût s'écrit

$$J[u, f_1, f_2] = \frac{1}{2}\int_0^T \int_0^1 \{ \langle v, Q(x,t)v \rangle + \langle u, R(x,t)u \rangle \} \, dx \, dt$$
$$+ \frac{1}{2}\int_0^T \{ \langle f_1, r(t) f_1 \rangle + \langle f_2, s(t) f_2 \rangle \} \, dt$$

où Q, R, s et r sont des matrices $n \times n$ définies positives et par morceaux $C^\gamma(0,1) \times C^{\gamma-1}(0,T)$ pour $\gamma \geq 1$ et $\langle \cdot, \cdot \rangle$ est le produit scalaire.

Ce système un peu plus complexe, où l'espace intervient en plus du temps, sera, lui-aussi, étudié dans ce travail.

§5 Remarques sur les régulateurs d'état.

D'abord, les deux systèmes introduits sous ce nom, doivent leur appellation au fait que le but de l'optimisation est de ramener l'état du système au repos (au zéro). On désire de plus que le contrôle ne prenne pas des valeurs trop élevées (dans plusieurs cas pratiques, l'énergie dépensée pour le contrôle est proportionnelle à $\int_0^T u(t)^2 dt$). Ces considérations expliquent la forme de la fonctionnelle de coût où les valeurs x et u apparaissent dans des formes quadratiques toujours positives.

Il est à remarquer qu'aucune contrainte explicite n'est là pour limiter la taille du contrôle. Seule la pénalisation dans le coût limite cette taille.

V Les méthodes de résolution

Dans ce paragraphe, nous aborderons d'abord une méthode générale de résolution du problème du régulateur d'état différentiel linéaire. Ensuite nous parlerons de la méthode de Riz-Galerkin, pour introduire le chapitre qui l'étudiera plus en détail.

1§ Le problème du régulateur d'état.

Rappelons les données générales du problème

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ x(0) = x_0 \end{cases}$$

$$J[u] = \frac{1}{2}\langle x(T), Fx(T)\rangle + \frac{1}{2}\int_0^T \{\langle x(t), Q(t)x(t)\rangle + \langle u(t), R(t)u(t)\rangle\}\, dt$$

comme plus haut.

Nous allons montrer que $u(t) = G(t)x(t)$, où $G(t)$ est une matrice $r \times n$.

Supposons d'abord que le contrôle optimal existe pour tout état initial. Nous allons nous servir du principe de Pontryagine pour obtenir des conditions nécessaires sur le contrôle optimal et tenter de le calculer à partir de là.

Ecrivons l'Hamiltonien du système

$$H = \frac{1}{2}\langle x(t), Qx(t)\rangle + \frac{1}{2}\langle u(t), Ru(t)\rangle + \langle Ax(t), \lambda(t)\rangle + \langle B(t)u(t), \lambda(t)\rangle$$

Le multiplicateur $\lambda(t)$ est solution de $\dot{\lambda}(t) = -\dfrac{\partial H}{\partial x(t)}$

soit $\dot{\lambda}(t) = -Q(t)x(t) - A^*(t)\lambda(t)$

Le long de la trajectoire optimale, on doit avoir $\frac{\partial H}{\partial u(t)} = 0$
ce qui équivaut à

$$\frac{\partial H}{\partial u(t)} = R(t)u(t) + B^t(t)\lambda(t) = 0$$

d'où on déduit $u(t) = -R^{-1}(t)B^t(t)\lambda(t)$.
Le fait que $R(t)$ est définie positive pour tout $t \in [0, T]$ implique l'existence de $R^{-1}(t)$ pour toutes ces valeurs.

Nous savons que l'Hamiltonien doit être minimum. La condition $\frac{\partial H}{\partial u(t)} = 0$ assure uniquement un extremum. Pour un minimum, il faut que $\frac{\partial^2 H}{\partial u^2(t)}$ soit une matrice définie positive.
Or

$$\frac{\partial^2 H}{\partial u^2(t)} = R(t)$$

qui est définie positive. Ce qui prouve donc que le contrôle $u(t)$ trouvé plus haut minimise bien H.

Remplaçons maintenant $u(t)$ par $-R^{-1}(t)B^t(t)\lambda(t)$

$$\dot{x}(t) = A(t)x(t) - B(t)R^{-1}(t)B^t(t)\lambda(t)$$

Si nous définissons $S(t) = B(t)R^{-1}(t)B^t(t)$, qui est évidemment une matrice $n \times n$ symétrique, nous pouvons écrire

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A(t) & -S(t) \\ -Q(t) & -A^t(t) \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}$$

qui est un système de $2n$ équations différentielles homogène linéaire. Nous savons obtenir une solution unique de ce système pourvu que nous ayons $2n$ conditions frontières. La condition initiale $x(0) = x_0$ nous en fournit n. D'autre part, la valeur $\lambda(T)$ n'étant pas encore déterminée

nous pouvons la choisir de manière à minimiser $\frac{1}{2}\langle x(T), Fx(T)\rangle$, soit

$$\lambda(T) = \frac{\partial}{\partial x(T)} \frac{1}{2}\langle x(T), Fx(T)\rangle = Fx(T)$$

ce qui nous fournit les n conditions nécessaires. (C'est la condition de transversalité, cfr (A-1) pp270-271)

Soit $\Omega(t,0)$ la matrice fondamentale (2n×2n) du système différentiel, on obtient donc

$$\begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \Omega(t,0) \begin{bmatrix} x(0) \\ \lambda(0) \end{bmatrix}$$

où $\lambda(0)$ est inconnu. En $t=T$, on obtient

$$\begin{bmatrix} x(T) \\ \lambda(T) \end{bmatrix} = \Omega(T,t) \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}$$

Partageons maintenant $\Omega(T,t)$ en 4 sous-matrices n×n

$$\Omega(T,t) = \begin{bmatrix} \Omega_{11}(T,t) & \Omega_{12}(T,t) \\ \Omega_{21}(T,t) & \Omega_{22}(T,t) \end{bmatrix}$$

Réécrivons l'équation précédente

$$x(T) = \Omega_{11}(T,t)\,x(t) + \Omega_{12}(T,t)\,\lambda(t)$$
$$\lambda(T) = \Omega_{21}(T,t)\,x(t) + \Omega_{22}(T,t)\,\lambda(t) = Fx(T)$$

Ce qui fournit, après un bref calcul

$$\lambda(t) = \left[\Omega_{22}(T,t) - F\Omega_{12}(T,t)\right]^{-1}\left[F\Omega_{11}(T,t) - \Omega_{21}(T,t)\right]x(t)$$

pourvu que l'inverse considéré existe.

Posons maintenant

$$K(t) = \left[\Omega_{22}(T,t) - F\Omega_{12}(T,t)\right]^{-1}\left[F\Omega_{11}(T,t) - \Omega_{21}(T,t)\right]$$

une matrice n×n dépendant de T et F, mais pas de l'état initial. Montrons maintenant que $\left[\Omega_{22}(T,t) - F\Omega_{12}(T,t)\right]^{-1}$ existe. En $t=T$ nous avons $\Omega(T,T) = I$, ce qui implique

$$\Omega_{11}(T,T) = \Omega_{22}(T,T) = I$$
$$\Omega_{12}(T,T) = \Omega_{21}(T,T) = 0$$

Donc $\Omega_{22}(T,T) - F\Omega_{12}(T,T) = I$ est inversible. De plus, on voit immédiatement que

$$K(T) = [\Omega_{22}(T,T) - F\Omega_{12}(T,T)]^{-1}[F\Omega_{11}(T,T) - \Omega_{21}(T,T)] = F$$

Comme nous venons de le voir, tout marche bien en $t = T$. On peut voir (Réf (k-1)) que l'inverse existe pour tout $t \in [t_0, T]$; nous pouvons alors écrire

$$\lambda(t) = K(t)\, x(t)$$

Quelques commentaires s'imposent ici. Quand les matrices A, S et Q dépendent du temps, il est généralement impossible d'obtenir une expression analytique de $K(t)$. Même lorsque ces matrices sont constantes, l'évaluation de $K(t)$ est un calcul très laborieux, spécialement lorsque l'ordre du système est élevé.

Nous allons maintenant mettre en évidence une propriété de $K(t)$ qui permettra de la calculer, sans devoir inverser une matrice $n \times n$.

Supposons que $x(t)$ et $\lambda(t)$ sont solutions du système différentiel avec $\lambda(t) = K(t)\, x(t)$. On obtient

$$\dot{\lambda}(t) = \dot{K}(t) x(t) + K(t)\, \dot{x}(t)$$

Or nous avons $\dot{x}(t) = A(t)x(t) - S(t)\lambda(t)$

$$\dot{\lambda}(t) = -Q(t)x(t) - A^t(t)\lambda(t)$$

En substituant, on obtient

$$\dot{x}(t) = [A(t) - S(t)K(t)]\, x(t)$$
$$\dot{\lambda}(t) = [\dot{K}(t) + K(t)A(t) - K(t)S(t)K(t)]\, x(t)$$

et encore

$$\dot{\lambda}(t) = [-Q(t) - A^t(t)K(t)]\, x(t)$$

ce qui fournit

$$[\dot{K}(t) + K(t)A(t) - K(t)S(t)K(t) + A^t(t)K(t) + Q(t)]\, x(t) = 0$$

pour tout $t \in [t_0, T]$

Comme cette équation doit être vraie pour tout choix des conditions initiales, car $K(t)$ ne dépend pas de celles-ci, on doit avoir

$$\dot{K}(t) + K(t)A(t) + A^t(t)K(t) - K(t)S(t)K(t) + Q(t) = 0$$

or $S(t) = B(t)R^{-1}(t)B^t(t)$, donc

$$\dot{K}(t) = -K(t)A(t) - A^t(t)K(t) + K(t)B(t)R^{-1}(t)B^t(t)K(t) - Q(t)$$

Nous avons donc établi

> Si $x(t)$ et $\lambda(t)$ sont les solutions du système différentiel et si $\lambda(t) = K(t)x(t)$ pour tout $t \in [t_0, T]$ et tout $x(t)$, alors $K(t)$ doit satisfaire
> $$\dot{K}(t) = -K(t)A(t) - A^t(t)K(t) + K(t)B(t)R^{-1}(t)B^t(t)K(t) - Q(t)$$

De plus, si $t = T$, on doit avoir (cfr supra) $\lambda(T) = Fx(T)$, mais aussi $\lambda(T) = K(T)x(T)$. Ce qui fournit

$$[K(T) - F]\,x(T) = 0$$

pour tout $x(T)$. Donc $K(T) = F$
(analogue à ce qui a déjà été vu).
Nous avons donc une équation de Riccati pour la matrice $K(t)$ avec une condition limite : nous avons donc une seule solution.
Pour réduire l'ordre de l'équation, voyons de plus que

> $K(t)$ est symétrique, pour $\forall t \in [t_0, T]$

Démonstration : Transposons les deux membres de l'équation de Riccati :

$$\left[\frac{d}{dt}K(t)\right]^t = -K^t(t)A(t) - A^t(t)K^t(t) + K^t(t)B(t)R^{-1}(t)B^t(t)K^t(t) - Q(t)$$

puisque $Q(t)$ et $B(t)R^{-1}(t)B^t(t)$ sont symétriques.
D'autre part, pour toute matrice $K(t)$, il est vrai que

$$\left[\frac{d}{dt}K(t)\right]^t = \frac{d}{dt}\left[K^t(t)\right]$$

En comparant alors l'équation originale et l'équation transposée, on voit que $K(t)$ et $K^t(t)$ satisfont la même équation. Comme F est symétrique, on conclut $K(T) = K^t(T) = F$.
Comme la solution de l'équation est unique, on a bien $K(t) = K^t(t)$ ■

Théorème II.1

Soit le système $\dot{x}(t) = A(t)x(t) + B(t)u(t)$
et la fonctionnelle de coût

$$J_1 = \frac{1}{2}\langle x(T), Fx(T)\rangle + \frac{1}{2}\int_0^T \{\langle x(t), Q(t)x(t)\rangle + \langle u(t), R(t)u(t)\rangle\}dt$$

où $u(t)$ n'est soumis à aucune contrainte, T est donné, F et $Q(t)$ et $R(t)$ sont définies positives et symétriques. Alors, l'unique contrôle optimal est donné par

$$u(t) = -R^{-1}(t)B^t(t)K(t)x(t)$$

où la matrice $K(t)$ $(n \times n)$ est symétrique et seule solution de l'équation

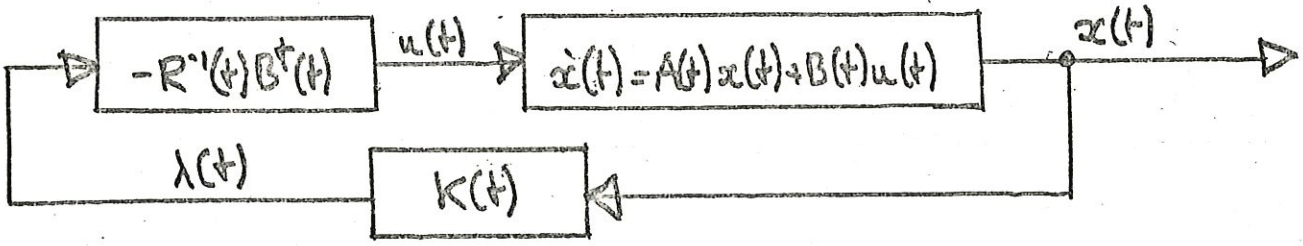
$$\dot{K}(t) = -K(t)A(t) - A^t(t)K(t) + K(t)B(t)R^{-1}(t)B^t(t)K(t) - Q(t)$$

satisfaisant la condition limite $K(T) = F$
L'état $x(t)$ est donné par l'équation

$$\dot{x}(t) = [A(t) - B(t)R^{-1}(t)B^t(t)K(t)]x(t) \qquad x(0) = x_0$$

Nous avons déjà établi la condition nécessaire. La condition suffisante, qui nécessite de longs développements, ne sera pas traitée ici. Une démonstration extensive peut être trouvée dans (A-1), pp 763 sq.

Pour terminer ce paragraphe, nous donnons ci dessous un schéma montrant la structure du système étudié.



§2 L'approximation de Ritz Galerkin

Dans des cas plus généraux que celui traité dans le §1, les calculs deviennent évidemment nettement plus complexes. Même dans le cas précédent, la masse de calcul peut devenir extrêmement lourde.

La difficulté réside principalement dans le fait que tous les espaces fonctionnels considérés (principalement U et G) sont de dimension infinie : on ne peut donc exprimer les fonctions considérées comme des combinaisons linéaires finies de fonctions qu'on se donne à priori.

Pour remédier à cet inconvénient, nous introduirons l'approximation suivante. Choisissons dans U et dans G deux sous-espaces de dimension finie, soit U_m et G_m respectivement. Résolvons le problème sur ces espaces uniquement. Nous espérons obtenir ainsi une solution approchée, d'autant meilleure que U_m et G_m seront plus proches de U et G. Comme nous construisons U_m et G_m de manière à pouvoir

y choisi des base finies, nous pouvons écrire

$$\begin{cases} x(t) = \sum_{j=1}^{m} \alpha_j \, G_j^m(t) \\ \lambda(t) = \sum_{j=1}^{m} \beta_j \, G_j^m(t) \\ u(t) = \sum_{j=1}^{m} \gamma_j \, U_j^m(t) \end{cases}$$

et, dans ce cas, nous écrirons la contrainte de l'équation d'état, lorsque cette écriture a un sens,

$$\int_{t_0}^{t} < G_j^m(\hat{t}), \, x(\hat{t}) - \phi[\hat{t}, u]_{t_0,t}], x(t_0)] > d\hat{t} = 0$$

avec $x(t_0) = x_0$

L'idée qui consiste à choisir des sous-espaces de dimension finie est-elle bonne? Les solutions trouvées par cette méthode convergent-elles, dans un sens encore à définir, vers les solutions du problème non approximé?

Voilà quelques questions auxquelles nous tenterons de répondre dans les pages qui suivent, en considérant successivement les trois problèmes décrits dans le paragraphe IV.

Chapitre III

La méthode de Ritz Galerkin sur des modèles de contrôle optimal

I Introduction

Nous allons maintenant considérer successivement les différents problèmes abordés dans le paragraphe IV du Chapitre II. Nous tenterons de trouver, pour l'approximation de Ritz Galerkin, des bornes d'erreur à priori, qui permettront de prouver la convergence des solutions approchées vers "la" solution du problème non approximé.

II Le régulateur d'état différentiel linéaire

Il s'agit du système présenté au Chapitre II, IV, §3. Rappelons brièvement les données du problème.

Soit $Q(t)$ une matrice $n \times n$, symétrique, définie positive et $R(t)$ une matrice $r \times r$ symétrique, définie positive, toutes deux continues sur l'intervalle de temps $[0, T]$ où T est fixé. Pour chaque $k \geq 0$, définissons

$$\Phi^k = \{ f : [0,T] \longrightarrow \mathbb{R}^k \mid f \text{ est différentiable par morceaux et à dérivée bornée} \}$$

Le problème est de trouver un contrôle optimal $u^* \in \Phi^r$ et un état $x^* \in \Phi^n$ qui minimisent

$$J[u,x] \equiv \frac{1}{2} \int_0^T \{ \langle x(t), Q(t)x(t) \rangle_n + \langle u(t), R(t)u(t) \rangle_n \} \, dt \qquad \text{(III.1)}$$

parmi tous les $u \in \Phi^r$, où $x(t)$ est donné par

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) & t>0 \\ x(0) = x_0 \end{cases} \qquad \text{(III.2)}$$

avec $|y|_n^2 = \langle y, y\rangle_n = \sum_{i=1}^{n} y_i^2$ pour tout $y \in \mathbb{R}^n$ et, de même $|z|_r^2 = \langle z, z\rangle_r = \sum_{i=1}^{r} z_i^2$ pour tout $z \in \mathbb{R}^r$. De plus, $A(t)$ est une matrice $n \times n$ et $B(t)$ une matrice $n \times r$, toutes deux continues par morceaux sur $[0,T]$

1§ Procédure variationnelle

Comme vu précédemment, le problème ci-dessus est équivalent à celui de trouver $\lambda^* \in \Phi^n$ qui extrémise le Lagrangien

$$L[u,x;\lambda,\gamma] = J[u,x] + \int_0^T \langle \lambda(t), -\dot{x}(t) + A(t)x(t) + B(t)u(t)\rangle_n \, dt$$
$$+ \langle \gamma, x(0) - x_0\rangle_n \qquad \text{(III.3)}$$

sous la condition de transversalité $\lambda(T) = 0$ (III.4), où $\gamma, u(t)$ et $x(t)$ sont donnés respectivement par

$$\gamma = -\lambda(0) \qquad \text{(III.5)}$$
$$u(t) = -R^{-1}(t)B^t(t)\lambda(t) \quad t \in [0,T] \qquad \text{(III.6)}$$
$$x(t) = -Q^{-1}(t)\left[\dot{\lambda} + A^t(t)\lambda(t)\right] \quad t \in [0,T] \qquad \text{(III.7)}$$

Effectuons à présent quelques changements d'écriture.

De (III.3) on tire :

$$L = \frac{1}{2}\int_0^T \langle x, Qx\rangle + \frac{1}{2}\int_0^T \langle u, Ru\rangle + \int_0^T \langle \lambda, -\dot{x}\rangle + \int_0^T \langle \lambda, Ax\rangle + \int_0^T \langle \lambda, Bu\rangle$$
$$+ \langle -\lambda(0), x(0) - x_0\rangle$$

$$L = \frac{1}{2}\int_0^T \langle x, Qx\rangle + \frac{1}{2}\int_0^T \langle u, Ru\rangle - \Big[\langle \lambda, x\rangle\Big]_0^T + \int_0^T \langle \dot\lambda, x\rangle + \int_0^T \langle \lambda, Ax\rangle + \int_0^T \langle \lambda, Bu\rangle$$

$$+ \langle -\lambda(0), x(0) - x_0\rangle$$

$$= \frac{1}{2}\int_0^T \langle x, Qx\rangle + \frac{1}{2}\int_0^T \langle u, Ru\rangle - \langle \lambda(T), x(T)\rangle + \langle \lambda(0), x(0)\rangle$$

$$+ \int_0^T \langle \dot\lambda + A^t\lambda, x\rangle + \int_0^T \langle B^t\lambda, u\rangle - \langle \lambda(0), x(0)\rangle + \langle \lambda(0), x_0\rangle$$

$$= \frac{1}{2}\int_0^T \langle x, Qx\rangle + \frac{1}{2}\int_0^T \langle u, Ru\rangle - \int_0^T \langle u, Qx\rangle - \int_0^T \langle u, Ru\rangle + \langle \lambda(0), x_0\rangle$$

En effet, $\lambda(T) = 0$ et (III.6) implique $-Ru = B^T\lambda$, tandis que de (III.7) on tire $-Qx = \dot\lambda + A^t\lambda$

Donc

$$L = -J[u, x] + \langle \lambda(0), x_0\rangle$$

Les conditions de continuité sur A, B, R, Q assurent l'existence des intégrales considérées.

Mais

$$\frac{1}{2}\int_0^T \langle u, Ru\rangle_1 dt = \frac{1}{2}\int_0^T \langle R^{-1}B^t\lambda, RR^{-1}B^t\lambda\rangle_1 dt = \frac{1}{2}\int_0^T \langle BR^{-1}B^t\lambda, \lambda\rangle_n dt$$

en utilisant (III.6).

De même, par (III.7), on obtient

$$\frac{1}{2}\int_0^T \langle x, Qx\rangle_1 dt = \frac{1}{2}\int_0^T \langle Q^{-1}A^t\lambda + Q^{-1}\dot\lambda,\ A^t\lambda + \dot\lambda\rangle_n dt$$

$$= \frac{1}{2}\int_0^T \langle Q^{-1}\dot\lambda, \dot\lambda\rangle dt + \frac{1}{2}\int_0^T \langle Q^{-1}A^t\lambda, A^t\lambda\rangle dt$$

$$+ \frac{1}{2}\int_0^T \langle Q^{-1}A^t\lambda, \dot\lambda\rangle dt + \frac{1}{2}\int_0^T \langle Q^{-1}\dot\lambda, A^t\lambda\rangle dt$$

$$= \frac{1}{2}\int_0^T \langle Q^{-1}\dot\lambda, \dot\lambda\rangle dt + \frac{1}{2}\int_0^T \langle AQ^{-1}A^t\lambda, \lambda\rangle dt$$

$$+ \int_0^T \langle AQ^{-1}\dot\lambda, \lambda\rangle dt$$

car φ est symétrique

Donc, si nous définissons, pour tout λ et η dans $\underline{\mathcal{F}}^n$

$$[\lambda,\eta] = \int_0^T \langle \varphi^{-1}\dot{\lambda},\dot{\eta}\rangle_n dt + \int_0^T \langle A\varphi^{-1}A^t\lambda,\eta\rangle_n dt$$
$$+ \int_0^T \langle BR^{-1}B^t\lambda,\eta\rangle_n dt + \int_0^T \{\langle A\varphi^{-1}\dot{\lambda},\eta\rangle_n + \langle A\varphi^{-1}\dot{\eta},\lambda\rangle_n\} dt \qquad (III.8)$$

nous pouvons bien écrire

$$F[\lambda] \equiv -L[u,x;\lambda,\gamma] = \frac{1}{2}\int_0^T \langle \varphi^{-1}\dot{\lambda},\lambda\rangle dt + \frac{1}{2}\int_0^T \langle A\varphi^{-1}\lambda,\lambda\rangle dt$$
$$+ \frac{1}{2}\int_0^T \langle BR^{-1}B^t\lambda,\lambda\rangle dt + \int_0^T \langle A\varphi^{-1}\dot{\lambda},\lambda\rangle dt - \langle\lambda(0),x_0\rangle_n$$

$$F[\lambda] = +\frac{1}{2}[\lambda,\lambda] - \langle\lambda(0),x_0\rangle_n \qquad (III.9)$$

qu'il faut minimiser par rapport à λ.

2§ Quelques théorèmes

Définissons les notations suivantes

$$|M|_t = \max\{|Mx|_t \mid x\in\mathbb{R}^t \text{ et } |x|_t = 1\}$$

$$\lambda_\varphi = \min_{t\in[0,T]}\{\lambda(t) \mid \lambda(t) \text{ est une valeur propre de } \varphi(t)\}$$

$$\|\varphi\|_2^2 = \int_0^T |\varphi(t)|_n^2\, dt$$

$$\|A^t\|_2^2 = \int_0^T |A^t(t)|_n^2\, dt$$

$$\|A^t\|_\infty = \sup_{t\in[0,T]} |A^t(t)|_n$$

et enfin

$$\varrho = \|g\|_2 \left[2\|A^t\|_\infty\right]^{-1/2} \exp\left[\|A^t\|_\infty T\right]$$

On peut alors démontrer le théorème suivant

Théorème III.1

Le multiplicateur de Lagrange optimal λ existe et est l'unique solution dans $\underline{\Phi}_0^n = \{\phi \in \underline{\Phi}^n \mid \phi(T)=0\}$ du problème généralisé ($[\lambda,\eta]$ est donné par (11))

$$[\lambda,\eta] = \langle \eta(0), x_0\rangle_n \quad \text{pour tout } \eta \in \underline{\Phi}_0^n \tag{III.10}$$

De plus, $[\lambda,\eta]$ est symétrique et

$$\|\mathring{\lambda}\|_2^2 = \int_0^T \langle \mathring{\lambda},\mathring{\lambda}\rangle_n dt \leq \lambda_g^{-1}\left[\|g\|_2 + \varrho\|A^t\|_2\right]^2 [\lambda,\lambda] \tag{III.11}$$

$$\|\lambda\|_2^2 \leq \lambda_g^{-1}\varrho^2[\lambda,\lambda] \tag{III.12}$$

Démonstration : La partie "existence" du théorème nous est déjà connue. Il s'agit d'ailleurs d'un résultat classique du calcul des variations (cfr. Ref A-1 et B-1).

Si $\eta \in \underline{\Phi}_0^n$ et $\alpha \in \mathbb{R}$, alors $F[\lambda^* + \alpha\eta] \geq F[\lambda^*]$ avec l'égalité ssi $\alpha = 0$, où λ^* est le multiplicateur de Lagrange optimal. Nous devons donc avoir

$$\frac{\partial F}{\partial \alpha}[\lambda^* + \alpha\eta](0) = 0$$

$$F[\lambda^*+\alpha\eta] = \frac{1}{2}\Big\{\int_0^T \langle g^{-1}(\mathring{\lambda}^*+\alpha\mathring{\eta}), \mathring{\lambda}+\alpha\mathring{\eta}\rangle + \int_0^T \langle AG^{-1}A^t(\lambda^*+\alpha\eta), \lambda^*+\alpha\eta\rangle$$

$$+ \int_0^T \langle BR^{-1}B^t(\lambda^*+\alpha\eta), \lambda^*+\alpha\eta\rangle\Big\} + \int_0^T \langle AG^{-1}(\mathring{\lambda}^*+\alpha\mathring{\eta}), \lambda^*+\alpha\eta\rangle$$

$$- \langle \lambda(0)+\alpha\eta(0), x_0\rangle$$

Dérivons terme à terme sous le signe "intégrale"

a) Soit $T_1 \equiv \langle g^{-1}\dot{\lambda}^*, \dot{\lambda}^* \rangle + \alpha \langle g^{-1}\dot{\lambda}^*, \dot{\eta} \rangle + \alpha \langle g^{-1}\dot{\eta}, \dot{\lambda}^* \rangle + \alpha^2 \langle g^{-1}\dot{\eta}, \dot{\eta} \rangle$
or g est symétrique, ce qui implique $\langle g^{-1}\dot{\lambda}^*, \dot{\eta} \rangle = \langle g^{-1}\dot{\eta}, \dot{\lambda}^* \rangle$
On a donc

$$\frac{\partial T_1}{\partial \alpha} = 2\langle g^{-1}\dot{\lambda}^*, \dot{\eta} \rangle + 2\alpha \langle g^{-1}\dot{\eta}, \dot{\lambda}^* \rangle$$

$$\left.\frac{\partial T_1}{\partial \alpha}\right|_{\alpha=0} = 2\langle g^{-1}\dot{\lambda}^*, \dot{\eta} \rangle$$

b) Soit $T_2 \equiv \langle Ag^{-1}A^t\dot{\lambda}^*, \lambda \rangle + \alpha \langle Ag^{-1}A^t\dot{\lambda}^*, \eta \rangle + \alpha \langle Ag^{-1}A^t\dot{\eta}, \lambda^* \rangle + \alpha^2 \langle Ag^{-1}A^t\eta, \eta \rangle$
Comme supra, on a $\langle Ag^{-1}A^t\dot{\lambda}^*, \eta \rangle = \langle Ag^{-1}A^t\eta, \dot{\lambda}^* \rangle$
Ce qui donne

$$\frac{\partial T_2}{\partial \alpha} = 2\langle Ag^{-1}A^t\dot{\lambda}^*, \eta \rangle + 2\alpha \langle Ag^{-1}A^t\eta, \eta \rangle$$

$$\left.\frac{\partial T_2}{\partial \alpha}\right|_{\alpha=0} = 2\langle Ag^{-1}A^t\dot{\lambda}^*, \eta \rangle$$

c) Soit $T_3 \equiv \langle BR^{-1}B^t\dot{\lambda}^*, \dot{\lambda}^* \rangle + \alpha \langle BR^{-1}B^t\dot{\lambda}^*, \eta \rangle + \alpha \langle BR^{-1}B^t\dot{\eta}, \lambda^* \rangle + \alpha^2 \langle BR^{-1}B^t\eta, \eta \rangle$
La symétrie de R implique $\langle BR^{-1}B^t\dot{\lambda}^*, \eta \rangle = \langle BR^{-1}B^t\eta, \dot{\lambda}^* \rangle$
Nous obtenons donc

$$\left.\frac{\partial T_3}{\partial \alpha}\right|_{\alpha=0} = 2\langle BR^{-1}B^t\dot{\lambda}^*, \eta \rangle$$

d) Soit $T_4 \equiv \langle Ag^{-1}\dot{\lambda}^*, \lambda \rangle + \alpha \langle Ag^{-1}\dot{\lambda}^*, \eta \rangle + \alpha \langle Ag^{-1}\dot{\eta}, \lambda^* \rangle + \alpha^2 \langle Ag^{-1}\dot{\eta}, \eta \rangle$
On obtient

$$\frac{\partial T_4}{\partial \alpha} = \langle Ag^{-1}\dot{\lambda}^*, \eta \rangle + \langle Ag^{-1}\dot{\eta}, \lambda^* \rangle + 2\alpha \langle Ag^{-1}\dot{\eta}, \eta \rangle$$

$$\left.\frac{\partial T_4}{\partial \alpha}\right|_{\alpha=0} = \langle Ag^{-1}\dot{\lambda}^*, \eta \rangle + \langle Ag^{-1}\dot{\eta}, \lambda^* \rangle$$

e) Soit enfin $T_5 = \langle \lambda(0), x_0 \rangle + \alpha \langle \eta(0), x_0 \rangle$

Donc $$\left.\frac{\partial T_5}{\partial \alpha}\right|_{\alpha=0} = \langle \eta(0), x_0 \rangle$$

Rassemblant tous les termes on obtient

$$\frac{\partial F}{\partial \alpha}[\lambda^* + \alpha\eta](0) = \int_0^T \langle Q^{-1}\dot{\lambda}^*, \dot{\eta}\rangle + \int_0^T \langle AQ^{-1}A^t\lambda^*, \eta\rangle + \int_0^T \langle BR^{-1}B^t\lambda^*, \eta\rangle$$
$$+ \int_0^T \{\langle AQ^{-1}\dot{\lambda}^*, \eta\rangle + \langle AQ^{-1}\dot{\eta}, \lambda^*\rangle\} - \langle \eta(0), x_0\rangle = 0$$

Ce qui fournit, en utilisant (III.8)

$$[\lambda, \eta] - \langle \eta(0), x_0\rangle = 0 \quad \text{pour tout } \eta \in \underline{\Phi}_0^n$$

ce qui est équivalent à (III.10)

D'autre part, on voit aisément dans (III.8) que $[\lambda, \eta]$ est symétrique en λ et η car Q et R sont symétriques.

De plus $\frac{1}{2}[\lambda, \lambda] = J[u, x] = \frac{1}{2}\int_0^T \langle x, Qx\rangle dt + \frac{1}{2}\int_0^T \langle u, Ru\rangle dt$

$$\geq \frac{1}{2}\int_0^T \langle x, Qx\rangle dt \geq \frac{1}{2}\lambda_Q \int_0^T \langle x, x\rangle dt = \frac{1}{2}\lambda_Q \|x\|_2^2 \qquad \text{(III.13)}$$

De (III.7) et (III.4) on peut déduire

$$\lambda(t) = -\int_t^T \frac{d\lambda}{ds}\, ds = \int_t^T [Qx + A^t\lambda]\, ds$$

Donc $$|\lambda(t)| \equiv \langle \lambda(t), \lambda(t)\rangle^{1/2} = \left| \int_t^T [Qx + A^t\lambda]\, ds \right|$$

car $|\cdot|$ est la norme associée à $\langle \cdot, \cdot \rangle$.

Ce qui donne

$$|\lambda(t)| \leq \left|\int_t^T Qx\right| + \left|\int_t^T A^t\lambda\right| \leq \int_t^T |Qx| + \left|\int_t^T A^t\lambda\right|$$
$$\leq \|Q\|_2 \|x\|_2 + \int_t^T |A^t(s)| \cdot |\lambda(s)|\, ds$$

en utilisant l'inégalité de Cauchy-Schwarz.

L'inégalité de Gronwall (Ref B-5) implique alors que

$$\|\lambda\|_2 \leq \|Q\|_2 [2\|A^t\|_\infty]^{-1/2} \exp\left[\|A^t\|_\infty T\right] \|x\|_2 = \xi \|x\|_2$$

Donc $[\lambda, \lambda] \geq \lambda_Q \xi^{-2} \|\lambda\|_2^2$, ce qui prouve (III.12)

De plus

$$\|\lambda\|_2 \leq \|g\|_2 \|x\|_2 + \|A^t\|_2 \|\lambda\|_2 \leq \left[\|g\|_2 + \rho \|A^t\|_2\right] \|x\|_2$$

qui, avec (III.8), prouve (III.11).
Pour l'unicité, si λ et μ vérifient (III.10), on a

$$0 = \left[\lambda-\mu, \lambda-\mu\right] \geq \lambda_0 \rho^{-2} \|\lambda-\mu\|_2^2$$

ce qui implique que $\lambda = \mu$ ∎

Comme nous allons travailler dans le cadre de l'approximation de Ritz-Galerkin, nous allons nous restreindre à S sous espace de dimension finie de Φ_0^b pour chercher les solutions. Nous déterminons ainsi une approximation λ_S de λ^* en minimisant F sur S et nous déterminons une approximation u_S de u par (III,6). En appliquant le contrôle obtenu dans l'équation (III.2) on obtient l'état x_S qui est généralement différent de celui qu'on obtiendrait par (III.7)

Montrons d'abord que cette approximation fournit une solution unique

Théorème III 2

> Il existe un seul $\lambda_S \in S$ qui minimise $F[\lambda]$ sur S

Démonstration : Choisissons $\{B_i(t)\}_{i=1,M}$ une base de S (de dimension M)

$$F\left[\sum_{i=1}^{M} \beta_i B_i\right] = \frac{1}{2}\left[\sum_{i=1}^{M} \beta_i B_i, \sum_{i=1}^{M} \beta_i B_i\right] - \left\langle \sum_{i=1}^{M} \beta_i B_i(0), x_0 \right\rangle_n$$

est une fonction de $\beta \in \mathbb{R}^M$, continûment différentiable, et donc F admet un minimum en β^* ssi

$$\begin{cases} \dfrac{\partial F}{\partial \beta_i}[\beta^*] = 0 & 1 \leq i \leq M \\ H = \left[\dfrac{\partial^2 F}{\partial x_i \partial x_j}\right] & \text{est définie positive} \end{cases} \qquad \text{(III.14)}$$

En calculant (III.14) comme précédemment, on obtient

$$\frac{\partial F}{\partial \beta_i}[\beta^*] = \sum_{j=1}^{M} \beta_j [B_i, B_j] - \langle B_i(0), x_0 \rangle \quad 1 \le i \le M \qquad \text{(III.15)}$$

c'est à dire $A\beta^* = k$ avec $A = \{[B_i, B_j]\}$ et $k_i = \langle B_i(0), x_0 \rangle$.
A est évidemment symétrique, car $[\lambda, \eta]$ l'est.
De plus, si $\beta \neq 0$

$$\beta^t A \beta = \left[\sum_{i=1}^{M} \beta_i B_i, \sum_{i=1}^{M} \beta_i B_i\right] \ge \lambda_0 \rho^{-2} \left\| \sum_{i=1}^{M} \beta_i B_i \right\|_2 > 0$$

en utilisant (III.13) et A est donc définie positive.
De plus, il résulte de (III.15) que $H = A$, et que β^* est, par conséquent, l'unique minimum de F sur $\mathbb{R}^M$ ∎

De plus, on peut obtenir une borne générale de l'erreur :

Théorème III.3

> Si λ_S est l'élément qui minimise $F[\lambda]$ sur S
> alors $|\lambda^* - \lambda_S| = [\lambda^* - \lambda_S, \lambda^* - \lambda_S]^{1/2} = \inf_{w \in S} |\lambda^* - w|$ (III.16)

Démonstration : Soit $w \in S$, alors $F[w] = \frac{1}{2}[w,w] - \langle w(0), x_0 \rangle$
Donc $F[w] - F[\lambda^*] = \frac{1}{2}[w,w] - \frac{1}{2}[\lambda^*, \lambda^*] + \langle x_0, \lambda^*(0) - w(0) \rangle$
Posons $w = \eta$ dans (12). On obtient $[\lambda^*, \eta] = [\lambda^*, w] = \langle x_0, w(0) \rangle$
De même, en posant $\lambda^* = \eta$, on obtient $[\lambda^*, \lambda^*] = \langle x_0, \lambda^*(0) \rangle$
Donc

$$\begin{aligned} F[w] - F[\lambda^*] &= \tfrac{1}{2}[w,w] + \tfrac{1}{2}[\lambda^*, \lambda^*] + \langle x_0, -w(0) \rangle \\ &= \tfrac{1}{2}[w,w] + \tfrac{1}{2}[\lambda^*, \lambda^*] - [\lambda^*, w] \\ &= \tfrac{1}{2}[\lambda^* - w, \lambda^* - w] = \tfrac{1}{2}|\lambda^* - w|^2 \end{aligned}$$

en utilisant la linéarité de $[\lambda, \eta]$

Par conséquent

$$|\lambda^* - \lambda_S|^2 = 2\left[F[\lambda_S] - F[\lambda^*]\right] \leq 2\left[F[w] - F[\lambda^*]\right] = |\lambda^* - w|^2$$

et on obtient

$$\inf_{w \in S} |\lambda^* - w| \leq |\lambda^* - \lambda_S| \leq \inf_{w \in S} |\lambda^* - w|$$

ce qui prouve le théorème ■

Des théorèmes précédents, on peut alors déduire le

Théorème III 4

> Si λ_S est l'élément qui minimise $F[\lambda]$ sur S
> alors $\|\lambda^* - \lambda_S\|_2 \leq [\lambda_Q]^{-1/2} \inf_{w \in S} |\lambda^* - w|$ (III.17)
>
> $$\|\dot{\lambda}^* - \dot{\lambda}_S\|_2 \leq \lambda_Q^{-1/2} \left[\|Q\| + \|A^t\|_2\right] \inf_{w \in S} |\lambda^* - w| \quad \text{(III.18)}$$

Nous pouvons remarquer que (III.16) implique que λ_S est la projection de λ^* par rapport au produit scalaire $[\cdot,\cdot]$.
Du résultat précédent, on peut déduire les résultats suivants

Théorème III.5

> Si $u_S(t) \equiv -R^{-1}(t) B^t(t) \lambda_S(t)$, $t \in [0,T]$, est le contrôle calculé (optimal dans S), alors
>
> $$\|u^* - u_S\|_2 \leq \|R^{-1} B^t\|_\infty \lambda_Q^{-1/2} \inf_{w \in S} |\lambda^* - w| \quad \text{(III.19)}$$
>
> où $\|R^{-1}B^t\|_\infty = \sup_{t \in [0,T]} |R^{-1} B^t(t)|_r$

Démonstration : Soit $\delta_S(t) = u^*(t) - u_S(t)$

alors $\delta_s(t) = -R^{-1}(t) B^t(t) [\lambda^*(t) - \lambda_s(t)]$

par linéarité de (III 6)

Donc $\|\delta_s(t)\|_2 = \|R^{-1} B^t (\lambda^* - \lambda_s)\|_2 \leq \|R^{-1} B^t\|_\infty \|\lambda^* - \lambda_s\|_2$

et $\|\lambda^* - \lambda_s\|_2 \leq \lambda \underline{q}^{-1/2} \inf_{w \in S} |\lambda^* - w|$ (III.17)

Ce qui prouve (III.19) ∎

Définissons maintenant les quantités suivantes

$$\Gamma = T \|B\|_\infty \exp\left[\int_0^T |A(z)|_n \, dz\right]$$

$$\|A\|_\infty = \sup_{t \in [0,T]} |A(t)|_n$$

$$\|B\|_\infty = \sup_{t \in [0,T]} \left\{ |B(t) w|_n \mid w \in \mathbb{R}^r \text{ et } |w|_r = 1 \right\}$$

Nous pouvons alors démontrer le

Théorème III 6

Si $\dot{x}_s(t) = A(t) x_s(t) + B(t) u_s(t) \quad t \in [0,T]$
et $x_s(0) = x_0$, alors

$$\|x^* - x_s\|_2 \leq \Gamma \|R^{-1} B^t\|_\infty \lambda \underline{q}^{-1/2} \inf_{w \in S} |\lambda^* - w| \quad \text{(III 20)}$$

$$\|\dot{x}^* - \dot{x}_s\|_2 \leq \left[\Gamma \|A\|_\infty + \|B\|_\infty\right] \|R^{-1} B^t\|_\infty \lambda \underline{q}^{-1/2} \inf_{w \in S} |\lambda^* - w| \quad \text{(III 21)}$$

Démonstration : Posons $\varepsilon_s(t) = x^*(t) - x_s(t)$ pour $t \in [0,T]$.
Alors $\dot{\varepsilon}_s(t) = A\varepsilon_s + B[u^* - u_s]$ pour $t \in [0,T]$ par linéarité, et, de plus, $\varepsilon_s(0) = 0$ (par hypothèse)

Donc $\varepsilon_s(t) = \int_0^t A(z)\,\varepsilon_s(z)\,dz + \int_0^t B(z)\,\delta_s(z)\,dz$

En appliquant alors l'inégalité de Gronwall (Réf. B-2), on obtient

$$|\varepsilon_s(t)|_n \le T^{1/2}\,\|B\|_\infty\,\|\delta_s\|_2 \exp\left[\int_0^T |A(z)|_n\,dz\right]$$

ce qui donne

$$\|\varepsilon_s\|_2^2 = \int_0^T |\varepsilon_s(t)|_n^2\,dt \le T^2\,\|B\|_\infty^2\,\|\delta_s\|_2^2 \exp\left[2\int_0^T |A(z)|_n\,dz\right]$$

soit $\|\varepsilon_s\|_2^2 \le \Gamma^2\,\|u^* - u_s\|_2^2$

qui, combiné avec (III.19), prouve (III.20).

De plus,

$$|\dot\varepsilon_s(t)|_n \le \|A\|_\infty\,|\varepsilon_s(t)|_n + \|B\|_\infty\,|u^*(t) - u_s(t)|_r \quad \text{pour } t \in [0,T]$$

Donc

$$\|\dot\varepsilon_s\|_2^2 = \int_0^T |\dot\varepsilon_s(t)|_n^2\,dt \le \int_0^T \left\{\|A\|_\infty\,|\varepsilon_s| + \|B\|_\infty\,|u^* - u_s|\right\}^2 dt$$

$$\le \left\| \left\{\|A\|_\infty\,|\varepsilon_s| + \|B\|_\infty\,|u^* - u_s|\right\} \right\|_2^2$$

Ce qui donne

$$\|\dot\varepsilon_s\|_2 \le \left\| \|A\|_\infty\,|\varepsilon_s| + \|B\|_\infty\,|u^* - u_s| \right\|_2 \le \|A\|_\infty\,\|\varepsilon_s\|_2 + \|B_\infty\|\,\|u^* - u_s\|_2$$

$$\le \left[\Gamma\,\|A\|_\infty + \|B\|_\infty\right] \|u^* - u_s\|_2$$

qui, avec (III.19), prouve (III.21) ■

Ce théorème nous donne donc des bornes sur les erreurs de trajectoires et de vitesse du système approximé.

Nous pouvons étendre ce genre de résultat en donnant une borne d'erreur pour la fonctionnelle de coût $J[u,x]$

Théorème III.7

Sous les hypothèses précédentes

$$0 \le J[u_s, x_s] - J[u^*, x^*] \le \tfrac{1}{2}\,\|R^{-1}B^t\|_\infty^2\,\lambda_g^{-1}\,\zeta^2\left[\|Q\|_\infty\,\Gamma^2 + \|R\|_\infty\right] \inf_{w \in \Omega_s} |\lambda^* - w|$$

(III.22)

Démonstration : Si $\delta_s(t)$ et $\varepsilon_s(t)$ sont définis comme plus haut, alors :

$$J[u_s, x_s] = \frac{1}{2}\int_0^T \langle x_s, Q x_s\rangle_n \, dt + \frac{1}{2}\int_0^T \langle u_s, R u_s\rangle_r \, dt$$

$$= \frac{1}{2}\int_0^T \langle x^* + \varepsilon_s, Q(x^*+\varepsilon_s)\rangle_n \, dt + \frac{1}{2}\int_0^T \langle u^* + \delta_s, R(u^*+\delta_s)\rangle_r \, dt$$

$$= J[u^*, x^*] + \int_0^T \langle \delta_s, R u^*\rangle_r dt + \int_0^T \langle \varepsilon_s, Q x^*\rangle_n dt$$

$$+ \frac{1}{2}\int_0^T \langle \delta_s, R\delta_s\rangle_r dt + \frac{1}{2}\int_0^T \langle \varepsilon_s, Q\varepsilon_s\rangle_n dt$$

car Q est symétrique.
Mais, par (III.6), on trouve

$$\int_0^T \langle \delta_s, R u^*\rangle_r dt = -\int_0^T \langle \delta_s, B^t \lambda^*\rangle_n dt = -\int_0^T \langle B\delta_s, \lambda^*\rangle_n dt$$

Or $\dot{\varepsilon}_s(t) = A(t)\varepsilon_s(t) + B(t)\delta_s(t)$, on obtient

$$\int_0^T \langle \delta_s, R u^*\rangle_r dt = -\int_0^T \langle \dot{\varepsilon}_s - A\varepsilon_s, \lambda^*\rangle_n dt$$

$$= -\left[\langle \varepsilon_s, \lambda^*\rangle\right]_0^T + \int_0^T \langle \varepsilon_s, \dot{\lambda}^*\rangle_n dt + \int_0^T \langle A\varepsilon_s, \lambda^*\rangle_n dt$$

$$= 0 + \int_0^T \langle \varepsilon_s, \dot{\lambda}^* + A^t\lambda^*\rangle_n = -\int_0^T \langle \varepsilon_s, Q x^*\rangle_n dt$$

car $\varepsilon_s(0) = 0$ et $\lambda^*(T) = 0$ car $\lambda^* \in \underline{\Phi}_0^n$

Donc

$$J[u_s, x_s] = J[u^*, x^*] + \frac{1}{2}\int_0^T \langle \delta_s, R\delta_s\rangle dt + \frac{1}{2}\int_0^T \langle \varepsilon_s, Q\varepsilon_s\rangle dt$$

$$\leq J[u^*, x^*] + \frac{1}{2}\|R\|_\infty \|\delta_s\|_2^2 + \frac{1}{2}\|Q\|_\infty \|\varepsilon_s\|_2^2$$

$$\leq J[u^*, x^*] + \frac{1}{2}\|R^{-1}B^t\|_\infty^2 \lambda_S^{-1} \xi_S^t \inf_{w\in S}|\lambda^* - w| \left[\|Q\|_\infty \Gamma^2 + \|R\|_\infty\right]$$

en utilisant (III.8) et (III.20). ∎

§3 Une particularisation de S

Il est clair que le choix d'une base finie pour engendrer le sous-espace S de $\mathcal{D}_0'$ est d'une grande importance pratique. Le calcul des solutions optimales en est, en effet, fort tributaire.

Nous nous contenterons de donner quelques renseignements supplémentaires pour ce choix : considérons l'espace (de dimension finie) des fonctions splines

$$S = S_d(\Delta) = \left\{ \sum_{i=1}^{p} \beta_i B_{d,i}(t) \mid \beta_i \in \mathbb{R}^n \right\}$$

En fait, si l'on choisit $\Delta : 0 = x_0 < x_1 < \cdots < x_{N+1} = T$ et $h = \max_{0 \le i \le N} (x_{i+1} - x_i)$

$$S_d(\Delta) = \{ S(x) \text{ polynômes par morceaux, de degré } d, \text{ tels que } S(x) \in C^d[0,T] \text{ et } S(T) = 0 \}$$

Dans ce cas, on peut montrer (Cfr Ref S-1) que si chaque composante de λ^* est $d+1$ fois continûment différentiable par morceaux par rapport à t, il existe une constante K_d positive et indépendante de la subdivision Δ telle que

$$\inf_{w \in S} |\lambda^* - w| \le K_d h^d \| D^{d+1} \lambda^* \|_2$$

où h est le "module" de la subdivision Δ.
On obtient alors le théorème suivant

Théorème III.8

Sous les hypothèses précédentes, et si chaque composante de λ^* est $d+1$ fois continûment différentiable par morceaux par rapport à t, il existe une constante $K_d \ge 0$ indépendante de la subdivision Δ telle que

$$\| \lambda^* - \lambda_{S_d(\Delta)} \|_2 \le \lambda_g^{-1/2} \varsigma K_d h^d \| D^{d+1} \lambda^* \|_2$$

$$\| u^* - u_{S_d(\Delta)} \|_2 \le \| R^{-1} B^t \|_\infty \lambda_g^{-1/2} \varsigma K_d h^d \| D^{d+1} \lambda^* \|_2$$

$$\|x^* - x_{d(\Delta)}\|_2 \leq \Gamma \|R^{-1}B^t\|_\infty \lambda_g^{-1/2} \varsigma K_d h^d \|D^{d+1}\lambda^*\|_2$$

$$\|\dot{x}^* - \dot{x}_{d(\Delta)}\|_2 \leq \left(\Gamma\|A\|_\infty + \|B\|_\infty\right)\|R^{-1}B^t\|_\infty \lambda_g^{-1/2} \varsigma K_d h^d \|D^{d+1}\lambda^*\|_2$$

$$J[u_{d(\Delta)}, x_{d(\Delta)}] - J[u^*, x^*] \geq 0$$

$$\geq \frac{1}{2}\lambda_g^{-1}\varsigma^2 \|R^{-1}B^t\|_\infty^2 \left(\|g\|_\infty \Gamma^2 + \|R\|_\infty\right) K_d h^d \|D^{d+1}\lambda^*\|_2$$

La démonstration est évidente en utilisant tous les théorèmes démontrés précédemment.

III Système différentiel non linéaire

Nous allons maintenant considérer un système plus général : celui introduit au Chapitre II, IV, § 2

Reposons brièvement le problème.

Considérons la fonctionnelle de coût

$$J[u] = \int_0^T g(x,u,t)\,dt \qquad \text{(III 23)}$$

et les équations du système

$$\dot{x}(t) = f(x,u,t) \qquad x(0) = x_0\,, \quad t \in [0,T] \qquad \text{(III 24)}$$

où $x(t)$ est un vecteur à n dimensions, $u(t)$ un vecteur à r dimensions $f(x,u,t)$ un vecteur à n dimensions et $g(x,u,t)$ une fonction scalaire.

Nous supposons de plus que

$$U = \{u\} = \{W_2^q[0,T], \mathbb{R}^r\} \qquad \text{(III 25)}$$

$$G = \{x\} = \{W_2^q[0,T], \mathbb{R}^n\} \qquad \text{(III 26)}$$

où $\{W_2^q[0,T], \mathbb{R}^k\}$ est l'espace de Sobolev des fonctions vectorielles de dimension k définies sur $[0,T]$ qui ont leurs dérivées au sens des distributions dans $\{L^2[0,T], \mathbb{R}^k\}$ jusqu'à l'ordre q.

On peut aussi caractériser $\{W_2^a[0,T], \mathbb{R}^k\}$ comme la collection des fonctions z définies sur $[0,T]$ telles que $z^{(j)}$, $j \leq a-1$, est absolument continue dans $[0,T]$ et $z^{(a)} \in \{L^2[0,T], \mathbb{R}^k\}$. La norme de Sobolev sur cet espace est donnée par

$$\|z\|_{2,a}^2 \equiv \sum_{i=1}^{k} \int_0^T \sum_{j=0}^{a} [z_i^{(j)}(t)]^2 dt \qquad (III\,27)$$

Nous utiliserons aussi la norme L^2 ($\|.\|_{2,0}$ ou $\|.\|_2$)

1 § Formulations du problème et hypothèses

Notre problème consiste donc à minimiser $J[u]$ sur tous les u appartenant à $\{W_2^a[0,T], \mathbb{R}^r\}$

Nous pouvons l'envisager sous l'angle du calcul des variations et construire le Lagrangien :

$$L[u,x;\lambda,\gamma] \equiv J[u] + \int_0^T \langle \lambda(t), -\dot{x}(t) + f(x,u,t) \rangle dt + \langle x(0) - x_0, \gamma \rangle \qquad (III\,28)$$

où λ et γ appartiennent respectivement à $\{W_2^a[0,T], \mathbb{R}^n\}'$ et $\mathbb{R}^n$. Comme ces espaces sont des Hilbert, $\gamma \in \mathbb{R}^n$ et $\lambda \in \{W_2^a[0,T], \mathbb{R}^n\}$ par identification (th. de Riesz). Il s'agit donc maintenant de trouver x^*, u^*, λ^* et γ^* tels que

$$L[u^*,x^*;\lambda^*,\gamma^*] = \sup_{\substack{\lambda \in G \\ \gamma \in \mathbb{R}^n}} \left\{ \inf_{\substack{u \in U \\ x \in G}} L[u,x;\lambda,\gamma] \right\} \qquad (III\,29)$$

Il faut donc que les dérivées partielles (au sens de Fréchet) vérifient

$$\frac{\partial L}{\partial u}[u^*,x^*;\lambda^*,\gamma^*] = 0 \qquad (III\,30)$$

$$\frac{\partial L}{\partial x}[u^*,x^*;\lambda^*,\gamma^*] = 0 \qquad (III\,31)$$

$$\frac{\partial L}{\partial \lambda}[u^*,x^*;\lambda^*,\gamma^*] = 0 \qquad (III\,32)$$

$$\frac{\partial L}{\partial \gamma}[u^*,x^*;\lambda^*,\gamma^*] = 0 \qquad (III\,33)$$

Faisons maintenant des hypothèses sur la régularité du problème (sous la première forme). Posons la

Définition

> Le problème $\in BC^\alpha[0,T]$ (α est un entier positif) ssi
>
> (i) $f(x,u,t)$ et $g(x,u,t)$ sont $\alpha+1$ fois continûment différentiables en x et u et α fois en t pour $t \in [0,T]$
>
> (ii) les opérateurs (non linéaires) $\dfrac{\partial^{i+j} f}{\partial x^i \partial u^j}$ et $\dfrac{\partial^{i+j} g}{\partial x^i \partial u^j}$ (i et $j \in \mathbb{N}_0$ et $i+j \leq \alpha+1$) transforment des domaines bornés de $G \times U$ en domaines bornés de $\{L^2[0,T], \mathbb{R}^n\}$ et $L^2[0,T]$ respectivement

Nous supposons maintenant

Hypothèse 1 : le problème $\in BC^\alpha[0,T]$

Hypothèse 2 : Le lagrangien est extrémisé en $(u^*, x^*; \lambda^*, \gamma^*)$ satisfaisant les conditions (III.30) à (III.33)

Le principe de Pontryagine nous fournit les conditions nécessaires (cf I, II)

$$\begin{cases} \dfrac{\partial g^*}{\partial u} + \left[\dfrac{\partial f^*}{\partial u}\right]^t \lambda^* = 0 & \text{(III 34)} \\ \dot{\lambda}^* + \left[\dfrac{\partial f^*}{\partial x}\right]^t \lambda^* + \dfrac{\partial g^*}{\partial x} = 0 \quad \text{avec } \lambda^*(T) = 0 & \text{(III 35)} \\ -\dot{x}^* + f(x^*, u^*, t) = 0 \quad \text{avec } x^*(0) = x_0 & \text{(III 36)} \\ \gamma^* = -\lambda^*(0) & \text{(III 37)} \end{cases}$$

Hypothèse 3 : La seconde variation de L (par rapport à (u,x)) est fortement positive dans un voisinage convexe de (u^*, x^*, λ^*)

i.e. il existe des voisinages $N(u^*) \subseteq U$, $N(x^*) \subseteq G$ et $N(\lambda^*) \subseteq G$ tels que l'opérateur H défini par

$$H[u,x;\lambda] = \left[\begin{array}{c|c} \dfrac{\partial^2 g}{\partial u^2} + \left[\dfrac{\partial^2 f}{\partial u^2}\right]^t \lambda & \dfrac{\partial^2 g}{\partial u \partial x} + \left[\dfrac{\partial^2 f}{\partial u \partial x}\right]^t \lambda \\ \hline \dfrac{\partial^2 g}{\partial x \partial u} + \left[\dfrac{\partial^2 f}{\partial x \partial u}\right]^t \lambda & \dfrac{\partial^2 g}{\partial x^2} + \left[\dfrac{\partial^2 f}{\partial x^2}\right]^t \lambda \end{array}\right] \qquad \text{(III.38)}$$

satisfasse

$$\int_0^T \left\langle \begin{bmatrix} \delta u \\ \delta x \end{bmatrix}, H[\tilde{u},\tilde{x};\tilde{\lambda}] \begin{bmatrix} \delta u \\ \delta x \end{bmatrix} \right\rangle dt \geq \sigma \left[\|\delta u\|_2^2 + \|\delta x\|_2^2\right] \qquad \text{(III.39)}$$

où $\delta u = u - \tilde{u}$ et $\delta x = x - \tilde{x}$ avec $(\tilde{u},\tilde{x},\tilde{\lambda}) \in N(u^*) \times N(x^*) \times N(\lambda^*)$, et où σ est une constante strictement positive.

A ce stade, quelques remarques s'imposent. D'abord, les hypothèses 1, 2 et 3 constitue un ensemble de conditions suffisantes locales d'existence et d'unicité de la solution du problème (Cfr Réf A-1). Ensuite, on peut voir, à l'aide d'arguments classiques en équations différentielles, que $x^*, \lambda^* \in \{C^n[0,T], \mathbb{R}^n\}$ et $u^* \in \{C^n[0,T], \mathbb{R}^r\}$ (Cfr Réf C-1). Enfin, nous interchangerons sup et inf dans (III 29), pourvu qu'on reste dans un voisinage convexe convenable du point optimal (Cfr Réf L-1)

Décrivons maintenant le comportement du Lagrangien au point optimal

Théorème III 9

> Si $x \in N(x^*)$, $u \in N(u^*)$, $\lambda \in N(\lambda^*)$ et $\gamma^* \in N(\gamma^*)$, le Lagrangien a un point de selle en $(u^* x^* \lambda^* \gamma^*)$ avec
>
> $$L[u^*,x^*;\lambda,\gamma] = L[u^*,x^*;\lambda^*,\gamma^*] \leq L[u,x;\lambda^*,\gamma^*] \qquad \text{(III 40)}$$

Démonstration : L'égalité de gauche est immédiate, si l'on se rappelle que le couple (x^*, u^*) doit satisfaire (III 24) pour tout couple $(\lambda, \gamma) \in G \times \mathbb{R}^n$ admissible. L'inégalité de droite s'obtient en développant L en série de Taylor autour de *.

$$L[u,x;\lambda^*,\gamma^*] = L[u^*,x^*;\lambda^*,\gamma^*]$$
$$+ \int_0^T \left\langle \frac{\partial g(x^*,u^*,t)}{\partial u} + \left[\frac{\partial f(x^*,u^*,t)}{\partial u}\right]^t \lambda^*, \Delta u \right\rangle dt$$
$$+ \int_0^T \left\langle \frac{\partial g(x^*,u^*,t)}{\partial x} + \left[\frac{\partial f(x^*,u^*,t)}{\partial x}\right]^t \lambda^* + \dot{\lambda}^*, \Delta x \right\rangle dt$$
$$+ \langle \gamma^* + \lambda^*(0), \Delta x(0)\rangle - \langle \lambda^*(T), \Delta x(T)\rangle$$
$$+ \int_0^1 (1-\tau) \int_0^T \left\langle \begin{bmatrix}\Delta u \\ \Delta x\end{bmatrix}, H[\tilde{u},\tilde{x},\lambda^*] \begin{bmatrix}\Delta u \\ \Delta x\end{bmatrix} \right\rangle dt\, d\tau \qquad (III.41)$$

où $\tilde{u} = \tau u + (1-\tau)u^*$, $\tilde{x} = \tau x + (1-\tau)x^*$ et $\Delta u = u - u^*$, $\Delta x = x - x^*$ avec $\tau \in [0,1]$. Le développement provient de la Réf D-1. Notons que toutes les variations du premier ordre dans (III 41) sont nulles en vertu des conditions (III 34) à (III 37). De plus l'Hypothèse 3 implique

$$L[u,x;\lambda^*,\gamma^*] \geq L[u^*,x^*;\lambda^*,\gamma^*] + \frac{\sigma}{2}\left[\|\Delta u\|_2^2 + \|\Delta x\|_2^2\right]$$

qui démontre le théorème ∎

A partir de maintenant, nous éliminerons γ du Lagrangien en ne considérant plus que les variations qui satisfont $\lambda(0) = -\gamma$.

Nous allons maintenant choisir, comme dans le cas précédent, des sous-espaces de dimension finie

2§ Sous espaces de dimension finie et approximation de Ritz-Galerkin

Définissons les ensembles admissibles pour u, x et λ par D_u, D_x et D_λ, ensembles de fonctions vectorielles de dimension r et n, qui sont fonctions continûment différentiables par morceaux sur $[0,T]$

Soient

$$C \equiv \{u \mid u \in U \cap N(u^*) \wedge D_u\}$$

$$S \equiv \{x \mid x \in G \cap N(x^*) \cap D_x\}$$

$$M \equiv \{\lambda \mid \lambda \in G \cap N(\lambda^*) \cap D_\lambda\}$$

où $N(u^*)$ et $N(x^*)$ et $N(\lambda^*)$ sont déterminés par l'hypothèse de positivité forte.

Choisissons

$$C_m(w) \equiv \{u \mid u(t) = \sum_{i=1}^{m} a_i w_{i,m}(t),\ t \in [0,T]\} \subseteq U$$

$$S_m(w) \equiv \{x \mid x(t) = \sum_{i=1}^{m} \beta_i w_{i,m}(t),\ t \in [0,T]\} \subseteq G$$

$$M_m(w) \equiv \{\lambda \mid \lambda(t) = \sum_{i=1}^{m} \xi_i w_{i,m}(t),\ t \in [0,T]\} \subseteq G$$

où $\{w_{i,m}\}$ représente des bases qu'on peut toujours choisir orthonormales, et où les a_i sont des vecteurs de $\mathbb{R}^r$ et les β_i et ξ_i de $\mathbb{R}^n$.

Choisissons $\{w_{i,m}\}$ tels que

$$C_m = \{u \mid u \in C_m(w) \cap \bar{C}\}$$

$$S_m = \{x \mid x \in S_m(w) \cap \bar{S}\}$$

$$M_m = \{\lambda \mid \lambda \in M_m(w) \cap \bar{M}\}$$

soient non-vides. Comme C_m, S_m et M_m sont des convexes fermés de W_2^1, on peut voir que ces ensembles possèdent des meilleures approximations de u^*, x^* & λ^* uniques. (cf Réf L-1, C-1)
On a donc

$$\|\hat{u} - u^*\|_{2,1} = \inf_{u \in C_m} \|u - u^*\|_{2,1}$$

$$\|\hat{x} - x^*\|_{2,1} = \inf_{x \in S_m} \|x - x^*\|_{2,1}$$

$$\|\hat{\lambda} - \lambda^*\|_{2,1} = \inf_{\lambda \in M_m} \|\lambda - \lambda^*\|_{2,1}$$

relations qui déterminent les "meilleures approximations" $\hat{u}, \hat{x}$ et $\hat{\lambda}$.
Posons maintenant

$$\varepsilon(\hat{u}) = \hat{u} - u^* \quad , \quad \varepsilon(\hat{x}) = \hat{x} - x^* \quad , \quad \varepsilon(\hat{\lambda}) = \hat{\lambda} - \lambda^* \qquad (III\,42)$$

Nous supposerons évidemment que C_m, S_m et H_m sont de "bons" sous espaces approximants, c.à.d que $\|\varepsilon(\hat{u})\|_{2,1}$, $\|\varepsilon(\hat{x})\|_{2,1}$ et $\|\varepsilon(\hat{\lambda})\|_{2,1}$ tendent vers zéro quand m tend vers l'infini.

Posons maintenant le problème dans les sous-espaces considérés : il faut minimiser $J[u]$ sur C_m. Soit

$$J[\bar{u}] = \inf_{u \in C_m} J[u]$$

sous la contrainte

$$\begin{cases} \int_0^T \langle w_{j,m}(t), -\dot{\bar{x}}(t) + f(\bar{x}(t), \bar{u}(t), t) \rangle dt = 0 \\ \bar{x}(0) = x_0 \end{cases}$$

avec $\bar{x} \in S_m$ et $j = 1, m$
En terme de sup inf, cela donne

$$L[\bar{u}, \bar{x}; \bar{\lambda}] = \sup_{\lambda \in H_m} \{ \inf_{\substack{x \in S_m \\ u \in C_m}} L[u, x; \lambda] \}$$

Pour les dérivées, il faut trouver $(\bar{u}, \bar{x}, \bar{\lambda}) \in C_m \times S_m \times H_m$ tel que

$$\frac{\partial L}{\partial u}(\bar{u}, \bar{x}, \bar{\lambda}) = 0$$

$$\frac{\partial L}{\partial x}(\bar{u}, \bar{x}, \bar{\lambda}) = 0$$

$$\frac{\partial L}{\partial \lambda}(\bar{u}, \bar{x}, \bar{\lambda}) = 0$$

Ces différentes formulations, bien sûr équivalentes, fournissent le système d'équations algébriques suivant

$$\int_0^T \langle w_{j,m}(t),\ \frac{\partial g}{\partial u}(\bar{x},\bar{u},t) + \left[\frac{\partial f}{\partial u}(\bar{x},\bar{u},t)\right]^t \bar{\lambda} \rangle\, dt = 0 \qquad \text{(III 43)}$$

$$\int_0^T \langle w_{j,m}(t),\ \dot{\bar{\lambda}} + \left[\frac{\partial f}{\partial x}(\bar{x},\bar{u},t)\right]^t \bar{\lambda} + \frac{\partial g}{\partial x}(\bar{x},\bar{u},t) \rangle\, dt = 0 \qquad \text{(III 44)}$$

$$\bar{\lambda}(T) = 0 \qquad \text{(III 45)}$$

$$\int_0^T \langle w_{j,m}(t),\ -\dot{\bar{x}} + f(\bar{x},\bar{u},t) \rangle\, dt = 0 \qquad \text{(III 46)}$$

$$\bar{x}(0) = x_0 \qquad \text{(III 47)}$$

où $(\bar{u},\bar{x},\bar{\lambda}) \in C_m \times S_m \times H_m$ et $j = 1,m$.
Exactement comme plus haut, on peut déduire le

Théorème III 10

> Soit $(\bar{u},\bar{x},\bar{\lambda})$ la solution du système (III 43) à (III 47)
> Si $(u,x,\lambda) \in C_m \times S_m \times H_m$, le Lagrangien a un point de selle dégénéré en $(\bar{u},\bar{x},\bar{\lambda})$, avec
>
> $$L[\bar{u},\bar{x},\lambda] = L[\bar{u},\bar{x},\bar{\lambda}] \leq L[u,x,\bar{\lambda}] \qquad \text{(III 48)}$$

Nous allons maintenant démontrer la "qualité" de l'approximation de Ritz Galerkin en dérivant des bornes d'erreurs données en termes des normes de $\varepsilon(\hat{u})$, $\varepsilon(\hat{x})$, $\varepsilon(\hat{\lambda})$, $\varepsilon(\hat{\xi})$ et $\varepsilon(\hat{\zeta})$, où les deux dernières sont définies par

$$\varepsilon(\hat{\xi}) = \hat{\xi} - \bar{\xi} \quad , \quad \varepsilon(\hat{\zeta}) = \hat{\zeta} - \bar{\zeta}$$

avec $(\hat{\xi},\hat{\zeta})$ la meilleure approximation L^2 de $(\bar{\xi},\bar{\zeta})$. La définition formelle de ξ et ζ sera faite ultérieurement.

Définissons d'abord les quantités suivantes :

$$J^* = J[u^*] \equiv L[u^*, x^*, \lambda^*] \qquad (III\,49)$$

$$J_m \equiv J[\bar{u}] \equiv L[\bar{u}, \bar{x}; \bar{\lambda}] \qquad (III\,50)'$$

où $(\bar{u}, \bar{x}, \bar{\lambda})$ est la solution du système de Galerkin (III 43) à (III 47)

3§ Convergence du couple (x, u).

Nous démontrerons d'abord quelques résultats préparatoires

Théorème III 11

Soit $(u, x, \lambda) \in P_m = C_m \times S_m \times M_m$. Il existe alors des constantes $k_1, k_2 > 0$ telles que

$$\left\| \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \lambda \right\|_2 \leq k_1 (\|\Delta u\|_2 + \|\Delta x\|_2) + \|\ell\|_2 \qquad (III\,51)$$

et

$$\left\| \dot{\lambda} + \left[\frac{\partial f^*}{\partial x}\right]^t \lambda + \frac{\partial g^*}{\partial x} \right\|_2 \leq k_2 (\|\Delta u\|_2 + \|\Delta x\|_2) + \|\zeta\|_2 \qquad (III\,52)$$

où $\ell(t) = \ell(x, u, \lambda; t) = \dfrac{\partial g}{\partial u}(x, u, t) + \dfrac{\partial f}{\partial u}[(x, u, t)]^t \lambda$ (III 53)

$$\zeta(t) = \zeta(x, u, \lambda, t) = \dot{\lambda} + \left[\frac{\partial f}{\partial x}(x, u, t)\right]^t \lambda + \frac{\partial g}{\partial x}(x, u, t) \qquad (III\,54)$$

avec $\Delta x = x - x^*$ et $\Delta u = u - u^*$

Démonstration Nous avons évidemment

$$\frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \lambda = \ell(t) + \left[\frac{\partial g^*}{\partial u} - \frac{\partial g}{\partial u}(x, u, t)\right] + \left\{\left[\frac{\partial f^*}{\partial u}\right]^t - \left[\frac{\partial f}{\partial u}(x, u, t)\right]^t\right\} \lambda$$

En appliquant le théorème de la valeur moyenne pour les opérateurs, dont les hypothèses font intervenir les hypothèses 1, (cf Rél D) on obtient

$$\left\| \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \lambda \right\|_2 \leq \|\ell\|_2 + k_g [\|\Delta u\|_2 + \|\Delta x\|_2] + k_f [\|\Delta u\|_2 + \|\Delta x\|_2]$$

qui se ramène à (III 51) en posant $k_1 = k_g + k_f$.

La preuve de (III 52) est identique ∎

Théorème III 12

Si $(\bar{u},\bar{x},\bar{\lambda})$ est la solution du système de Galerkin sur P_m. Soient $\hat{\bar{\varepsilon}}$ et $\hat{\bar{\zeta}}$ les meilleures approximations L^2 de $\varepsilon(\bar{x},\bar{u},\bar{\lambda},t)$ et $\zeta(\bar{x},\bar{u},\bar{\lambda},t)$ respectivement. Alors

$$\|\bar{\varepsilon}\|_2 = \|\varepsilon(\hat{\bar{\varepsilon}})\|_2 \qquad \text{(III 55)}$$

$$\|\bar{\zeta}\|_2 = \|\varepsilon(\hat{\bar{\zeta}})\|_2 \qquad \text{(III 56)}$$

Démonstration : Il suffit de voir que $\hat{\bar{\varepsilon}}$ et $\hat{\bar{\zeta}}$ sont nulles. Calculons les coefficients de Fourier pour $\bar{\varepsilon}$ et $\bar{\zeta}$

$$\langle \bar{\varepsilon}, w_{i,m}\rangle_{L^2} = \int_0^T w_{i,m}(t)\, \varepsilon(\bar{x},\bar{u},\bar{\lambda},t)\, dt = 0$$

par (III 43) pour $j=1,m$.

$$\langle \bar{\zeta}, w_{i,m}\rangle_{L^2} = \int_0^T w_{i,m}(t)\, \zeta(\bar{x},\bar{u},\bar{\lambda},t)\, dt = 0$$

par (III.44) pour $j=1,m$. ∎

Le théorème a pour seul but d'uniformiser les notations
Démontrons maintenant la convergence

Théorème III 13

Si $(\bar{u},\bar{x},\bar{\lambda}) \in P_m$ est la solution du système de Galerkin, il existe des constantes $k,k',k''>0$ indépendantes de m, telles que

$$\|\bar{x} - x^*\|_2 \leq \eta_m \qquad \text{(III 57)}$$

$$\|\bar{u} - u^*\|_2 \leq \tau_m \qquad \text{(III.58)}$$

où les constantes η_m et τ_m sont données par les relations suivantes :

$$\eta_m = k''\left[\|\varepsilon(\hat{u})\|_2 + \|\varepsilon(\hat{x})\|_2 + \|\varepsilon(\hat{\lambda})\|_{2,1}\right]$$
$$+ k\left[\|\varepsilon(\hat{\xi})\|_2 \|\varepsilon(\hat{u})\|_2 + \|\varepsilon(\hat{\zeta})\|_2 \|\varepsilon(\hat{x})\|_2 + \|\varepsilon(\hat{x})\|_2^2 + \|\varepsilon(\hat{u})\|_2^2 + \|\varepsilon(\hat{\lambda})\|_{2,1}^2\right]^{1/2} \quad (III.59)$$

$$\sigma_m = k'\left[\|\varepsilon(\hat{u})\|_2 + \|\varepsilon(\hat{x})\|_2 + \|\varepsilon(\hat{\lambda})\|_2\right]$$
$$+ k\left[\|\varepsilon(\hat{\xi})\|_2 \|\varepsilon(\hat{u})\|_2 + \|\varepsilon(\hat{\zeta})\|_2 \|\varepsilon(\hat{x})\|_2 + \|\varepsilon(\hat{u})\|_2^2 + \|\varepsilon(\hat{x})\|_2^2 + \|\varepsilon(\hat{\lambda})\|_{2,1}^2\right]^{1/2} \quad (III.60)$$

Démonstration : Utilisons le théorème III 10 en posant $x = \hat{x}$ et $u = \hat{u}$. On obtient

$$J_m \leq L[\hat{u}, \hat{x}, \bar{\lambda}]$$

Développons maintenant le membre de droite en série de Taylor autour de $(u^*, x^*, \bar{\lambda})$. On obtient alors

$$J_m \leq L[u^*, x^*, \bar{\lambda}] + \int_0^T \left\langle \frac{\partial g^*}{\partial u} + \left[\frac{\partial f}{\partial u}\right]^t \bar{\lambda}, \varepsilon(\hat{u}) \right\rangle dt$$

$$+ \int_0^T \left\langle \dot{\bar{\lambda}} + \left[\frac{\partial f^*}{\partial x}\right]^t \bar{\lambda} + \frac{\partial g^*}{\partial x}, \varepsilon(\hat{x}) \right\rangle dt$$

$$+ \int_0^1 (1-\tau) \int_0^T \left\langle \begin{bmatrix} \varepsilon(\hat{\xi}) \\ \varepsilon(\hat{\zeta}) \end{bmatrix}, H(\tilde{u}, \tilde{x}; \bar{\lambda}) \begin{bmatrix} \varepsilon(\hat{u}) \\ \varepsilon(\hat{x}) \end{bmatrix} \right\rangle dt\, d\tau$$

où $\tilde{x} = \tau x^* + (1-\tau)\hat{x}$ et $\tilde{u} = \tau u^* + (1-\tau)\hat{u}$.

D'autre part, nous avons

$$L[u^*, x^*; \lambda^*] = J^* = L[u^*, x^*; \bar{\lambda}] = J[u^*, x^*; \hat{\lambda}] \quad (III.61)$$

en vertu du théorème III 9

Prenons les normes

$$J_m \leq J^* + \left\| \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \bar{\lambda} \right\|_2 \|\varepsilon(\hat{u})\|_2$$
$$+ \left\| \dot{\bar{\lambda}} + \left[\frac{\partial f^*}{\partial x}\right]^t \bar{\lambda} + \frac{\partial g^*}{\partial x} \right\|_2 \|\varepsilon(\hat{x})\|_2$$
$$+ k\left[\|\varepsilon(\hat{u})\|_2^2 + \|\varepsilon(\hat{x})\|_2^2\right]$$

En utilisant le théorème III 11 on déduit immédiatement

$$J_m \le J^* + c\left[\|\bar u - u^*\|_2 \left(\|\varepsilon(\bar u)\|_2 + \|\varepsilon(\hat x)\|_2\right) + \|\bar x - x^*\|_2 \left(\|\varepsilon(\hat u)\|_2 + \|\varepsilon(\hat x)\|_2\right)\right] \qquad (III.62)$$
$$+ \|\varepsilon(\hat\ell)\|_2 \|\varepsilon(\hat u)\|_2 + \|\varepsilon(\hat g)\|_2 \|\varepsilon(\hat u)\|_2$$
$$+ k\left(\|\varepsilon(\hat u)\|_2^2 + \|\varepsilon(\hat x)\|_2^2\right)$$

Utilisons maintenant l'égalité du théorème (III.10) où nous choisissons $\lambda = \hat\lambda$. Nous avons donc

$$J_m = L[\bar u, \bar x; \hat\lambda]$$

Développons le terme de droite en série de Taylor autour de $(u^*, x^*, \hat\lambda)$

$$J_m = L[u^*, x^*; \hat\lambda] + \int_0^T \left\langle \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \hat\lambda, \bar u - u^* \right\rangle dt$$
$$+ \int_0^T \left\langle \dot{\hat\lambda} + \left[\frac{\partial f^*}{\partial x}\right]^t \hat\lambda + \frac{\partial g^*}{\partial x}, \bar x - x^* \right\rangle dt$$
$$+ \int_0^1 (1-\tau) \int_0^T \left\langle \begin{bmatrix} \bar u - u^* \\ \bar x - x^* \end{bmatrix}, H(u_\tau, x_\tau; \hat\lambda) \begin{bmatrix} \bar u - u^* \\ \bar x - x^* \end{bmatrix} \right\rangle dt\, d\tau$$

où $u_\tau = \tau \bar u + (1-\tau) u^*$ et $x_\tau = \tau \bar x + (1-\tau) x^*$. En utilisant (III.61) et la positivité forte on obtient

$$J_m \ge J^* - \left| \int_0^T \left\langle \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \hat\lambda, \bar u - u^* \right\rangle \right| - \left| \int_0^T \left\langle \dot{\hat\lambda} + \left[\frac{\partial f^*}{\partial x}\right]^t \hat\lambda + \frac{\partial g^*}{\partial x}, \bar x - x^* \right\rangle \right|$$
$$+ \frac{\alpha}{2}\left[\|\bar u - u^*\|_2^2 + \|\bar x - x^*\|_2^2\right]$$
$$\ge J^* - \left\| \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \hat\lambda \right\|_2 \|\bar u - u^*\|_2 - \left\| \dot{\hat\lambda} + \left[\frac{\partial f^*}{\partial x}\right]^t \hat\lambda + \frac{\partial g^*}{\partial x} \right\|_2 \|\bar x - x^*\|_2$$
$$+ \frac{\alpha}{2}\left[\|\bar u - u^*\|_2^2 + \|\bar x - x^*\|_2^2\right]$$

Or, nous savons que

$$\left\| \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \hat\lambda \right\|_2 = \left\| -\frac{\partial g^*}{\partial u} - \left[\frac{\partial f^*}{\partial u}\right]^t \lambda^* + \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \hat\lambda \right\|_2$$
$$= \left\| \left[\frac{\partial f^*}{\partial u}\right]^t \varepsilon(\hat\lambda) \right\|_2 \le k_0 \|\varepsilon(\hat\lambda)\|_2$$

De même

$$\| \dot{\hat\lambda} + \left[\frac{\partial f^*}{\partial x}\right]^+ \hat\lambda + \frac{\partial g^*}{\partial x} \|_2 = \| - \frac{\partial g^*}{\partial x} - \left[\frac{\partial f^*}{\partial x}\right]^+ \lambda^* - \dot\lambda^* + \frac{\partial g^*}{\partial x} + \left[\frac{\partial f^*}{\partial x}\right]^+ \hat\lambda + \dot{\hat\lambda} \|_2$$

$$= \| \varepsilon(\dot{\hat\lambda}) + \left[\frac{\partial f^*}{\partial x}\right]^+ \varepsilon(\hat\lambda) \|_2 \leq \| \varepsilon(\dot{\hat\lambda}) \|_2 + \| \left[\frac{\partial f^*}{\partial x}\right]^+ \varepsilon(\hat\lambda) \|_2$$

$$\leq k' \left[\| \varepsilon(\dot{\hat\lambda}) \|_2 + \| \varepsilon(\hat\lambda) \|_2 \right] \leq k'' \| \varepsilon(\hat\lambda) \|_{2,1}$$

en effet

$$\| \varepsilon(\hat\lambda) \|_{2,1}^2 = \| \varepsilon(\hat\lambda) \|_2^2 + \| \varepsilon(\dot{\hat\lambda}) \|_2^2$$

implique

$$\| \varepsilon(\hat\lambda) \|_{2,1}^2 \geq \| \varepsilon(\hat\lambda) \|_2^2 \quad \text{et donc} \quad \| \varepsilon(\hat\lambda) \|_{2,1} \geq \| \varepsilon(\hat\lambda) \|_2$$

$$\| \varepsilon(\dot{\hat\lambda}) \|_{2,1}^2 \geq \| \varepsilon(\dot{\hat\lambda}) \|_2^2 \quad \text{et donc} \quad \| \varepsilon(\dot{\hat\lambda}) \|_{2,1} \geq \| \varepsilon(\dot{\hat\lambda}) \|_2$$

ce qui donne $\| \varepsilon(\dot{\hat\lambda}) \|_2 + \| \varepsilon(\hat\lambda) \|_2 \leq 2 \| \varepsilon(\hat\lambda) \|_{2,1}$

On a choisi les constantes comme suit : $k' = \sup[1, k]$; $k'' = 2k'$

Si on choisit $c_0 = \sup[k_0, k'']$, on obtient alors

$$J_m \geq J^* - c_0 \left[\| \bar u - u^* \|_2 \| \varepsilon(\hat\lambda) \|_2 + \| \bar x - x^* \|_2 \| \varepsilon(\hat\lambda) \|_{2,1} \right] \qquad (\text{III } 63)$$

$$+ \frac{\sigma}{2} \left[\| \bar u - u^* \|_2^2 + \| \bar x - x^* \|_2^2 \right]$$

En combinant cette inégalité avec (III 62), on obtient

$$\| \varepsilon(\hat\xi) \|_2 \| \varepsilon(\hat u) \|_2 + \| \varepsilon(\hat\zeta) \|_2 \| \varepsilon(\hat x) \|_2 + k \left[\| \varepsilon(\hat u) \|_2^2 + \| \varepsilon(\hat x) \|_2^2 \right]$$

$$\geq - \| \bar u - u^* \|_2 \left[c_0 \| \varepsilon(\hat\lambda) \|_2 + c \left(\| \varepsilon(\hat u) \|_2 + \| \varepsilon(\hat x) \|_2 \right) \right]$$

$$- \| \bar x - x^* \|_2 \left[c_0 \| \varepsilon(\hat\lambda) \|_{2,1} + c \left(\| \varepsilon(\hat u) \|_2 + \| \varepsilon(\hat x) \|_2 \right) \right]$$

$$+ \sigma \left[\| \bar u - u^* \|_2^2 + \| \bar x - x^* \|_2^2 \right]$$

Or

$$k_0^2 \left(\| \varepsilon(\hat\xi) \|_2 \| \varepsilon(\hat u) \|_2 + \| \varepsilon(\hat\zeta) \|_2 \| \varepsilon(\hat x) \|_2 + \| \varepsilon(\hat u) \|_2^2 + \| \varepsilon(\hat x) \|_2^2 + \| \varepsilon(\hat\lambda) \|_{2,1}^2 \right)$$

$$\geq \left[\| \bar u - u^* \|_2 - k'' \left(\| \varepsilon(\hat u) \|_2 + \| \varepsilon(\hat x) \|_2 + \| \varepsilon(\hat\lambda) \|_2 \right) \right]^2 \qquad (\text{III } 64)$$

$$+ \left[\| \bar x - x^* \| - k'' \left(\| \varepsilon(\hat u) \|_2 + \| \varepsilon(\hat x) \|_2 + \| \varepsilon(\hat\lambda) \|_{2,1} \right) \right]^2$$

Le second membre est plus grand que

$$\Big[\|\bar u-u^*\|_2^2 - 2k'\|\bar u-u^*\|_2\,(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_2)$$
$$+k'^2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_2)^2+\|\bar x-x^*\|_2^2$$
$$-2k''\|\bar x-x^*\|_2\,(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_{2,1})^2$$
$$+k''^2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_{2,1})^2\Big]$$
$$\geq \|\bar u-u^*\|_2^2-2k'\|\bar u-u^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_2)$$
$$+\|\bar x-x^*\|_2^2-2k''\|\bar x-x^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_{2,1})$$
$$+k''^2(\|\varepsilon(\hat\lambda)\|_{2,1}^2+\|\varepsilon(\hat x)\|_2^2+\|\varepsilon(\hat u)\|_2^2)$$

Donc

$$\|\varepsilon(\hat\xi)\|_2\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat\xi)\|_2\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat u)\|_2^2+\|\varepsilon(\hat x)\|_2^2$$
$$\geq\Big\{\|\bar u-u^*\|_2^2+\|\bar x-x^*\|_2^2+(k''^2-1)\|\varepsilon(\hat\lambda)\|_{2,1}^2+k''^2(\|\varepsilon(\hat x)\|_2^2+\|\varepsilon(\hat u)\|_2^2)$$
$$-2k'\|\bar u-u^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_2)$$
$$-2k''\|\bar x-x^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_{2,1})\Big\}\frac{1}{k_0^2}$$

C'est à dire

$$\|\varepsilon(\hat\xi)\|_2\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat\xi)\|_2+\|\varepsilon(\hat x)\|_2+k(\|\varepsilon(\hat u)\|_2^2+\|\varepsilon(\hat x)\|_2^2)$$
$$\geq(k+k''^2-1)(\|\varepsilon(\hat u)\|_2^2+\|\varepsilon(\hat x)\|_2^2)+\frac{1}{k_0^2}\Big\{\|\bar u-u^*\|_2^2+\|\bar x-x^*\|_2^2$$
$$-2k'\|\bar u-u^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_2)$$
$$-2k''\|\bar x-x^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_{2,1})\Big\}$$
$$\geq\frac{1}{k_0^2}\Big\{\|\bar x-x^*\|_2^2+\|\bar u-u^*\|_2^2-2k''\|\bar x-x^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_{2,1})$$
$$-2k'\|\bar u-u^*\|_2(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2+\|\varepsilon(\hat\lambda)\|_2)\Big\}$$
$$\geq\sigma(\|\bar x-x^*\|_2^2+\|\bar u-u^*\|_2^2)-\|\bar u-u^*\|_2\big[C_0\|\varepsilon(\hat\lambda)\|_2+c(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2)\big]$$
$$-\|\bar x-x^*\|_2\big[C_0\|\varepsilon(\hat\lambda)\|_{2,1}+c(\|\varepsilon(\hat u)\|_2+\|\varepsilon(\hat x)\|_2)\big]$$

en choisissant $k_0\geq\sigma^{-1/2}$, $k'=\sup[c_0,c]\frac{k_0^2}{2}$ et $k''=\sup[1,k']$, ce qui prouve (III 64).

Résolvons (III.64) pour $\bar{u}-u^*$ et $\bar{x}-x^*$ séparement. Nous obtenons alors rapidement la thèse. ■

4 §. Convergence de la fonctionnelle de coût J_m

Nous allons maintenant chercher des bornes d'erreur pour la fonctionnelle de coût. Pour ce faire, nous allons introduire une approximation intermédiaire : nous cherchons dans $N(u^*) \times N(x^*) \times H_m$ un triplet $(\tilde{u}, \tilde{x}, \tilde{\lambda})$ tel que

$$I_m \equiv L[\tilde{u}, \tilde{x}; \tilde{\lambda}] = \sup_{\lambda \in H_m} \left\{ \inf_{\substack{u \in N(u^*) \\ x \in N(x^*)}} L[u, x; \lambda] \right\}$$

Nous pouvons alors obtenir les résultats suivants

Théorème III 14

> Sous les hypothèses précédentes, il existe une constante $k_1 > 0$ telle que
>
> $$J^* - k_1 \|\varepsilon(\hat{\lambda})\|_{2,1}^2 \leq I_m \leq J^* \qquad \text{(III.65)}$$

Démonstration : L'inégalité de droite est immédiate.

D'autre part

$$I_m \geq \inf_{\substack{x \in N(x^*) \\ u \in N(u^*)}} L[u, x; \hat{\lambda}]$$

où $\hat{\lambda}$ est défini comme plus haut. Développons le membre de droite en série de Taylor autour de $(u^*, x^*, \hat{\lambda})$:

$$L[ux; \hat{\lambda}] = L[u^*, x^*; \hat{\lambda}] + \int_0^T \left\langle \frac{\partial g^*}{\partial u} + \left[\frac{\partial f^*}{\partial u}\right]^t \hat{\lambda}, \Delta u \right\rangle dt$$

$$+ \int_0^T \left\langle \dot{\hat{\lambda}} + \left[\frac{\partial f^*}{\partial x}\right]^t \hat{\lambda} + \frac{\partial g^*}{\partial x}, \Delta x \right\rangle dt \qquad \text{(cont.)}$$

$$+\int_0^1 (1-\tau)\int_0^T \left\langle \begin{bmatrix}\Delta u\\ \Delta x\end{bmatrix}, H(\bar{\bar{u}},\bar{\bar{x}};\hat{\lambda})\begin{bmatrix}\Delta u\\ \Delta x\end{bmatrix}\right\rangle dt\, d\tau$$

où $\Delta x = x - x^*$, $\Delta u = u - u^*$, $\bar{\bar{x}} = \tau x + (1-\tau)x^*$ et $\bar{\bar{u}} = \tau u + (1-\tau)u^*$.

Exactement comme dans le théorème III.13 nous obtenons

$$L[u,x;\hat{\lambda}] \geq L^* - k_1' \|\epsilon(\hat{\lambda})\|_2 \|\Delta u\|_2 - k_2' \|\epsilon(\hat{\lambda})\|_{2,1} \|\Delta x\|_2 + \frac{\sigma}{2}\left(\|\Delta u\|_2^2 + \|\Delta x\|_2^2\right)$$

Or

$$\begin{aligned}
&L^* - k_1 \|\epsilon(\hat{\lambda})\|_{2,1}^2 + \frac{k_1}{2}\left[\left(\|\Delta x\|_2 - \|\epsilon(\hat{\lambda})\|_{2,1}\right)^2 + \left(\|\Delta u\|_2 - \|\epsilon(\hat{\lambda})\|_2\right)^2\right] \\
&= L^* - k_1 \|\epsilon(\hat{\lambda})\|_{2,1}^2 + \frac{k_1}{2}\left[\|\Delta x\|_2^2 - 2\|\Delta x\|.\|\epsilon(\hat{\lambda})\|_{2,1} + \|\epsilon(\hat{\lambda})\|_{2,1}^2\right. \\
&\qquad \left. + \|\Delta u\|_2^2 - 2\|\Delta u\|.\|\epsilon(\hat{\lambda})\|_2 + \|\epsilon(\hat{\lambda})\|_2^2\right] \\
&= L^* + \frac{k_1}{2}\left[\|\Delta x\|_2^2 - 2\|\Delta x\|_2 \|\epsilon(\hat{\lambda})\|_{2,1} - \|\epsilon(\hat{\lambda})\|_{2,1}^2 + \|\Delta u\|_2^2\right. \\
&\qquad \left. - 2\|\Delta u\|_2 \|\epsilon(\hat{\lambda})\|_2 + \|\epsilon(\hat{\lambda})\|_2^2\right] \\
&\leq L^* + \frac{k_1}{2}\left[\|\Delta x\|_2^2 - 2\|\Delta x\|_2 \|\epsilon(\hat{\lambda})\|_{2,1} + \|\Delta u\|_2^2 - 2\|\Delta u\|_2 \|\epsilon(\hat{\lambda})\|_2\right]
\end{aligned}$$

$$\leq L^* - k_1' \|\epsilon(\hat{\lambda})\|_2 \|\Delta u\|_2 - k_2' \|\epsilon(\hat{\lambda})\|_{2,1} \|\Delta x\|_2 + \frac{\sigma}{2}\left(\|\Delta u\|_2^2 + \|\Delta x\|_2^2\right)$$

si l'on choisit $k_1 = \inf\left[\sigma, \sup(k_1', k_2')\right]$

Nous avons donc

$$L[u\,x;\hat{\lambda}] \geq L^* - k_1 \|\epsilon(\hat{\lambda})\|_{2,1}^2 + \frac{k_1}{2}\left\{\left(\|\Delta x\|_2 - \|\epsilon(\hat{\lambda})\|_{2,1}\right)^2 + \left(\|\Delta u\|_2 - \|\epsilon(\hat{\lambda})\|_2\right)^2\right\}$$

Comme les termes entre $\{\}$ sont toujours positifs, on peut les supprimer et l'on obtient bien la thèse ■

Théorème III.15

Si $J_m = L[\bar{u},\bar{x},\bar{\lambda}]$ avec $(\bar{u},\bar{x},\bar{\lambda}) \in P_m$, il existe des constantes $k_1, k_2 > 0$ telles que

$$-\mu_m + J^* \leq J_m \leq J^* + \mu_m \qquad \text{(III C6)}$$

où $\mu_m = k_1 \|\varepsilon(\hat{\lambda})\|_{2,1}^2$ (III 67)

et

$$\nu_m = k_2\Big[(\eta_m+\sigma_m+\|\varepsilon(\hat{\rho})\|_2)\|\varepsilon(\hat{u})\|_2 + (\eta_m+\sigma_m+\|\varepsilon(\hat{\xi})\|_2)\|\varepsilon(\hat{x})\|_2+\|\varepsilon(\hat{u})\|_2^2+\|\varepsilon(\hat{x})\|_2^2\Big] \quad \text{(III 68)}$$

avec η_m et σ_m définis par (III 59) et (III 60)

Démonstration Commençons par remarquer que

$$I_m \leq J_m .$$

Le théorème précédent nous permet alors d'affirmer

$$J^* \doteq k_1\|\varepsilon(\hat{\lambda})\|_{2,1}^2 \leq I_m \leq J_m$$

ce qui prouve la borne inférieure.

D'autre part, en substituant (III 57) et (III 58) dans (III 63) on obtient

$$J_m \leq J^* + c\Big[\eta_m(\|\varepsilon(\hat{u})\|_2+\|\varepsilon(\hat{x})\|_2) + \sigma_m(\|\varepsilon(\hat{u})\|_2+\|\varepsilon(\hat{x})\|_2)+\|\varepsilon(\hat{\rho})\|_2\|\varepsilon(\hat{u})\|_2 + \|\varepsilon(\hat{\xi})\|_2\|\varepsilon(\hat{x})\|_2 + k(\|\varepsilon(\hat{u})\|_2^2+\|\varepsilon(\hat{x})\|_2^2)\Big]$$

ce qui donne, en rassemblant les termes, la borne supérieure ■

5§ Une particularisation de P_m

Nous allons maintenant indiquer quelques résultats pour un choix particulier du sous espace P_m : un espace de fonctions splines.

Définition

Supposons H_m^α ($\alpha \geq 1, m \geq 1$) un espace de dimension m de polynomes par morceaux d'ordre $\alpha-1$, tel que

1) $\exists\ \Sigma_m$ opérateur linéaire : $P^\alpha[0,T] \to H_m^\alpha$

2) $\forall f \in P^\alpha[0,T]\quad \|\Sigma_m f - f\|_2 = O(m^{-\alpha})$

3) $\forall f \in PC^\alpha[0,T] \quad \| \frac{d^\beta}{dt^\beta}(\mathcal{L}_m f - f)\|_2 = O(m^{\beta-\alpha}), \ 1 \le \beta \le \alpha$

4) $(\mathcal{L}_m f)^{(i)}(0) = f^{(i)}(0)$ et $(\mathcal{L}_m f)^{(i)}(T) = f^{(i)}(T)$, $i = 0, 1$

5) $H_m^\alpha \subseteq \mathcal{U}$ où $\mathcal{U}$ est un "ensemble admissible" (ex: G, U)

avec $PC^\alpha[0,T] = \{ f \mid f^{(\alpha)}$ est continue par morceaux sur $[0,T] \}$

Il est classique que de tels sous-espaces existent, à condition d'avoir des "bonnes conditions" sur $\mathcal{U}$. Comme, dans notre cas, G et U sont inclus dans W_2^α, toutes ces conditions sont satisfaites.

Si nous définissons un découpage de $[0,T]$ tel que

$$0 = t_0 < t_1 < \cdots < t_q = T$$

avec $\quad h = |\Pi| = \max_{i=1,q} |t_i - t_{i-1}|$

nous pouvons alors choisir $H_m^\alpha(\Pi)$ l'espace des splines d'ordre α pour lesquelles

$$\| \frac{d^\beta}{dt^\beta}(\mathcal{L}_m f - f)\|_2 = O(h^{\alpha-\beta}) \qquad 0 \le \beta \le \alpha$$

On peut alors démontrer les résultats suivants

Théorème III 16

Si $u_s, x_s, \lambda_s \in H_m^\alpha(\Pi)$ où (u_s, x_s, λ_s) est l'approximation spline de (u^*, x^*, λ^*), on a, pour $\alpha \ge 1$

$$\| \varepsilon(u_s)\|_{2,\beta} \le O(h^{\alpha-\beta}) \qquad 0 \le \beta \le \alpha$$
$$\| \varepsilon(x_s)\|_{2,\beta} \le O(h^{\alpha-\beta}) \qquad 0 \le \beta \le \alpha$$
$$\| \varepsilon(\lambda_s)\|_{2,\beta} \le O(h^{\alpha-\beta}) \qquad 0 \le \beta \le \alpha$$

Théorème III 17

Sous les hypothèses précédentes

$$\|\varepsilon(\hat{u})\|_{2,1} \leq \|\varepsilon(u_s)\|_{2,1}$$
$$\|\varepsilon(\hat{x})\|_{2,1} \leq \|\varepsilon(x_s)\|_{2,1}$$
$$\|\varepsilon(\hat{\lambda})\|_{2,1} \leq \|\varepsilon(\lambda_s)\|_{2,1}$$

Théorème III 18

Si $\bar{u}, \bar{x}$ et $\bar{\lambda} \in H_m^n(\Pi)$ sont solutions du système de Galerkin, alors

$$\|\varepsilon(\hat{\xi})\|_2 = \|\hat{\xi} - \bar{\xi}\|_2 \leq O(h^n)$$
$$\|\varepsilon(\hat{\zeta})\|_2 = \|\hat{\zeta} - \bar{\zeta}\|_2 \leq O(h^{n-1})$$

Ce dernier théorème fait intervenir, entre autres, l'hypothèse 1.

En utilisant ces résultats, nous obtenons les bornes suivantes

Théorème III 19

Si $\bar{u}, \bar{x}$ et $\bar{\lambda} \in H_m^n(\Pi)$ sont solutions du système de Galerkin, alors

$$\|\bar{x} - x^n\|_2 \leq O(h^{n-1})$$
$$\|\bar{u} - u^n\|_2 \leq O(h^{n-1})$$
$$J^* - O(h^{2(n-1)}) \leq J_m \leq J^* + O(h^{2(n-1)})$$

6 § Application au cas linéaire quadratique

Nous envisageons de nouveau la minimisation de

$$J[u,x] = \frac{1}{2}\int_0^T \{ \langle x, Qx \rangle + \langle u, Ru \rangle \}\, dt$$

sous la contrainte

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \qquad x(0) = x_0$$

avec $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^r$ pour tout $t \in [0,T]$

Nous supposons de plus

1) $R(t)$ et $Q(t)$ sont des matrices réelles symétriques et définies positives

2) $R(t), Q(t), A(t)$ et $B(t) \in PC^0[0,T]$

Les Hypothèses 1 à 3 sont alors évidemment vérifiées. La théorie générale fournit, en implémentant le système de Galerkin à l'aide des expressions

$$\bar{x}(t) = \sum_{i=1}^{m} a_i w_i(t)$$

$$\bar{u}(t) = \sum_{i=1}^{m} b_i w_i(t)$$

$$\bar{\lambda}(t) = \sum_{i=1}^{m} c_i w_i(t)$$

on obtient, après quelques calculs simples, le système

$$\begin{aligned} G\underline{a} + A\underline{a} + B\underline{b} &= -\underline{d} \\ G^t\underline{c} + A^t\underline{c} + Q\underline{a} &= 0 \\ R\underline{b} + B^t\underline{c} &= 0 \end{aligned}$$

où $G = (G_{ij})$ avec $G_{ij} = \int_0^T \langle \dot{w}_j(t), w_i(t) \rangle\, dt$

$A = (A_{ij})$ avec $A_{ij} = \int_0^T \langle w_j(t), A(t) w_i(t) \rangle\, dt$

et $\underset{\sim}{B}$, $\underset{\sim}{Q}$ et $\underset{\sim}{R}$ définis de manière analogue à $\underset{\sim}{A}$.
Nous avons aussi choisi

$$\underline{a}^t = (a_1^t, a_2^t, \dots, a_m^t)$$
$$\underline{b}^t = (b_1^t, b_2^t, \dots, b_m^t)$$
$$\underline{c}^t = (c_1^t, c_2^t, \dots, c_m^t)$$

et

$$\underline{d}^t = (x_0^t w_1(0), x_0^t w_2(0), \dots, x_0^t w_m(0))^t$$

Comme R et Q sont définies positives, $\underset{\sim}{R}$ et $\underset{\sim}{Q}$ le sont aussi. Nous pouvons alors réduire le système en

$$\left[(G+\underset{\sim}{A})\underset{\sim}{Q}^{-1}(G+\underset{\sim}{A})^t + \underset{\sim}{B}\underset{\sim}{R}^{-1}\underset{\sim}{B}^t\right]\underline{c} = \underline{d}$$

ou plus simplement

$$\boxed{F\underline{c} = \underline{d}}$$

où nous avons posé

$$F = \left[(G+\underset{\sim}{A})\underset{\sim}{Q}^{-1}(G+\underset{\sim}{A})^t + \underset{\sim}{B}\underset{\sim}{R}^{-1}\underset{\sim}{B}^t\right]$$

une matrice $(m.n) \times (m.n)$ symétrique et définie positive qui possède, quand on choisit les espaces "approximations" de manière adéquate (et possible!), des propriétés numériques intéressantes.

La résolution du système fournit alors tous les renseignements désirés.

IV Le régulateur d'état parabolique linéaire

Rappelons brièvement les données du problème.
Nous considérons le système parabolique linéaire suivant

$$\frac{\partial v}{\partial t} = A(x,t)\,\frac{\partial^2 v}{\partial x^2} + B(x,t)\,u(x,t) \qquad \text{(III 69)}$$

avec la condition initiale $\lim_{t\to 0^+} v(x,t) = v_0(x)$

et les conditions-limites

$$\alpha v(0,t) + \frac{\partial v(0,t)}{\partial x} = c\,f_1(t) \qquad \text{(III 70)}$$

$$\beta v(1,t) + \frac{\partial v(1,t)}{\partial x} = d\,f_2(t) \qquad \text{(III 71)}$$

où $x \in [0,1]$ et $t \in [0,T]$. La fonctionnelle du coût est

$$J[u,f_1,f_2] \equiv \frac{1}{2}\int_0^T \int_0^1 [\langle v, Q(x,t)v\rangle + \langle u, R(x,t)u\rangle]\,dx\,dt + \frac{1}{2}\int_0^T [\langle f_1, r(t)f_1\rangle + \langle f_2, s(t)f_2\rangle]\,dt \qquad \text{(III 72)}$$

où $v(\cdot)$ est un vecteur d'état à n dimensions, $u(\cdot)$ un vecteur de contrôle à r dimensions et $f_1(\cdot)$ et $f_2(\cdot)$ sont des vecteurs de contrôle aux frontières à n dimensions. Nous supposons de plus que $A(\cdot)$ est une matrice $n\times n$, $B(\cdot)$ une matrice $n\times r$, $Q(\cdot)$, $r(\cdot)$, $S(\cdot)$ des matrices $n\times n$ définies positives, et que $R(\cdot)$ est une matrice $r\times r$ définie positive. α,β,c et d sont des scalaires. De plus, nous exigeons que A, B, Q et $R \in PC^{\gamma}[0,1]\times PC^{\gamma-1}[0,T]$ pour $\gamma \geq 1$ et que $v_0(x) \in C^{\gamma}[0,1]$. T est fixé, $0 < r \leq n$. Nous n'imposons pas de contrainte supplémentaire sur $u(\cdot)$, $f_1(\cdot)$ et $f_2(\cdot)$. De plus, Q et R sont symétriques.

Nous imposons encore que A_u (notre ensemble de contrôles u) soit une partie de $\{V_2^0[(0,1)\times(0,T)], \mathbb{R}^r\}$ et que A_f (l'ensemble des f_i) soit une partie de $W_2^0[0,T]$. Nous désirons trouver (u^*, f_1^*, f_2^*, v^*) tel que

$$J[u^*,f_1^*,f_2^*] = \min_{\substack{u\in A_u \\ f_1\in A_f,\, f_2\in A_f}} J[u,f_1,f_2] \qquad \text{(III 73)}$$

1§ Formulations du problème.

Sous l'angle du calcul des variations, on peut attaquer le problème en introduisant des multiplicateurs de Lagrange vectoriels $\lambda_1, \lambda_2, \lambda_3$ et λ_4 et définir le lagrangien comme suit

$$\begin{aligned} L[u, f_1, f_2, v; \lambda_1, \lambda_2, \lambda_3, \lambda_4] &= L[\underline{u}, v; \underline{\lambda}] \\ &= J[\underline{u}] + \int_0^T \int_0^1 \left\langle \lambda_1(x,t), \frac{\partial v}{\partial t} + A \frac{\partial^2 v}{\partial x^2} + B u \right\rangle dx\, dt \\ &\quad + \int_0^T \left\langle \lambda_3(t), c f_1 - a v(0,t) - \frac{\partial v}{\partial x}(0,t) \right\rangle dt \\ &\quad + \int_0^T \left\langle \lambda_4(t), d f_2 - \beta v(1,t) - \frac{\partial v(1,t)}{\partial x} \right\rangle dt \\ &\quad + \int_0^1 \left\langle \lambda_2(t), v(x,0) - v_0(x) \right\rangle dx \end{aligned} \qquad (\text{III } 74)$$

où $\underline{u} \equiv (u, f_1, f_2) \in V_2^0[(0,1)\times(0,T)] \times W_2^0[0,T]^2$ et $\underline{\lambda} \equiv (\lambda_1, \lambda_2, \lambda_3, \lambda_4) \in V_2^1[(0,1)\times(0,T)] \times W_2^0[0,T]^2$.

On peut aussi formuler le problème de manière équivalente comme suit : trouver $v^*, \underline{u}^*$ et $\underline{\lambda}^*$ tels que

$$L[\underline{u}^*, v^*, \underline{\lambda}^*] = \max_{\underline{\lambda} \in A_{\underline{\lambda}}} \min_{\substack{u \in A_u \\ v \in A_v}} L[\underline{u}, v, \underline{\lambda}] \qquad (\text{III } 75)$$

où $A_{\underline{\lambda}} = A_{\lambda_1} \times A_{\lambda_2} \times A_{\lambda_3} \times A_{\lambda_4}$ avec $A_{\lambda_i} \subseteq \{V_2^0[(0,1)\times(0,T)] \times W_2^0[0,T], \mathbb{R}^1\}$ et $A_{\underline{u}} = A_u \times A_{f^2}$

Nous pouvons aussi demander de trouver $v^*, \underline{u}^*$ et $\underline{\lambda}^*$ tels que

$$L[\underline{u}^*, v^*, \underline{\lambda}^*] = \max_{\lambda \in A\underline{\lambda}} L[\underline{u}, v, \underline{\lambda}] \qquad (III\,76)$$

sous les conditions

$$\frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial v} = \frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial u} = \frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial f_1} = \frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial f_2} = 0 \qquad (III\,77)$$

(dérivées au sens de Fréchet).

L'équivalence de ces formulations est un résultat du calcul des variations (cfr. supra).

2§ Conditions nécessaires d'optimalité.

Effectuons d'abord quelques transformations sur le Lagrangien

$$\int_0^T \int_0^l \langle \lambda_1, -\frac{\partial v}{\partial t} + A\frac{\partial^2 v}{\partial x^2} + Bu \rangle \, dx\,dt = \int_0^l \int_0^T \langle \lambda_1, -\frac{\partial v}{\partial t} \rangle dx\,dt + \int_0^T \int_0^l \langle A^t\lambda_1, \frac{\partial^2 v}{\partial x^2} \rangle dx\,dt$$

$$+ \int_0^T \int_0^l \langle \lambda_1, Bu \rangle dx\,dt$$

$$= -\int_0^l \langle \lambda_1, v \rangle \Big]_0^T dx + \int_0^l \int_0^T \langle v, \frac{\partial \lambda_1}{\partial t} \rangle dt\,dx + \int_0^T \langle A^t\lambda_1, \frac{\partial v}{\partial x} \rangle \Big]_0^l dt$$

$$- \int_0^T \int_0^l \langle \frac{\partial A^t\lambda_1}{\partial x}, \frac{\partial v}{\partial x} \rangle dx\,dt + \int_0^T \int_0^l \langle \lambda_1, Bu \rangle dx\,dt$$

$$= -\int_0^l \langle \lambda_1, v \rangle \Big]_0^T dx + \int_0^T \int_0^l \langle v, \frac{\partial \lambda_1}{\partial t} \rangle dx\,dt + \int_0^T \langle A^t\lambda_1, \frac{\partial v}{\partial x} \rangle \Big]_0^l dt$$

$$- \int_0^T \langle v, \frac{\partial A^t\lambda_1}{\partial x} \rangle \Big]_0^l dt + \int_0^T \int_0^l \langle v, \frac{\partial^2 A^t\lambda_1}{\partial x^2} \rangle dx\,dt + \int_0^T \int_0^l \langle \lambda_1, Bu \rangle dx\,dt$$

$$= -\int_0^l \langle \lambda_1, v \rangle \Big]_0^T dx + \int_0^T \Big\{ \langle A^t\lambda_1, \frac{\partial v}{\partial x} \rangle - \langle v, \frac{\partial A^t\lambda_1}{\partial x} \rangle \Big\} \Big]_0^l dt$$

$$+ \int_0^T \int_0^l \Big\{ \langle v, \frac{\partial \lambda_1}{\partial t} + \frac{\partial^2 A^t\lambda_1}{\partial x^2} \rangle + \langle \lambda_1, Bu \rangle \Big\} dx\,dt$$

ce qui donne évidemment

$$L[\underline{u},v,\underline{\lambda}] = J[\underline{u}] - \int_0^1 \langle \lambda_1, v\rangle \Big|_0^T dx$$
$$+ \int_0^T \left\{ \langle A^t\lambda_1, \frac{\partial v}{\partial x}\rangle - \langle v, \frac{\partial A^t\lambda_1}{\partial x}\rangle \right\} \Big|_0^1 dt$$
$$+ \int_0^T \int_0^1 \left\{ \langle v, \frac{\partial \lambda_1}{\partial t} + \frac{\partial^2 A^t\lambda_1}{\partial x^2}\rangle + \langle \lambda_1, Bu\rangle \right\} dx\,dt$$
$$+ \int_0^T \left\{ \langle \lambda_3, c f_1 - \alpha v(0,t) - \frac{\partial v(0,t)}{\partial x}\rangle \right.$$
$$\left. + \langle \lambda_4, d f_2 - \beta v(1,t) - \frac{\partial v(1,t)}{\partial x}\rangle \right\} dt$$
$$+ \int_0^1 \langle \lambda_2, v(x,0) - v_0(x)\rangle\, dx \qquad \text{(III 78)}$$

De manière analogue aux problèmes précédents, on obtient les conditions nécessaires suivantes

$$\frac{\partial \lambda_1}{\partial t} + \frac{\partial^2 A^t\lambda_1}{\partial x^2} + Q(x,t)\, v(x,t) = 0 \qquad \text{(III 79)}$$

$$R(x,t)\, u(x,t) + B^t \lambda_1(x,t) = 0 \qquad \text{(III 80)}$$

$$\lambda_1(x,T) = 0 \qquad \lambda_2(x) + \lambda_1(x,0) = 0 \qquad \text{(III 81)}$$

$$A^t(0,t)\,\lambda_1(0,t) + \lambda_3(t) = 0 \qquad \text{(III 82)}$$

$$A^t(1,t)\,\lambda_1(1,t) - \lambda_4(t) = 0 \qquad \text{(III 83)}$$

$$c\,\lambda_3(t) + c(t)\, f_1(t) = 0 \qquad \text{(III 84)}$$

$$d\lambda_4(t) + s(t)\,\ell_2(t) = 0 \qquad \text{(III 85)}$$

$$\alpha(A^t\lambda_1)(0,t) + \frac{\partial(A^t\lambda_1)(0,t)}{\partial x} = 0 \qquad \text{(III 86)}$$

$$\beta(A^t\lambda_1)(1,t) + \frac{\partial(A^t\lambda_1)(1,t)}{\partial x} = 0 \qquad \text{(III 87)}$$

à l'optimum cherché.

Modifions le Lagrangien. Remarquons que

$$\int_0^1 \langle -\lambda_1(x,0), v(x,0)\rangle + \int_0^1 \langle \lambda_1(x,0), v_0(x)\rangle = \int_0^1 \langle \lambda_2, v(x,0) - v_0(x)\rangle$$

en utilisant (III 81) ; et donc

$$\int_0^1 \langle \lambda_2, v(x,0) - v_0(x)\rangle + \int_0^1 \langle \lambda_1(x,0), v(x,0)\rangle = \int_0^1 \langle \lambda_1(x,0), v_0(x)\rangle$$

De plus (III.81) implique

$$\int_0^1 \langle \lambda_1(x,T), v(x,T)\rangle = 0$$

Nous avons aussi que

$$\int_0^T \left\{ \int_0^1 \left\langle v, \frac{\partial \lambda_1}{\partial t} + \frac{\partial^2 A^t \lambda_1}{\partial x^2} \right\rangle + \langle B^t\lambda_1, u\rangle \right\} = -\int_0^T \left\{ \int_0^1 \langle Kv, gv\rangle + \langle u, Ru\rangle \right\}$$

en utilisant (III 79) et (III 80). De même, avec (III 84) et (III 85),

$$\int_0^T \left\{ \langle \lambda_3, c\ell_1\rangle + \langle \lambda_4, d\ell_2\rangle \right\} = -\int_0^T \left\{ \langle \ell_1, c\ell_1\rangle + \langle \ell_2, s\ell_2\rangle \right\}$$

Nous pouvons aussi déduire

$$\int_0^T \{ \langle \lambda_3, -\alpha v(0,t) - \frac{\partial v(0,t)}{\partial x} \rangle - \langle (A^t \lambda_1)(0,t), \frac{\partial v(0,t)}{\partial x} \rangle + \langle v(0,t), \frac{\partial (A^t \lambda_1)(0,t)}{\partial x} \rangle \}$$

$$= \int_0^T \langle \alpha (A^t \lambda_1)(0,t) + \frac{\partial (A^t \lambda_1)(0,t)}{\partial x}, v(0,t) \rangle = 0$$

en utilisant (III 82) et (III 86).
De même

$$\int_0^T \{ \langle \lambda_4, -\beta v(1,t) - \frac{\partial v(1,t)}{\partial x} \rangle + \langle (A^t \lambda_1)(1,t), \frac{\partial v(1,t)}{\partial x} \rangle - \langle v(1,t), \frac{\partial (A^t \lambda_1)(1,t)}{\partial x} \rangle \}$$

$$= - \int_0^T \langle \beta (A^t \lambda_1)(1,t) + \frac{\partial (A^t \lambda_1)(1,t)}{\partial x}, v(1,t) \rangle = 0$$

en utilisant (III.83) et (III 87).
En rassemblant toutes ces simplifications, on obtient

$$L[u, v, \lambda] = J[u] - \int_0^T \int_0^1 \{ \langle v, gv \rangle + \langle u, Ru \rangle \} dx\, dt$$

$$- \int_0^T \{ \langle f_1, r f_2 \rangle + \langle f_2, s f_2 \rangle \} dt + \int_0^1 \langle \lambda_1(x,0), v_0(x) \rangle dx$$

$$= - J[u] + \int_0^1 \langle \lambda_1(x,0), v_0(x) \rangle dx \qquad \text{(III 88)}$$

3§ Sous espace de dimension finie et approximation de Ritz Galerkin.

Comme dans les deux problèmes précédents, nous allons introduire un sous espace de dimension finie et tenter de résoudre le problème

initial sur ce sous-espace.

Soit S_h^γ, $\gamma \geq 1$ un espace de polynomes par morceaux d'ordre $\gamma - 1$ et d'une seule variable. h est ici le pas de la discrétisation. Alors, S_h^γ possède les propriétés suivantes

1. $\exists L_h$ linéaire $L_h : C^\gamma[a,b] \longrightarrow S_h^\gamma$
2. $\forall f \in C^\gamma[a,b] \quad \|L_h f - f\|_2 = O(h^\gamma)$
3. $\forall f \in C^\gamma[a,b] \quad \| \frac{d}{dt}(L_h f - f) \|_2 = O(h^{\gamma-1})$

Il est classique que de tels espaces existent, les espaces de fonctions splines par exemple.

Etendons maintenant, par un procédé simple, ces caractérisations, pour obtenir un espace de polynômes à deux variables.

Considérons d'abord Ω un rectangle fermé dans $\mathbb{R}^2$ défini par

$$\Omega = \{(p_1, p_2) \mid a_i \leq p_i \leq b_i, \ i = 1,2\}$$

où a_i et $b_i \in \mathbb{R}$ pour chaque i.

Soit Γ la frontière de Ω. Introduisons une partition π de Ω comme suit

$$\pi_i \mid a_i = x_i^0 < x_i^1 < \cdots < x_i^{m_i} = b_i \qquad i = 1,2$$

Soit, de plus

$$h_i = |\pi_i| = \max_{j \in J_{m_i}} \{x_j^i - x_{j-1}^i\}$$

Posons maintenant $\sum_M^\gamma = S_{m_1}^{\gamma_1} \times S_{m_2}^{\gamma_2}$

où $\gamma = \{\gamma_1, \gamma_2\}$ et $M = \{h_1, h_2\}$.

On en déduit immédiatement que $\sum_M^\gamma$ possède les propriétés

1. $\exists L_S$ bilinéaire $L_S : C^\gamma(\Omega) \longrightarrow \sum_M^\gamma$
2. $\forall f \in C^\gamma[\Omega]$

$$\|L_S f - f\|_2 = \sum_{j=1}^{2} O(h_j^{\gamma_j}) \equiv O(h^\gamma)$$

3. $\forall f \in C^{\gamma}[\Omega]$

$$\left\| \frac{\partial}{\partial x_j} (L_h f - f) \right\|_2 = \sum_{j=1}^{2} O(h_j^{\gamma_j - 1}) = O(h^{\gamma - 1})$$

où $\gamma = (\gamma_1, \gamma_2)$ et $h = (h_1, h_2)$.
Notre espace est donc construit.
Définissons maintenant:

Le problème initial $\in C^{\gamma}[(0,1) \times (0,T)]$

$\Longleftrightarrow$

(i) A, B, Q et R $\in C^{\gamma}$ par rapport à t
(ii) A, B, Q et R $\in C^{\gamma+1}$ par rapport à x
(iii) $V_0(x) \in C^{\gamma+1}$ par rapport à x
(iv) r(·) et s(·) $\in PC^{\gamma}$ par rapport à t

Nous pouvons alors voir le théorème suivant

Théorème III.20

Si le problème $\in C^{\gamma-1}[(0,1) \times (0,T)]$, alors
(i) $u^*, v^*, \lambda_1^* \in C^{\gamma}[(0,1) \times (0,T)]$
(ii) $f_1^*, f_2^*, \lambda_3^*, \lambda_4^* \in C^{\gamma}[0,T]$
(iii) $\lambda_2^* \in C^{\gamma}[0,1]$

Démonstration: Par (III.19) et (III.79), on voit immédiatement que $v^*(\cdot)$ et λ_1^* satisfont le système

$$\begin{bmatrix} \frac{\partial v^*}{\partial t} \\ \frac{\partial \lambda_1^*}{\partial t} \end{bmatrix} = \begin{bmatrix} A\frac{\partial^2(\cdot)}{\partial x^2} & -S(x,t) \\ -Q(x,t) & -\frac{\partial^2 A^t(\cdot)}{\partial x^2} \end{bmatrix} \begin{bmatrix} v^* \\ \lambda_1^* \end{bmatrix} \qquad (III\ 89)$$

avec les conditions initiales et aux limites données par (III 69)(III 70)(III 71) et (III 81 à 87). (III 80) implique $u^*(x,t) = -R^{-1}B^t\lambda_1^*$ et nous définissons alors

$$S(x,t) = B(x,t)R^{-1}(x,t)B^t(x,t)$$

Comme $B, R \in C^{\delta-1}[(0,1)\times(0,T)]$ et R est définie positive, $R^{-1} \in C^{\delta}[(0,1)\times(0,T)]$ et donc $S \in C^{\delta-1}[(0,1)\times(0,T)]$. Du système (III 89), on obtient évidemment que $v^*, \lambda_1^* \in C^{\delta}[(0,1)\times(0,T)]$

De plus $u^* = -R^{-1}B^t\lambda_1^*$ implique $u^* \in C^{\delta}[(0,1)\times(0,T)]$. (III 82) et $A \in C^{\delta}$ implique $\lambda_3 \in C^{\delta}[0,T]$, puisque (III 83), on a aussi $\lambda_4 \in C^{\delta}[0,T]$. (III. 84) et (III. 85) impliquent respectivement $f_1 \in C^{\delta}[0,T]$ et $f_2 \in C^{\delta}[0,T]$. Enfin (III. 81) implique $\lambda_2 \in C^{\delta}[0,T]$ ■

Comme précédemment, nous pouvons poser le problème: trouver $(\bar{u}, \bar{f}_1, \bar{f}_2) = \underline{\bar{u}}$ tel que

$$J[\underline{\bar{u}}] = \min_{\underline{u} \in A_{\underline{u}}} J[\underline{u}]$$

sous les contraintes

$$\int_0^1 \langle w_j(x) w_k(0), (\bar{v}(x,0) - v_0(x))_i \rangle dx = 0$$

$$\int_0^T \langle w_j(0), w_k(t), \left(c\bar{f}_1(t) - \alpha\bar{v}(0,t) - \frac{\partial \bar{v}}{\partial x}(0,t)\right)_i \rangle dt = 0$$

$$\int_0^T \langle w_j(1)\, w_k(1), \left(d\bar{f}_2(t) - \beta\bar{v}(1,t) - \frac{\partial \bar{v}}{\partial x}(1,t)\right)_i \rangle\, dt = 0$$

$$\int_0^T \int_0^1 \langle w_j(x)\, w_k(t), \left(-\frac{\partial \bar{v}}{\partial t} + A\frac{\partial^2 \bar{v}}{\partial x^2} + B\bar{v}\right)_i \rangle\, dx\, dt = 0$$

le système de Galerkin pour chaque fonction de base $w_j(x), w_k(t) \in \Sigma_H^r$, $j \in J_{m_1}$, $k \in J_{m_2}$ avec $i \in J_n$

On peut évidemment formuler cela comme suit :

trouver $\bar{u}, \bar{v}, \bar{f}_1$ et $\bar{f}_2$ tels que

$$L[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}] = \max_{\underline{\lambda} \in \Sigma_H^r} \min_{\substack{\underline{u} \in A_{\underline{u}} \\ v \in A_v}} L[\underline{u}, v, \underline{\lambda}] \qquad \text{(III 90)}$$

où $\underline{\lambda} \in \Sigma_H^r$ s'interprète naturellement comme $(\lambda_1)_i \in \Sigma_H^r$, $(\lambda_2)_i \in S_{m_1}^{\gamma_1}$, $(\lambda_3)_i \in S_{m_2}^{\gamma_2}$ et $(\lambda_4)_i \in S_{m_2}^{\delta_2}$ $(i = 1 - n)$.

On peut aussi chercher $\bar{\lambda}$ tel que

$$L[\underline{u}, v, \underline{\bar{\lambda}}] = \max_{\underline{\lambda} \in \Sigma_H^r} L[\underline{u}, v, \underline{\lambda}] \qquad \text{(III 91)}$$

sous les contraintes

$$\frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial v} = \frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial f_1} = \frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial f_2} = \frac{\partial L[\underline{u}, v, \underline{\lambda}]}{\partial u} = 0 \qquad \text{(III 92)}$$

Un algorithme effectif sera développé plus bas, en utilisant cette dernière formulation.

Appelons $\underline{\bar{u}}, \underline{\bar{\lambda}}$ et $\bar{v}$ les solutions du système de Galerkin et abordons le point suivant.

4§ Convergence du coût pour le quadruple $(\bar{u}, \bar{v}, \bar{f}_1, \bar{f}_2)$

Notre but est de voir comment $J[\underline{\bar{u}}]$ approxime $J[\underline{u}^*]$ où $\underline{u}^*$ est la solution optimale non approximée.

Théorème III.21

Si le problème initial $\in C^{\gamma-1}[(0,1)\times(0,T)]$ et si $\bar{u}, \bar{v}, \bar{f}_1, \bar{f}_2$ et $\underline{\bar{\lambda}}$ sont les solutions du système de Galerkin pour $\sum_M^\gamma$.
Si $\lambda_S \in \sum_M^\gamma$ où λ_S est une approximation de λ_1^* à $O(h^\gamma)$.
Si $(u_S, v_S, f_{1S}, f_{2S})$ sont engendrés par λ_S d'après (III.79 à 87)
Alors, pour $\gamma \geq 2$

$$\|\varepsilon_u\|_2 = O(h^\gamma) \qquad \|\varepsilon_v\| = O(h^{\gamma-2}) \qquad \text{(III.93)}$$
$$\|\varepsilon f_1\|_2 = O(h^\gamma) \qquad \|\varepsilon f_2\| = O(h^\gamma)$$

et

$$J[\underline{u}^*] \geq J[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}] \geq J[\underline{u}^*] + L[\varepsilon_u, \varepsilon_v, \varepsilon\underline{\lambda}] + \int_0^t \langle \varepsilon\lambda_2, v_0(x)\rangle dx \qquad \text{(III.94)}$$

où

$$\varepsilon_{\underline{u}} = \underline{u}_S - \underline{u}^*, \quad \varepsilon_v = v_S - v^* \text{ et } \varepsilon\underline{\lambda} = \underline{\lambda}_S - \underline{\lambda}^*$$

Démonstration: Puisque le problème initial $\in C^{\gamma-1}$, nous avons $\lambda_1^* \in C^\gamma[(0,1)\times(0,T)]$. Or par hypothèse, il existe un $\lambda_S \in \sum_M^\gamma$ tel que $\|\lambda_S - \lambda_1^*\|_2 = O(h^\gamma)$. Donc, de (III.79) à (III.87) on tire

$$\|\lambda_{3S} - \lambda_3^*\|_2 = \|A^t(0,t)(\lambda_S(0,t) - \lambda_1^*(0,t))\|_2 \leq O(h^\gamma)$$
$$\|\lambda_{4S} - \lambda_4^*\|_2 = \|A^t(1,t)(\lambda_1^*(1,t) - \lambda_S(1,t))\|_2 \leq O(h^\gamma)$$
$$\|u_S - u^*\|_2 = \|R^{-1}B^t(\lambda_1^* - \lambda_S)\|_2 \leq O(h^\gamma)$$
$$\|v_S - v^*\|_2 = \|g^{-1}\left[\frac{\partial}{\partial t}(\lambda_S - \lambda_1^*) + \frac{\partial^2}{\partial x^2}A^t(\lambda_S - \lambda_1^*)\right]\|_2 \leq O(h^{\gamma-2})$$
$$\|f_{1S} - f_1^*\|_2 = c\|r^{-1}(\lambda_{3S} - \lambda_3^*)\|_2 \leq O(h^\gamma)$$

$$\|f_{2s} - f_2^*\|_2 = d\,\|S^{-1}(\lambda_{4s} - \lambda_4^*)\|_2 \leq o(h^\gamma)$$

en se rappelant que r, s, A, B, R et Q sont continus sur des compacts, donc bornés. Nous avons donc prouvé (III 94)
Si nous nous souvenons qu'à l'extremum, le lagrangien est égal au coût, nous pouvons écrire, en utilisant (III.73)

$$J[\underline{u}^*] = L[\underline{u}^*, v^*, \underline{\lambda}^*] \geq L[\underline{u}_s, v_s, \underline{\lambda}_s]$$

puisque $\Sigma_M^\Gamma \subseteq A_{\underline{\lambda}}$.
De (III 91) on déduit que $L[\underline{u}_s, v, \underline{\lambda}_s] \geq L[\underline{u}_s, v_s, \underline{\lambda}_s]$
Mais

$$L[\underline{u}_s, v_s, \underline{\lambda}_s] = L[\underline{u}^* + \varepsilon\underline{u}, v^* + \varepsilon v, \underline{\lambda}^* + \varepsilon\underline{\lambda}]$$

$$= \frac{1}{2}\int_0^T \int_0^1 [\langle v^* + \varepsilon v, Q(v^* + \varepsilon v)\rangle + \langle u^* + \varepsilon u, R(u^* + \varepsilon u)\rangle]$$

$$+ \frac{1}{2}\int_0^T [\langle f_1^* + \varepsilon f_1, r(f_1^* + \varepsilon f_1)\rangle + \langle f_2^* + \varepsilon f_2, s(f_2^* + \varepsilon f_2)\rangle]$$

$$+ \int_0^T \int_0^1 \langle \lambda_1^* + \varepsilon\lambda_1, -\frac{\partial v^*}{\partial t} + A\frac{\partial^2 v^*}{\partial x^2} + Bu^* - \frac{\partial \varepsilon v}{\partial t} + A\frac{\partial^2 \varepsilon v}{\partial x^2} + B\varepsilon u\rangle$$

$$+ \int_0^T \langle \lambda_3^* + \varepsilon\lambda_3, c(f_1^* + \varepsilon f_1) - \alpha(v^*(0,t) + \varepsilon v(0,t)) - \frac{\partial v^*}{\partial x}(0,t) - \frac{\partial \varepsilon v}{\partial x}(0,t)\rangle$$

$$+ \int_0^T \langle \lambda_4^* + \varepsilon\lambda_4, d(f_2^* + \varepsilon f_2) - \beta(v^*(d,t) + \varepsilon v(d,t)) - \frac{\partial v^*}{\partial x}(d,t) - \frac{\partial \varepsilon v}{\partial x}(d,t)\rangle$$

$$+ \int_0^1 \langle \lambda_2^* + \varepsilon\lambda_2, v^*(x,0) + \varepsilon v(x,0) - v_0(x)\rangle$$

$$= \frac{1}{2}\int_0^T \int_0^1 [\langle v^*, Qv^*\rangle + \langle u^*, Ru^*\rangle] + \frac{1}{2}\int_0^T [\langle f_1^*, rf_1^*\rangle + \langle f_2^*, sf_2^*\rangle]$$

cont.

$$+\int_0^T\int_0^1 <\lambda_1^*, -\frac{\partial v^*}{\partial t}+A\frac{\partial^2 v^*}{\partial x^2}+Bw^*> + \int_0^T <\lambda_3^*, c f_1^* - \alpha v^*(0,t) - \frac{\partial v^*}{\partial x}(0,t)>$$

$$+\int_0^T <\lambda_4^*, d f_2^* - \beta v^*(1,t) - \frac{\partial v^*}{\partial x}(1,t)> + \int_0^1 <\lambda_2^*, v^*(x,0) - v_0(x)>$$

$$+\frac{1}{2}\int_0^T\int_0^1 [<\varepsilon_v, Q\varepsilon_v> + <\varepsilon_u, R\varepsilon_u>] + \frac{1}{2}\int_0^T [<\varepsilon f_1, \Gamma \varepsilon f_1> + <\varepsilon f_2, S\varepsilon f_2>]$$

$$+\int_0^T\int_0^1 <\varepsilon\lambda_1, -\frac{\partial \varepsilon v}{\partial t}+A\frac{\partial^2 \varepsilon v}{\partial x^2}+B\varepsilon_u> + \int_0^T <\varepsilon\lambda_3, c\varepsilon f_1 - \alpha \varepsilon_v(0,t) - \frac{\partial \varepsilon v}{\partial x}(0,t)>$$

$$+\int_0^T <\varepsilon\lambda_4, d\varepsilon f_2 - \beta\varepsilon_v(1,t) - \frac{\partial \varepsilon v}{\partial x}(1,t)> + \int_0^1 <\varepsilon\lambda_2, \varepsilon_v(x,0) - v_0(x)>$$

$$+\int_0^1 <\varepsilon\lambda_2, v_0(x)> + \frac{1}{2}\int_0^T\int_0^1 [<\varepsilon_v, Qv^*> + <\varepsilon_u, Rw^*>] + \frac{1}{2}\int_0^T [<\varepsilon f_1, \Gamma f_1^*> + <\varepsilon f_2, S f_2^*>]$$

$$+\int_0^T\int_0^1 \underline{<\varepsilon\lambda_1, -\frac{\partial v^*}{\partial t}+A\frac{\partial^2 v^*}{\partial x^2}+Bw^*>} + \int_0^T \underline{<\varepsilon\lambda_3, c f_1^* - \alpha v^*(0,t) - \frac{\partial v^*}{\partial x}(0,t)>}$$

$$+\int_0^T \underline{<\varepsilon\lambda_4, d f_2^* - \beta v^*(1,t) - \frac{\partial v^*}{\partial x}(1,t)>} + \int_0^1 \underline{<\varepsilon\lambda_2, v^*(x,0) - v_0(x)>}$$

$$+\frac{1}{2}\int_0^T\int_0^1 [<v^*, Q\varepsilon_v> + <u^*, R\varepsilon_u>] + \frac{1}{2}\int_0^T [<f_1^*, \Gamma\varepsilon f_1> + <f_2^*, S\varepsilon f_2>]$$

$$+\int_0^T <\lambda_3^*, -\frac{\partial \varepsilon v}{\partial t}+A\frac{\partial^2 \varepsilon v}{\partial x^2}+B\varepsilon_u> + \int_0^T <\lambda_3^*, c\varepsilon f_1 - \alpha\varepsilon_v(0,t) - \frac{\partial \varepsilon v}{\partial x}(0,t)>$$

$$+\int_0^T <\lambda_4^*, d\varepsilon f_2 - \beta\varepsilon_v(1,t) - \frac{\partial \varepsilon v}{\partial x}(1,t)> + \int_0^1 <\lambda_2^*, \varepsilon_v(x,0)> .$$

Les termes soulignés sont nuls car v^* est optimal et satisfait les conditions nécessaires (III.79) à (III.87).

D'autre part, puisque R et Q sont symétriques, on obtient bien que

$$L[u_s, v_s, \lambda_s] = L[u^*, v^*, \lambda^*] + L[\varepsilon_u, \varepsilon_v, \varepsilon\lambda]$$

$$+ \int_0^T \int_0^1 \underline{[\langle v^*, g\varepsilon_v\rangle + \langle u^*, R\varepsilon_u\rangle]} + \underline{\int_0^T [\langle f_1^*, r\varepsilon_{f_1}\rangle + \langle f_2^*, s\varepsilon_{f_2}\rangle]}$$

$$+ \int_0^1 [\langle \lambda_2^*, \varepsilon_v(x,0)\rangle + \langle \varepsilon_{\lambda_2}, v_0(x)\rangle] + \int_0^T \langle \lambda_3^*, c\varepsilon_{f_1} - \alpha\varepsilon_v(0,t) - \frac{\partial \varepsilon_v}{\partial x}(0,t)\rangle$$

$$+ \int_0^T \langle \lambda_4^*, d\varepsilon_{f_2} - \beta\varepsilon_v(1,t) - \frac{\partial \varepsilon_v}{\partial x}(1,t)\rangle + \int_0^T \int_0^1 \langle \lambda_1^*, -\frac{\partial \varepsilon_v}{\partial t} + A\frac{\partial^2 \varepsilon_v}{\partial x^2} + B\varepsilon_u\rangle$$

Dans cette égalité, le terme souligné vaut $2J[\varepsilon_u]$ par définition et les termes suivants $-2J[\varepsilon_u]$ par un calcul exactement semblable à celui qui a permis de déduire (III 88) (On excepte $\int_0^1 \langle \varepsilon_{\lambda_2}, v_0(x)\rangle$)

Nous avons donc

$$L[u_s, v_s, \lambda_s] = J[u^*] + L[\varepsilon_u, \varepsilon_v, \varepsilon\lambda] + \int_0^1 \langle \varepsilon_{\lambda_2}, v_0(x)\rangle \, dx$$

ce qui achève la démonstration ∎

Nous pouvons alors démontrer le corollaire suivant

Théorème III 22

> Sous les hypothèses du théorème III 21, on a
>
> $$0 \leq J[u^*] - J[\bar{u}] \leq O(h^{2(k-2)}) \qquad \text{(III 95)}$$

Démonstration : Remarquons d'abord, comme précédemment, que $J[\bar{u}] = L[\bar{u}, \bar{v}, \bar{\lambda}]$.

De plus, le théorème précédent implique

$$0 \le J[u^*] - L[\underline{\mu}, \bar{v}, \underline{\lambda}] \le -L[\varepsilon_{\underline{\mu}}, \varepsilon_v, \varepsilon_{\underline{\lambda}}] - \int_0^1 \langle \varepsilon\lambda_2, v_0(x) \rangle \, dx$$

Mais, en utilisant (III 74),

$$L[\varepsilon_{\underline{\mu}}, \varepsilon_v, \varepsilon_{\underline{\lambda}}] = \frac{1}{2}\int_0^T \int_0^1 [\langle \varepsilon_u, R\varepsilon_u \rangle + \langle \varepsilon_v, Q\varepsilon_v \rangle] + \frac{1}{2}\int_0^T [\langle \varepsilon\mu_1, r\varepsilon\mu_1 \rangle + \langle \varepsilon\mu_2, s\varepsilon\mu_2 \rangle]$$

$$+ \int_0^T \int_0^1 \langle \varepsilon\lambda_1, -\frac{\partial \varepsilon_v}{\partial t} + A\frac{\partial^2 \varepsilon_v}{\partial x^2} + B\varepsilon_u \rangle + \int_0^T \Big\{ \langle \varepsilon\lambda_3, \varepsilon\mu_3 - \alpha\varepsilon_v(0,t) - \frac{\partial \varepsilon_v(0,t)}{\partial x} \rangle$$

$$+ \langle \varepsilon\lambda_4, \alpha\varepsilon\mu_2 - \beta\varepsilon_v(1,t) - \frac{\partial \varepsilon_v(1,t)}{\partial x} \rangle \Big\} + \int_0^1 \langle \varepsilon\lambda_2, \varepsilon_v(x,0) - v_0(x) \rangle$$

En appliquant les conditions (III 79) à (III 85), car $\varepsilon_v = v_s - v^*$, on obtient

$$L[\varepsilon_{\underline{\mu}}, \varepsilon_v, \varepsilon_{\underline{\lambda}}] = -\frac{1}{2}\int_0^T \int_0^1 [\langle \varepsilon_v, Q\varepsilon_v \rangle + \langle \varepsilon_u, R\varepsilon_u \rangle] - \frac{1}{2}\int_0^T [\langle \varepsilon\mu_1, r\varepsilon\mu_1 \rangle + \langle \varepsilon\mu_2, s\varepsilon\mu_2 \rangle]$$

$$+ \int_0^1 \{ \langle \varepsilon\lambda_1(x,0), v_0(x) \rangle - \langle \varepsilon\lambda_1(x,T), \varepsilon_v(x,T) \rangle \}$$

$$+ \int_0^T \langle \varepsilon_v(0,t), \frac{\partial (A^t \varepsilon\lambda_1)(0,t)}{\partial x} + \alpha (A^t \varepsilon\lambda_1)(0,t) \rangle$$

$$- \int_0^T \langle \varepsilon_v(1,t), \frac{\partial (A^t \varepsilon\lambda_1)(1,t)}{\partial x} + \beta (A^t \varepsilon\lambda_1)(1,t) \rangle$$

En appliquant maintenant l'inégalité de Cauchy-Schwartz aux deux dernières intégrales, on déduit

$$0 \le J[u^*] - L[\underline{\mu}, \bar{v}, \underline{\lambda}] \le \frac{1}{2}\|\varepsilon_v\|_Q^2 + \frac{1}{2}\|\varepsilon_u\|_R^2 + \frac{1}{2}\|\varepsilon\mu_1\|_r^2 + \frac{1}{2}\|\varepsilon\mu_2\|_s^2$$

$$+ \|\varepsilon\lambda_1(x,0)\|_2 \|v_0(x)\|_2 + \|\varepsilon\lambda_1(x,T)\|_2 \|\varepsilon_v(x,T)\|_2$$

$$+ \|\varepsilon_v(0,t)\|_2 \Big\| \frac{\partial (A^t \varepsilon\lambda_1)(0,t)}{\partial x} + \alpha (A^t \varepsilon\lambda_1)(0,t) \Big\|_2$$

cont.

$$+\|\epsilon_v(1,t)\|_2 \left\| \frac{\partial (A^t \epsilon\lambda_1)(1,t)}{\partial x} + \beta (A^t \epsilon\lambda_1)(1,t) \right\|_2$$

$$- \|\epsilon\lambda_2\|_2 \|V_0(x)\|$$

$$\leq O(h^{2(\gamma-2)}) + O(h^{2\gamma}) + O(h^{2\gamma}) + O(h^{2\gamma}) + O(h^{\gamma}).O(h^{\gamma-2})$$

$$+O(h^{\gamma-2})O(h^{\gamma-2}) + O(h^{\gamma-2})O(h^{\gamma-2}) = O(h^{2(\gamma-2)})$$

Comme les normes $\|\cdot\|_g$, $\|\cdot\|_R$, $\|\cdot\|_r$ et $\|\cdot\|_s$ sont bien équivalentes aux normes $\|\cdot\|_2$ respectives (cfr th III 1), nous avons bien démontré la thèse ■

De la même manière, on peut voir le

Théorème III 23

Supposons vérifiées les hypothèses du théorème III 21.
Imposons de plus à Σ_m^r les conditions suivantes

$$\forall \lambda_1 \in \Sigma_m^r \quad \lambda_1(x,T) = 0 \qquad x \in (0,1)$$

$$a(A^t\lambda_1)(0,t) + \frac{\partial (A^t\lambda_1)(0,t)}{\partial x} = 0 \qquad t \in (0,T)$$

$$\beta (A^t\lambda_1)(1,t) + \frac{\partial (A^t\lambda_1)(1,t)}{\partial x} = 0 \qquad t \in (0,T) \qquad \text{(III 36)}$$

alors $0 \leq J[u^*] - J[\bar{u}] \leq O(h^{2(\gamma-2)})$

où $\bar{\lambda}, \bar{u}, \bar{v}$ est la solution de Galerkin sur le nouveau Σ_M^r

La démonstration est identique à celle du théorème précédent. ■

§5 Convergence en norme du quadruple $(\bar{u}, \bar{v}, \bar{f}_1, \bar{f}_2)$

Nous nous intéressons maintenant à la convergence en norme du quadruple $\bar{u}, \bar{v}, \bar{f}_1, \bar{f}_2$ vers u^*, v^*, f_1^*, f_2^*.
Prouvons le théorème suivant

Théorème III 24

> Si le problème initial $\in C^\gamma[(0,1)\times(0,T)]$,
> si $\underline{\bar{u}}, \bar{v}$ et $\bar{\lambda}$ sont les solutions sur le nouveau $\sum_M^\Gamma$ (i.e. $\sum_m^\Gamma$ muni des conditions du th. III 22), alors
> $$\|\bar{u} - u^*\|_R \leq O(h^{\gamma-2})$$
> $$\|\bar{f}_1 - f_1^*\|_r \leq O(h^{\gamma-2})$$
> $$\|\bar{f}_2 - f_2^*\|_s \leq O(h^{\gamma-2})$$
> $$\|\bar{v} - v^*\|_q \leq O(h^{\gamma-2})$$

Démonstration Développons $L[\underline{u}^*, v, \underline{\lambda}^*]$ en série de Taylor autour de $(\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}})$

$$L[\underline{u}^*, v^*, \underline{\lambda}^*] - L[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}] = L[\underline{u}^*, v^*, \underline{\bar{\lambda}}] - L[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}]$$
$$+ \frac{\partial L[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}]}{\partial u}(u^* - \bar{u}) + \frac{\partial L[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}]}{\partial f_1}(f_1^* - \bar{f}_1)$$
$$+ \frac{\partial L[\underline{\bar{u}}, \bar{v}, \bar{\lambda}]}{\partial f_2}(f_2^* - \bar{f}_2) + \frac{\partial L[\underline{\bar{u}}, \bar{v}, \underline{\bar{\lambda}}]}{\partial v}(v^* - \bar{v})$$
$$+ \int_0^T \int_0^1 (u^* - \bar{u} \;\; f_1^* - \bar{f}_1 \;\; f_2^* - \bar{f}_2 \;\; v^* - \bar{v}) H(x,t) \begin{bmatrix} u^* - \bar{u} \\ f_1^* - \bar{f}_1 \\ f_2^* - \bar{f}_2 \\ v^* - \bar{v} \end{bmatrix} dx\, dt$$

où $M(x,t)$ est défini par

$$M(x,t) = \begin{bmatrix} R(x,t) & 0 & 0 & 0 \\ 0 & r(t) & 0 & 0 \\ 0 & 0 & s(t) & 0 \\ 0 & 0 & 0 & q(x,t) \end{bmatrix}$$

(III 93) et (III 96) nous permettent d'assurer que les dérivées (au sens de Fréchet) du Lagrangien sont nulles en $(\underline{u}, v, \underline{\lambda})$. (III 95) donne alors

$$0 \leq L[\underline{u}^*, v^*, \underline{\lambda}^*] - L[\underline{u}, v, \underline{\lambda}] = \|u^* - u\|_R^2 + \|f_1^* - f_1\|_r^2 + \|f_2^* - f_2\|_s^2$$

$$+ \|v^* - v\|_q^2 \leq O(h^{2(\gamma-\varepsilon)}) \quad \blacksquare$$

La même remarque que supra s'impose à propos de l'équivalence des normes $\|\cdot\|_R \|\cdot\|_q \|\cdot\|_r \|\cdot\|_s$ avec la norme $\|\cdot\|_2$, puisque R, q, s, r sont définies positives.

Nous avons donc prouvé tout ce que nous désirions.

6§ Algorithme numérique de résolution

Remarquons d'abord que (III 79) (III 80) et (III 88) nous permettent d'écrire $L[\underline{u}, v, \underline{\lambda}] = \mathcal{L}[\lambda_1]$.

Nous voulons

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 0$$

Comme tout $(\lambda_1)_i \in \sum_M^r$, on peut exprimer λ_1 comme une combinaison linéaire

$$\lambda_1(x,t) = \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} c_{jk} w_j(x) w_k(t) \qquad x \in (0,1), t \in (0,T)$$

où chaque c_{jk} est un vecteur de dimension n.
On obtient alors successivement

$$\int_0^T \int_0^1 \langle u, Ru \rangle = \int_0^1 \int_0^T \langle R^{-1} B^t \lambda_1, B^t \lambda_1 \rangle = \int_0^T \int_0^1 \langle \lambda_1, B R^{-1} B^t \lambda_1 \rangle$$

$$= \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} \sum_{p=1}^{n_1} \sum_{q=1}^{n_2} c_{jk}^t \left[\int_0^T \int_0^1 E_{jkpq} \, dx \, dt \right] c_{pq}$$

où $E_{jkpq}(x,t) = w_j(x) w_k(t) w_p(x) w_q(t) \, B(x,t) R^{-1}(x,t) B^t(x,t)$
$\left(= \langle w_j(x) w_k(t), B(x,t) R^{-1}(x,t) B^t(x,t) w_p(x) w_q(t) \rangle \right)$

De même

$$\int_0^T \int_0^1 \langle v, Qv \rangle = \int_0^T \int_0^1 \left\langle Q^{-1} \left(\frac{\partial \lambda_1}{\partial t} + \frac{\partial^2 A^t \lambda_1}{\partial x^2} \right), \frac{\partial \lambda_1}{\partial t} + \frac{\partial^2 A^t \lambda_1}{\partial x^2} \right\rangle$$

$$= \sum_j \sum_k \sum_p \sum_q c_{jk}^t \left[\int_0^T \int_0^1 F_{jkpq} \, dx \, dt \right] c_{pq}$$

où $F_{jkpq}(x,t) = w_j(x) \dot{w}_k(t) w_p(x) \dot{w}_q(t) Q^{-1}$

$$+ w_p(x) \dot{w}_k(t) w_q(t) Q^{-1} \frac{\partial^2 w_j(x) A^t(x,t)}{\partial x^2}$$

$$+ w_j(x) w_k(t) \dot{w}_q(t) Q^{-1} \frac{\partial^2 w_p(x) A^t(x,t)}{\partial x^2}$$

$$+ w_k(t) w_q(t) \frac{\partial^2 w_j(x) A^t(x,t)}{\partial x^2} Q^{-1} \frac{\partial^2 w_p(x) A^t(x,t)}{\partial x^2}$$

et aussi, de la même façon,

$$\int_0^T \langle \psi_{1,r}(t), f_1 \rangle = \sum_j \sum_k \sum_p \sum_q c_{jk}^+ \left[\int_0^T G_{jkpq}\, dt \right] c_{pq}$$

où $G_{jkpq} = c^2 w_j(0) w_k(t) w_p(0) w_q(t)\, A(0,t)\, r^{-1}(t)\, A^+(0,t)$

et enfin

$$\int_0^T \langle f_2, s(t) f_2 \rangle = \sum_j \sum_k \sum_p \sum_q c_{jk}^+ \left[\int_0^T G'_{jkpq}\, dt \right] c_{pq}$$

où $G'_{jkpq} = d^2 w_j(1) w_k(t) w_p(1) w_q(t)\, A(1,t)\, s^{-1}(t)\, A^+(1,t)$

et

$$\int_0^1 \langle \lambda_1(x,0), v_0(x) \rangle dx = \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} c_{jk} \int_0^1 w_j(x) w_k(0) v_0(x)\, dx$$

Maximisons maintenant $L[c_{11}, c_{12}, \dots, c_{21}, \dots, c_{m_1 m_2}]$

$$\frac{\partial L}{\partial c_{pq}} = -\frac{1}{2} \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} \left[\int_0^T \int_0^1 \{ E_{jkpq} + E^+_{jkpq} + F_{jkpq} + F^+_{jkpq} \}\, dx\, dt \right.$$

$$\left. + \int_0^T \{ G_{jkpq} + G^+_{jkpq} + G'_{jkpq} + G'^+_{jkpq} \}\, dt \right] c^+_{jk}$$

$$+ \int_0^1 w_p(x)\, w_q(0)\, v_0(x)\, dx \qquad \text{(III 97)}$$

pour $p \in J_{m_1}$ et $q \in J_{m_2}$.

Définissons H_{il} par

$$H_{il} = \frac{1}{2}\int_0^T \int_0^1 [E_{jkpq} + E^t_{jkpq} + F_{jkpq} + F^t_{jkpq}]\, dx\, dt + \int_0^T [G_{jkpq} + G^t_{jkpq} + G'_{jkpq} + G'^t_{jkpq}]\, dt$$

et $b_i = \int_0^1 w_p(x)\, w_q(0)\, v_0(x)\, dx$

où $i = (p-1)m_2 + q$ et $l = (j-1)m_2 + k$ pour $j,p \in J_{m_1}$ et $k,q \in J_{m_2}$.
Résoudre (III 97) revient alors à résoudre

$$\boxed{Hy = b} \qquad \text{(III 98)}$$

où la matrice symétrique $H((m_1 m_2 n)\times(m_1 m_2 n))$ est définie par

$$H = \begin{bmatrix} H_{11} & \cdots & H_{1\,m_1 m_2} \\ \vdots & & \vdots \\ H_{m_1 m_2\,1} & \cdots & H_{m_1 m_2 m_1 m_2} \end{bmatrix} \qquad \text{(III 99)}$$

avec $H_{ij} = H^t_{ji}$, et où y est donné par

$$y = \begin{bmatrix} c_{11} \\ \vdots \\ c_{m_1 m_2} \end{bmatrix}$$

On peut, de plus, démontrer une propriété numérique supplémentaire du système algébrique (III 98)

Théorème III 25

> Si H est définie comme en (III 99), alors H est définie positive

Démonstration : Supposons $y \in \mathbb{R}^{m_1 m_2 n}$ et $y \neq 0$. Soit λ correspondant à y par construction et soient u, f et v correspondant à λ par (III 79) à (III 87). Nous avons alors

$$y^t H y = -\frac{1}{2} y^t \left\{ \int_0^T \int_\Omega [E + E^t + F + F^t] \, dx \, dl + \int_0^T \int_\Gamma [G + G^t] \, dx \, dt \right\} y$$

$$= \int_0^T \int_0^l \langle u, Ru \rangle \, dx \, dl \, dt + \int_0^T \int_0^l \langle v, gu \rangle \, dx \, dl \, dt + \int_0^T \langle f, Sf \rangle \, dt$$

$$= 2J[y, v] \geq 0$$

par construction de J. ∎

Ainsi se termine l'étude des trois problèmes annoncés.

Conclusion

Il est clair que les problèmes abordés dans ces quelques pages sont loin d'être épuisés. De nombreuses portes restent encore ouvertes pour une recherche plus approfondie.

Nous aurions personnellement beaucoup souhaité pouvoir traiter le côté plus numérique de la chose, mais les difficultés en matière d'informatique en 1973-1974 n'ont malheureusement pas permis l'implémentation directe des méthodes décrites.

Il nous reste à souhaiter que les circonstances futures permettront de revenir à ces problèmes qui, nous semble-t-il, peuvent avoir de fécondes répercussions non seulement dans l'application directe, mais aussi dans la théorie mathématique elle-même.

Bibliographie

A-1 ATHANS M. et FALB P.L.: "Optimal control: an introduction to the theory and its applications", Lincoln Laboratory Pub. (McGraw Hill), NY, 1966

B-1 BLISS G.A.: "Lectures on the calculus of variations", Phoenix Science series, Chicago, 1968

B-2 BOSARGE W.E. Jr, JOHNSON O.G., MC KNIGHT R.S. and TIMLAKES W.P.: "The Galerkin procedure for non linear control problems", SIAM J. Numer. Anal., Vol 10, No 1, March 1973, pp 94 sq

B-3 BOSARGE W.E. Jr, JOHNSON O.G. and SMITH C.L.: "A direct method: Approximation to the linear parabolic regulator problem over multivariate spline bases", SIAM J. Numer. Anal., Vol 10, No 1, March 1973, pp 35 sq

B-4 BRAUER F. and NOHEL J.: "Ordinary differential equations". W.A. Benjamin N.Y., 1966

B-5 BRYSON A.E. and HO Y.C.: "Applied optimal control", Blaisdell Pub. Company, Waltham, Massachusetts, 1969

B-6 BOURBAKI N. "Elements de mathématique" (Livre V: Espaces vectoriels topologiques), Hermann, Paris, 1955

C-1 CHENEY E.W. "Introduction to Approximation theory", Mc Graw Hill, N.Y., 1966

C-2 CODDINGTON E.A. and LEVINSON N. "Theory of Ordinary Differential equations", Mc Graw Hill, N.Y., 1955

D-1 DIEUDONNÉ J.: "Les fondements de l'analyse moderne", Gauthier-Villars, Paris, 1965

E-1 EDWARDS R.E.: "Functional analysis: theory and applications.", Holt, Rinehart and Winston Inc., N.Y., 1965

K-1 KALMAN R.E. "Contributions to the theory of optimal control". Bol. Soc. Mat. Mex., vol 5, 1960, pp 102-119

L-1 LUENBERGER D.G.: "Optimization by vector space methods.", J. Wiley, N.Y., 1969

R-1 RAVIART P.A. "Methodes des éléments finis", cours de III^e cycle à Paris VI et CNRS, Paris, 1971

S-1 SCHULTZ M.H.: "Multivariate spline functions and elliptic problems", ed. I.J. SCHOENBERG, Academic Press, N.Y., 1969

S-2 SCHULTZ M.H.: "A Ritz method for an optimal control problem", Journ. of Opt. theory and appl., vol 11, n°3, 1973

Y-1 YOSIDA K.: "Functional analysis", Springer Verlag, Berlin, 1971

Z-1 ZADEH L.A. and DESOER C.A.: "Linear System theory: the state space approach", Mc Graw Hill, N.Y., 1963