

Introduction au contrôle optimal

Aspects théoriques et numériques

Master 2 - Calcul Scientifique et Mathématiques de l'Information

Année universitaire 2022-2023

polycopié rédigé par Yannick PRIVAT

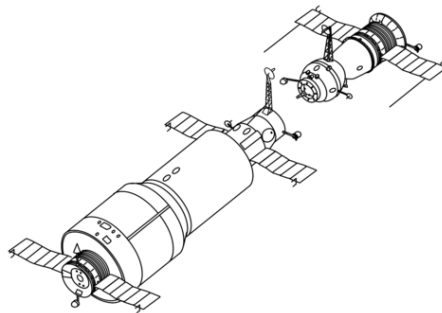


Table des matières

Préambule	1
Introduction	3
I. Contrôlabilité de systèmes différentiels ordinaires	5
I.1 Contrôlabilité des systèmes linéaires	5
<i>I.1.1 Critère de Kalman</i>	6
<i>I.1.2 Ensemble accessible</i>	7
I.2 Contrôlabilité des systèmes non-linéaires	9
<i>I.2.1 Rappels sur le théorème de Cauchy-Lipschitz</i>	9
<i>I.2.2 Contrôlabilité locale de systèmes autonomes</i>	10
I.3 Exercices du chapitre	12
II. Principe du maximum de Pontryagin	15
II.1 Commande en temps minimum de systèmes linéaires	16
II.2 Problème LQ	17
<i>II.2.1 Le principe du maximum de Pontryagin</i>	17
<i>II.2.2 Équation de Riccati et contrôle de boucle fermée</i>	19
II.3 Cas de systèmes non-linéaires	20
II.4 Exercices du chapitre	25
III. Méthodes d'adjoint pour les contraintes EDP	29
III.1 Optimisation dans des espaces de Hilbert	29
<i>III.1.1 Rappels sur les fonctions convexes</i>	29
<i>III.1.2 Théorème d'existence et inéquation d'Euler</i>	30
III.2 Contrôle optimal de problèmes elliptiques	31
<i>III.2.1 Contrôle et observation distribués</i>	31
<i>III.2.2 Algorithme de minimisation pour le problème (P)</i>	33
<i>III.2.3 Contrôle frontière et observation distribuée</i>	33
III.3 Contrôle optimal de l'équation de la chaleur	34
<i>III.3.1 Préliminaires</i>	35
<i>III.3.2 Contrôle et observation distribués</i>	36
<i>III.3.3 Algorithme de minimisation pour le problème (III.7)</i>	37
<i>III.3.4 Contrôle frontière et observation distribuée.</i>	38
III.4 Exercices du chapitre	40
Appendice	44

A. Rappels sur l'exponentielle d'endomorphismes	45
A.1 Définition	45
A.2 Méthodes pratiques de calcul d'exponentielles de matrices	45
A.3 Systèmes différentiels linéaires à coefficients constants	46
B. Rappels sur la régularité des ouverts et les équations elliptiques	49
B.1 Espaces de Sobolev	49
B.2 Régularité des ouverts	49
B.3 Intégration par parties, formule de Green	50
B.4 Théorème de Rellich et inégalité de Poincaré	50
B.5 Théorème de Lax-Milgram	51
B.6 Régularité elliptique	52
C. Optimisation dans les espaces de Hilbert, topologie faible	55
C.1 Topologie faible dans les espaces de Hilbert	55
C.2 Semi-continuité inférieure	56
C.3 Optimisation dans un espace de Hilbert	57
D. Filtre de Kalman déterministe	59
Bibliographie	63

Préambule

Ce cours présente les bases théoriques et numériques du contrôle optimal. Le premier chapitre traite de la notion de contrôlabilité (exacte, approchée). Le second chapitre est dédié à l'analyse des problèmes de contrôle optimal en dimension finie, autrement dit mettant en jeu les solutions de systèmes différentiels. L'obtention de conditions d'optimalité pour ce type de problème repose sur le Principe du Maximum de Pontryagin (PMP). Le troisième et dernier chapitre étend l'étude précédente à l'analyse de problèmes d'optimisation mettant en jeu la solution d'équations aux dérivées partielles (on s'intéressera tout particulièrement aux EDP elliptiques et paraboliques).

Quelques algorithmes numériques sont également introduits à l'occasion de séances de travaux pratiques.

Divers exercices accompagnent le présent document afin d'assimiler les notions étudiées en cours.

Ce polycopié est volontairement succinct. Sa lecture ne peut exempter d'une présence assidue en cours.

Introduction

Afin de mieux comprendre de quoi il retourne dans ce cours, débutons par quelques exemples significatifs de problèmes de contrôle optimal.

Rendez-vous de deux vaisseaux spatiaux au voisinage de la terre. Supposons que le véhicule 1 est passif et de trajectoire circulaire, et le véhicule 2 est activé par un moteur exerçant une poussée $v = (v_1, v_2)$ pour rattraper le véhicule 1. Désignons par x le vecteur linéarisé de la position du véhicule 2 dans le repère mobile d'origine le véhicule 1. Alors $x = (x_1, x_2)$ obéit aux équations de Hill

$$\begin{cases} \ddot{x}_1(t) = 3\omega^2 x_1(t) + 2\omega \dot{x}_2(t) + v_1(t) \\ \ddot{x}_2(t) = -2\omega \dot{x}_1(t) + v_2(t) \\ (x_1(0), \dot{x}_1(0), x_2(0), \dot{x}_2(0)) = X_0 \in \mathbb{R}^4 \end{cases}$$

Posons $X_v = (x_1, \dot{x}_1, x_2, \dot{x}_2)$, où les x_i sont les solutions du système ci-dessus, autrement dit les positions et vitesses relatives des vaisseaux dans le repère mobile. La pulsation ω correspond à la période de révolution $T = 2\pi/\omega$ du véhicule 1, elle est égale à 5480 secondes pour la station spatiale internationale. Modélisons le problème. Pour effectuer un rendez-vous spatial, plusieurs approches sont possibles :

- (i) *Temps minimum.* On prend en compte la puissance du moteur en introduisant l'ensemble (admissible) de contraintes sur le contrôle

$$\mathcal{U}_{\text{ad}} = \{v \in C^0 M([0, T]) \mid v(s) \in U\},$$

où U est un compact de \mathbb{R}^2 , par exemple, on suppose que les moteurs sont indépendants tels que $|v_i(s)| \leq 1$, $i = 1, 2$. On cherche un contrôle v qui ramène le vaisseau de X_0 à $X(T) = 0$ en un temps minimal T , autrement dit

minimiser T tel que $X_v(T) = 0$ parmi les contrôles $v \in \mathcal{U}_{\text{ad}}$.

- (ii) *Compromis énergie-précision.* Cette fois, on fixe un horizon de temps $T > 0$, et on souhaite que les deux vaisseaux se rencontrent à l'horizon de temps T . Une façon de procéder consiste à définir la fonctionnelle

$$J(v) = \frac{\varepsilon}{2} \int_0^T (v_1(t)^2 + v_2(t)^2) dt + (1 - \varepsilon)(x_1(T)^2 + \dot{x}_1(T)^2 + x_2(T)^2 + \dot{x}_2(T)^2)$$

où $\varepsilon \in]0, 1[$ est un poids donné. Le premier terme de la somme traduit le “coût du contrôle” tandis que le second cherche à rendre les positions et vitesses finales relatives nulles. On cherche à minimiser J sur un certain ensemble des contrôles admissibles \mathcal{U}_{ad} , i.e. on résout

minimiser $J(v)$ parmi les contrôles $v \in \mathcal{U}_{\text{ad}}$.

On remarque que dans cette approche, la cible finale $X_v(T)$ n'est pas forcément atteinte, mais se présente dans la fonction coût comme une pénalisation.

Consommation minimale d'énergie. Soit Ω un ouvert borné de \mathbb{R}^n qui représente un corps thermiquement conducteur. L'état du système considéré est le champ des températures dans Ω , c'est-à-dire une fonction $y : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ solution de l'équation de la chaleur

$$\begin{cases} \frac{\partial y}{\partial t}(t, x) - \Delta y(t, x) = f(x) + v(t, x) & (t, x) \in [0, T] \times \Omega \\ y(t, x) = 0 & (t, x) \in [0, T] \times \partial\Omega \\ y(0, \cdot) = y_0(\cdot) \end{cases}$$

D'un point de vue physique, la condition de Dirichlet homogène signifie tout simplement que la température est maintenue constante sur le bord et que cette constante est utilisée comme origine pour l'échelle de température. La fonction f modélise des sources de chaleur sur lesquelles aucune action n'est possible. En revanche la fonction contrôle v représente des sources de chaleur que l'on peut contrôler afin d'agir sur la température dans le domaine Ω . D'autre part, on se donne un profil de température z_d , et on introduit le critère quadratique

$$J(v) = \frac{1}{2} \int_0^T \int_{\Omega} |y(t, x) - z_d(x)|^2 dx dt + \frac{\alpha}{2} \int_0^T \int_{\Omega} |v(t, x)|^2 dx dt$$

où le premier terme mesure l'écart entre l'état y et le profil z_d , tandis que le second mesure d'une certaine manière la consommation d'énergie, et $\alpha > 0$ est une constante de pondération entre les deux termes. Enfin, le contrôle v sera pris dans un ensemble des contraintes (contacts unilatéraux, sources localisées...). Typiquement, on a

$$\mathcal{U}_{\text{ad}} = \{v \in L^\infty([0, T]) \mid a \leq v(s) \leq b \text{ p.p. dans }]0, T[\times \omega\},$$

où ω désigne un ouvert de Ω donné.

Le problème de contrôle optimal s'écrit : *trouver un contrôle $u \in \mathcal{U}_{\text{ad}}$ minimisant J sur l'ensemble des contraintes \mathcal{U}_{ad} .*

De cette manière, on cherche à régulariser la température avec le moindre coût de consommation d'énergie. En résumé, le problème du contrôle optimal en dimension finie ou infinie est en effet un problème d'optimisation. La nouveauté est que la fonction coût J ne s'exprime pas directement en fonction du contrôle v , mais à travers la variable d'état qui est obtenue en résolvant l'équation d'état dans laquelle intervient directement le contrôle v . Ceci posera de sérieux problèmes dans le calcul du gradient de J qui joue un rôle très important dans tous les problèmes d'optimisation. L'idée principale dans les démarches méthodiques par rapport aux problèmes d'optimisation standards consiste en l'introduction d'un état adjoint qui permet de calculer efficacement le gradient de J ainsi d'exprimer la condition d'optimalité sous une forme facile à manipuler.

Contrôlabilité de systèmes différentiels ordinaires

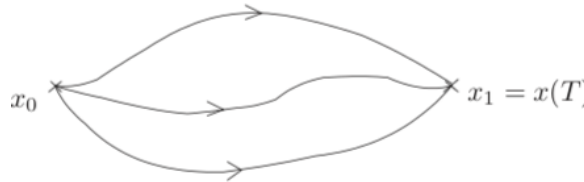
Le problème général étudié dans cette partie est le suivant. Soient $(n, m) \in (\mathbb{N}^*)^2$, I un intervalle de \mathbb{R} , et $f : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ une application choisie suffisamment régulière pour que tout ce qui suit ait un sens. Soit U un sous-ensemble de \mathbb{R}^m et soit $x_0 \in \mathbb{R}^n$. Le système de contrôle linéaire auquel on s'intéresse est

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)) & t \in I \\ x(0) = x_0, \end{cases} \quad (\text{I.1})$$

où l'ensemble des contrôles u considérés est l'ensemble des applications mesurables et bornées sur I , à valeurs dans le sous-ensemble $U \subset \mathbb{R}^m$.

Les théorèmes classiques d'existence de solutions d'équations différentielles nous assurent que, pour tout contrôle u , le système (I.1) admet une unique solution $x(\cdot) : I \rightarrow \mathbb{R}^n$, absolument continue.

La solution x_u du système ci-dessus dépend de u . Donc si on change la fonction u , on obtient une autre trajectoire $t \mapsto x(t)$ dans \mathbb{R}^n .



Deux questions se posent alors naturellement :

- Etant donné un point $x_1 \in \mathbb{R}^n$, existe-t-il un contrôle u tel que la trajectoire associée à ce contrôle joigne x_0 à x_1 en un temps fini T ? **C'est le problème de contrôlabilité exacte.**
- Si la condition précédente est remplie, existe-t-il un contrôle joignant x_0 à x_1 , et qui de plus minimise une certaine fonctionnelle $u \mapsto C(u)$?

C'est le problème de contrôle optimal.

La fonctionnelle $u \mapsto C(u)$ est un critère d'optimisation, on l'appelle le *coût*. Par exemple ce coût peut être égal au temps de parcours; dans ce cas c'est le problème du *temps minimal*.

I.1 Contrôlabilité des systèmes linéaires

I.1.1 Critère de Kalman

Soit $T > 0$ un horizon temporel fixé. On considère un système dynamique dont l'état $x(t) \in \mathbb{R}^n$ pour tout $t \in [0, T]$ est régi par le système différentiel

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) & t \in [0, T] \\ x(0) = x_0, \end{cases} \quad (\text{I.2})$$

où $A \in \mathbb{M}_n(\mathbb{R})$, $B \in \mathbb{M}_{n,m}(\mathbb{R})$ et $u(\cdot) : [0, T] \rightarrow \mathbb{R}^m$ est une fonction de **contrôle**. Puisque la trajectoire x , solution de (I.2), dépend du contrôle u , nous la noterons souvent x_u . Il n'est pas toujours pertinent de chercher un contrôle u continu, on ne cherchera pas une trajectoire de classe $C^1([0, T]; \mathbb{R}^n)$. Un bon cadre fonctionnel pour la trajectoire est celui des fonctions absolument continues¹ sur $[0, T]$.

Proposition I.1.1. Formule de Duhamel

Pour tout contrôle $u \in L^1([0, T]; \mathbb{R}^m)$, il existe une unique trajectoire $x_u \in AC([0, T], \mathbb{R}^n)$ solution de (I.2) au sens où cette trajectoire vérifie la condition initiale $x_u(0) = 0$ et le système différentiel de (I.2) p.p. sur $[0, T]$. Cette trajectoire est donnée par la formule de Duhamel

$$x_u(t) = e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu(s)ds, \quad t \in [0, T].$$

Intéressons-nous à la notion de contrôlabilité exacte que nous précisons ci-dessous.

Définition I.1.2. Contrôlabilité exacte

On dit que le système (I.2) est contrôlable en temps T à partir de x_0 si

$$\forall x_1 \in \mathbb{R}^n, \exists u \in L^1(0, T; \mathbb{R}^m) \mid x_u(T) = x_1.$$

Un résultat remarquable, dû à Kalman, permet de caractériser la contrôlabilité du système différentiel ci-dessus à partir d'une condition purement algébrique ne faisant intervenir que les matrices A et B . On introduit la matrice de Kalman

$$\text{Kal} = [B, AB, \dots, A^{n-1}B] \in \mathbb{M}_{n,nm}(\mathbb{R}).$$

Théorème I.1.3. Critère de Kalman

Le système linéaire autonome (I.2) est (exactement) contrôlable pour tout $T > 0$ et pour tout $x_0 \in \mathbb{R}^n$ si et seulement si la matrice Kal est de rang maximal, i.e. $\text{rg}(\text{Kal}) = n$.

1. On dit qu'une fonction $F : [0, T] \rightarrow \mathbb{R}^n$ est absolument continue sur $[0, T]$ et on écrit $F \in AC([0, T]; \mathbb{R}^n)$ s'il existe $f \in L^1([0, T]; \mathbb{R}^n)$ telle que

$$F(t) - F(0) = \int_0^t f(t)dt \quad \forall t \in [0, T].$$

Si une fonction F est absolument continue sur $[0, T]$, alors elle est continue sur $[0, T]$ et elle est dérivable presque partout, de dérivée égale à f .

Remarque I.1.4

La condition de Kalman est indépendante de l'horizon temporel $T > 0$ et de la donnée initiale x_0 , ce qui implique en particulier que lorsqu'un système de contrôle linéaire autonome est contrôlable, on peut atteindre à partir d'une donnée initiale toute cible en un horizon temporel même très court. Attention toutefois : le prix à payer est que le contrôle peut prendre des valeurs arbitrairement grandes.

Exemple I.1.5

Pour quelles valeurs de α le système

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 2 & \alpha - 3 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ \alpha^2 - \alpha & 0 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

est-il contrôlable ?

Remarque I.1.6 Cas des systèmes linéaires non autonomes

Soient A , B et r trois applications L^∞ sur $[0, T]$ à valeurs respectivement dans $\mathbb{M}_n(\mathbb{R})$, $\mathbb{M}_{n,m}(\mathbb{R})$, et $\mathbb{M}_{n,1}(\mathbb{R})$. Soit U un sous-ensemble de \mathbb{R}^m et soit $x_0 \in \mathbb{R}^n$. Considérons le système contrôlé

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) & t \in [0, T] \\ x(0) = x_0 \in \mathbb{R}^n \end{cases} \quad (\text{I.3})$$

Pour de tels systèmes, la formule de Duhamel n'est plus valable et on utilise la notion de résolvante² que l'on notera R . On vérifie sans peine que la solution du système (I.3) est x_u définie par

$$x_u(t) = R(t)x_0 + R(t) \int_0^t R(s)^{-1} B(s) u(s) ds.$$

En adaptant soigneusement la preuve du critère de Kalman, on peut montrer que le système (I.3) est contrôlable en temps T si, et seulement si la matrice de contrôlabilité

$$G_T = \int_0^T R(s)^{-1} B(s) B(s)^\top (R(s)^{-1})^\top ds \in \mathbb{M}_n(\mathbb{R})$$

est inversible.

I.1.2 Ensemble accessible

Dans cette section, on considère le système (linéaire) contrôlé

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) & t \in [0, T] \\ x(0) = x_0 \in \mathbb{R}^n \end{cases} \quad (\text{I.4})$$

2. La résolvante R du système (I.3) est l'application $R : [0, T] \rightarrow \mathbb{M}_n(\mathbb{R})$ telle que

$$\begin{cases} \dot{R}(t) = A(t)R(t) & t \in [0, T] \\ R(0) = I_n \end{cases}$$

Il est important de noter que si $A \in L^1([0, T], \mathbb{M}_n(\mathbb{R}))$, alors R est absolument continue sur $[0, T]$. De plus, on peut montrer que $\frac{d}{dt} \det(R(t)) = \text{tr } A(t) \det(R(t))$, ce qui implique que $R(t)$ est inversible pour tout t puisque $\det R(0) = 1$.

où A et B sont deux applications L^∞ sur $[0, T]$ à valeurs respectivement dans $\mathbb{M}_n(\mathbb{R})$ et $\mathbb{M}_{n,m}(\mathbb{R})$. On suppose que le contrôle u est à valeurs dans un sous-ensemble non vide $U \subset \mathbb{R}^n$.

Définition I.1.7. Ensemble des points accessibles

L'ensemble des points accessibles à partir de x_0 en un temps $T > 0$ est défini par

$$\mathcal{A}(x_0, T) = \{x_u(T) \mid u \in L^\infty([0, T], U)\},$$

où $x_u(\cdot)$ est la solution du système (I.4) associée au contrôle u .

Intéressons-nous à présent à la topologie des ensembles accessibles

Théorème I.1.8.

On suppose que U est un convexe compact non vide de \mathbb{R}^m . Considérons le système de contrôle linéaire dans \mathbb{R}^n

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

Soient $T > 0$ et $x_0 \in \mathbb{R}^n$. Alors pour tout $t \in [0, T]$, $\mathcal{A}(x_0, t)$ est compact, convexe, et varie continûment avec t sur $[0, T]$.

Remarque I.1.9

De façon surprenante, la conclusion de ce théorème reste vraie si l'on remplace l'hypothèse “ U est convexe” par “ U est compact”. Dans ce cas, on peut même démontrer que les ensembles accessibles associés à U , sa frontière et son enveloppe convexe sont les mêmes.

Pour caractériser la variation continue d'un ensemble, nous allons utiliser la notion de *distance de Hausdorff* entre deux compacts que nous précisons : si $x \in \mathbb{R}^n$ et E est un sous-ensemble fermé, on notera $d_E(x)$ la distance de x à E , soit $d_E(x) = \text{dist}(x, E) := \inf\{|x - y|, y \in E\}$.

Définition I.1.10. Distance de Hausdorff entre deux fermés

Soient A et B , deux ensembles fermés de \mathbb{R}^d . La distance de Hausdorff entre A et B est définie par

$$d_H(A, B) = \max\{\sup_{x \in A} d_B(x), \sup_{x \in B} d_A(x)\}.$$

Exemple I.1.11

On considère un point matériel en mouvement rectiligne. On contrôle la vitesse de ce point par un contrôle à valeurs dans l'intervalle borné $U = [-1, 1]$, ce qui conduit à considérer le système

$$\dot{x}(t) = u(t), \quad t \in [0, T], \quad x(0) = 0, \quad u(t) \in U.$$

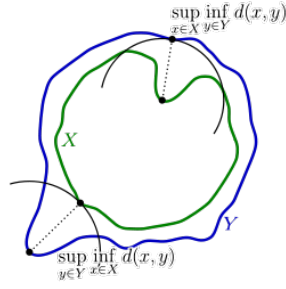


FIGURE I.1 : Distance de Hausdorff de deux compacts X et Y

où on a fixé l'origine à la position initiale du point matériel. L'ensemble atteignable est $\mathcal{A}(t, 0) = [-t, t]$ (qui est bien compact, convexe et varie continûment en t). Quel est l'ensemble des points atteignables ?

I.2 Contrôlabilité des systèmes non-linéaires

Considérons un système de la forme

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)) & t \in [0, T] \\ x(0) = x_0, \end{cases} \quad (\text{I.5})$$

où $u \in L^1([0, T], \mathbb{R}^m)$ et $f : [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$. On voit donc que même si l'application f est régulière par rapport à u , le fait que le contrôle ne dépende pas continûment du temps fait que l'application f ne sera pas nécessairement continue en t .

I.2.1 Rappels sur le théorème de Cauchy-Lipschitz

Théorème I.2.1. Cauchy-Lipschitz, cas mesurable et Lipschitz global

On suppose que :

- (i) L'application F est mesurable en t et continue en x , i.e., pour tout $x \in \mathbb{R}^n$, l'application $t \mapsto F(t, x)$ est mesurable et pour presque tout $t \in [0, T]$, l'application $x \mapsto F(t, x)$ est continue ;
- (ii) L'application F est intégrable en t , i.e.

$$\forall x \in \mathbb{R}^n, \quad \exists \beta \in L^1([0, T], \mathbb{R}_+) \quad | | \quad |F(t, x)| \leq \beta(t), \quad \forall t \in [0, T];$$

- (iii) L'application F est globalement lipschitzienne en x , i.e.

$$\exists C_0 \in L^1([0, T], \mathbb{R}_+) \quad | | \quad |F(t, x_1) - F(t, x_2)|_{\mathbb{R}^n} \leq C_0(t) |x_1 - x_2|_{\mathbb{R}^n}$$

pour presque tout $t \in [0, T]$ et tous $x_1, x_2 \in \mathbb{R}^n$.

Alors, il existe une unique solution au problème de Cauchy $\dot{x}(t) = F(t, x(t))$, $t \in [0, T]$ et $x(0) = x_0$ telle que $x \in AC([0, T], \mathbb{R}^n)$, satisfaisant

$$\forall t \in [0, T], \quad x(t) = x(0) + \int_0^t F(s, x(s)) ds.$$

Cette solution est définie sur l'intervalle $J \subset [0, T]$ et soit $J = [0, T]$, soit $J = [0, T^*[$ avec $T^* < T$ et $\lim_{t \rightarrow T^*} |x(t)|_{\mathbb{R}^n} = +\infty$.

Remarque I.2.2 Généralisation

Si l'on remplace l'hypothèse (iii) par : $\forall x \in \mathbb{R}^n, \quad \exists r > 0, \quad \exists C_0 \in L^1([0, T], \mathbb{R}_+)$ tels que p.p. $t \in [0, T], \quad \forall x_1, x_2 \in B(x, r), \quad |F(t, x_1) - F(t, x_2)|_{\mathbb{R}^n} \leq C_0(t) |x_1 - x_2|_{\mathbb{R}^n}$, alors il existe une unique solution maximale au problème de Cauchy précédent.

Exemple I.2.3 Explosion pour un système de contrôle

On se place dans \mathbb{R} ($n = 1$) et on considère le système de contrôle

$$\dot{x}(t) = f(t, x(t), u(t)), \quad t \in [0, T], \quad x(0) = 0, \quad u(t) \in U.$$

avec un contrôle à valeurs scalaires ($m = 1$) et l'application f telle que $f(t, x, u) = x^2 + u$ (qui ne dépend pas de t explicitement). On obtient alors le problème de Cauchy $\dot{x}(t) = x(t)^2 + u(t)$. On considère la donnée initiale $x_0 = 0$ et on suppose que le contrôle est constant en temps égal à $u_0 \in \mathbb{R}_+$. On vérifie sans peine que la trajectoire est donnée par $x(t) = \sqrt{u_0} \tan(\sqrt{u_0} t)$. On a donc explosion au temps fini $T^* = \pi/(2\sqrt{u_0})$ qui dépend de la valeur (constante) prise par le contrôle

I.2.2 Contrôlabilité locale de systèmes autonomes

On fixe l'horizon temporel $T > 0$ et la condition initiale $x_0 \in \mathbb{R}^n$, et on considère le système de contrôle non-linéaire

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) & t \in [0, T] \\ x(0) = x_0, \end{cases} \quad (\text{I.6})$$

où $u \in L^1([0, T], \mathbb{R}^m)$ et $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$.

Dans cette section, on suppose que la fonction f est de classe C^1 en (x, u) .

Soit $y \in \mathcal{A}(T, x_0)$. Par définition, il existe un contrôle $u_y \in \mathcal{U}_{T, x_0}$ amenant l'état de x_0 à y en temps T . Le problème de la contrôlabilité locale consiste à savoir si cette propriété reste satisfaite dans un voisinage du point $y \in \mathcal{A}(T, x_0)$.

On considère le cas particulier d'un **point d'équilibre** d'un système différentiel autonome, i.e., un couple (\bar{x}, \bar{u}) tel que $f(\bar{x}, \bar{u}) = 0$. Noter que $\bar{x} \in \mathcal{A}(t, \bar{x})$ en utilisant le contrôle constant égal à \bar{u} .

Définition I.2.4. Contrôlabilité locale

On dit que le système de contrôle non-linéaire (I.6) est contrôlable localement au point d'équilibre (\bar{x}, \bar{u}) s'il existe $\varepsilon > 0$ tel que pour tout point $x_1 \in \mathbb{R}^n$ vérifiant $\|x_1 - \bar{x}\| < \varepsilon$, il existe une commande u dans $L^\infty(0, T; \mathbb{R}^m)$ telle que $x_u(T) = x_1$, où x_u désigne la solution de (I.6) associée au choix d'état initial x_0 et à la commande u .

Pour énoncer le résultat de contrôlabilité locale, il est utile d'introduire la notion de système linéarisé autour de l'équilibre.

Définition I.2.5. Système linéarisé

Soit $A = \frac{\partial f}{\partial x}(x_0, u_0) \in \mathbb{M}_n(\mathbb{R})$ (matrice jacobienne de f par rapport à la variable x évaluée en (x_0, u_0)) et $B = \frac{\partial f}{\partial u}(x_0, u_0) \in \mathbb{M}_{n, m}(\mathbb{R})$ (matrice jacobienne de f par rapport à la variable u évaluée en (x_0, u_0)). Le système différentiel linéarisé autour de l'équilibre (x_0, u_0) est le système

$$\begin{cases} \frac{d(\delta x)}{dt} = A\delta x(t) + B\delta u(t) & t \in [0, T] \\ \delta x(0) = 0_{\mathbb{R}^n} \end{cases} \quad (\text{I.7})$$

où $\delta u \in L^\infty(0, T, \mathbb{R}^m)$ est donné.

Théorème I.2.6. Contrôlabilité locale

On suppose qu'aucune contrainte n'est imposée sur le contrôle. Si le système différentiel linéarisé est contrôlable en temps T (autrement dit s'il satisfait la condition de Kalman), alors le système différentiel non-linéaire est localement contrôlable en temps T à partir de x_0 .

Exemple I.2.7 Le pendule inversé

On considère l'exemple du pendule inversé (masse vers le haut, tige vers le bas) avec pour simplifier une masse et une longueur unités. On suppose que le pendule a un mouvement dans un plan et on repère l'extrémité supérieure du pendule par son angle θ avec la verticale (dans le sens horaire). On contrôle l'accélération horizontale du point inférieur de la tige. La dynamique s'écrit sous la forme

$$\ddot{\theta}(t) = \sin \theta(t) - u(t) \cos \theta(t).$$

Posons $x = (x_1, x_2) = (\theta, \dot{\theta})$, on se ramène au système d'ordre un (I.6) avec

$$f(x, u) = \begin{pmatrix} x_2 \\ \sin x_1 - u \cos x_1 \end{pmatrix}.$$

Le pendule inversé est-il **localement contrôlable** autour du point d'équilibre instable $(x_0, u_0) = ((0, 0)^\top, 0)$?

I.3 Exercices du chapitre

Rappels d'analyse et premiers pas en contrôle optimal

Les deux exercices qui suivent n'utilisent pas les notions introduites dans ce chapitre. Il servent de support à la révision de quelques notions étudiées les années précédentes et utiles pour ce cours.

I.1 Pour $x_0 \in \mathbb{R}$ fixé, on considère la solution x_u du système du premier ordre

$$\begin{cases} x'(t) = a(t)x(t) + u(t) & t \in]0, 1[\\ x(0) = x_0, \end{cases}$$

où a et u sont deux fonctions de $C^0([0, 1])$.

- (i) Décrire géométriquement l'ensemble des fonctions $u(\cdot)$ telles que $x_u(1) = 0$.
- (ii) Soit J , la fonctionnelle définie sur $C^0([0, 1])$ par $J(u) = \frac{1}{2} \int_0^1 u(t)^2 dt$ et soient u et h , deux éléments de $C^0([0, 1])$. Déterminer (si elles existent) les différentielles directionnelles au point u dans la direction h de J et $u \mapsto x_u(1)$.
- (iii) Résoudre le problème $\inf\{J(u), u \in C^0([0, 1]), x_u(1) = 0\}$.
Indication : on cherchera de prime abord à résoudre le problème $\inf\{J(u), u \in L^2([0, 1]), x_u(1) = 0\}$. On n'oubliera pas d'étudier l'existence et l'unicité de solutions.

I.2 Soit Ω , un ouvert connexe borné et régulier de \mathbb{R}^N . On considère y_Ω , la solution de l'équation aux dérivées partielles

$$\begin{cases} -\Delta y_\Omega(x) = u(x)f(x) & x \in \Omega \\ y_\Omega(x) = 0 & x \in \partial\Omega, \end{cases}$$

où $f \in C^0(\overline{\Omega})$ est une fonction strictement positive donnée et $u(\cdot) \in L^2(\Omega)$ est une loi de commande. On s'intéresse au problème

$$\sup_{u \in L^2(\Omega)} J(u) \quad \text{avec} \quad J(u) = \int_{\Omega} u(x)f(x)y_\Omega(x) dx \quad (\mathcal{P})$$

- (i) Rappeler pourquoi l'équation aux dérivées partielles ci-dessus possède une solution unique et préciser sa régularité.
- (ii) Montrer que le problème (\mathcal{P}) est équivalent au problème de minimisation

$$\inf_{u \in L^2(\Omega)} \inf_{y \in H_0^1(\Omega)} \hat{J}(u, y), \quad (\mathcal{P}')$$

où \hat{J} est une fonctionnelle définie sur $L^2(\Omega) \times H_0^1(\Omega)$ que l'on précisera.

- (iii) On suppose que le problème (\mathcal{P}') possède une solution (u^*, y^*) . Écrire les conditions d'optimalité du premier ordre sur u^* et sur y^* .

(iv) Que vient-on de démontrer?

Contrôle exact de systèmes linéaires

I.3 (*critère de Kalman*) Les systèmes suivants sont-ils contrôlables?

(i) **Vitesse angulaire d'un rotor.** $I\dot{\omega}(t) = u(t)$ (avec $I > 0$).

(ii) **Circuit RLC.** Soient $L > 0$ et $C > 0$. En appelant $i(t)$ l'intensité du courant et $q(t)$ la charge du condensateur, on a

$$\frac{dq}{dt} = i \quad \text{et} \quad L \frac{di}{dt} + Ri + \frac{q}{C} = u.$$

(iii) **Système de ressorts couplés (train à deux wagons).** Soient $k_1 > 0$ et $k_2 > 0$.

$$\begin{cases} \ddot{x} = -k_1 x + k_2 (y - x), \\ \ddot{y} = -k_2 (y - x) + u. \end{cases}$$

I.4 Soit le système

$$\begin{cases} \dot{x}(t) = -y(t) + u(t) \cos t, \\ \dot{y}(t) = x(t) + u(t) \sin t. \end{cases}$$

(i) Calculer le rang de la matrice de Kalman (dépendant de t) associée à ce système linéaire. Que peut-on en déduire?

(ii) Montrer par deux méthodes différentes que ce système est non contrôlable en temps $T > 0$.

Indication : que peut-on dire de la fonction $t \mapsto x(t) \sin t - y(t) \cos t$?

I.5 (*contrôles bang-bang*) Soit $T > 0$, un horizon de temps. Pour chacun des deux exemples ci-dessous, répondre aux questions suivantes :

(i) Le système est-il contrôlable en temps T ?

(ii) Le système reste-t-il contrôlable si l'on se restreint à des contrôles u constants par morceaux, tels que $u(t) \in \{-1, 1\}$ pour tout $t \in [0, T]$?

Position d'un tram. Sur une voie rectiligne, elle est repérée par sa position $x(t)$ et son accélération est commandée par la relation $\ddot{x} = u$.

Oscillateur harmonique linéaire. Le déplacement est régi par l'équation $m\ddot{x}(t) + kx(t) = u(t)$.

Indication : dessiner l'ensemble des trajectoires dans le plan de phase (x, \dot{x}) .

I.6 (*contrôle de norme L^2 minimale*) Soit $T > 0$. On considère le système commandé $x'_u(t) = Ax_u(t) + Bu(t)$, $t \in [0, T]$ où $A \in \mathbb{M}_n(\mathbb{R})$ et $B \in \mathbb{M}_{n,m}(\mathbb{R})$ et la commande $u(\cdot)$ est à valeurs dans \mathbb{R}^m . Soit G , la matrice de $\mathbb{M}_n(\mathbb{R})$ définie par

$$G = \int_0^T e^{(T-s)A} B B^\top (e^{(T-s)A})^\top ds.$$

(i) On suppose que la condition de Kalman est satisfaite. Montrer que G est inversible.

(ii) Soit \bar{u} , la fonction définie sur $[0, T]$ par $\bar{u}(s) = (e^{(T-s)A} B)^\top G^{-1} v$ avec $v \in \mathbb{R}^n$. Montrer que \bar{u} envoie $x_{\bar{u}}(0) = 0$ au temps $t = 0$ sur l'état $x_{\bar{u}}(T) = v$ au temps $t = T$.

- (iii) Montrer que \bar{u} est un minimiseur du critère J défini par $J(u) = \frac{1}{2} \int_0^T \|u(s)\|^2 ds$, parmi toutes les commandes $u \in C^0([0, T], \mathbb{R}^m)$ permettant d'amener le système de 0 en $t = 0$ à v en $t = T$.

Indication : Calculer $J(\bar{u}+h) - J(\bar{u})$ où h est choisi de sorte que $x_{\bar{u}+h}(0) = 0$ et $x_{\bar{u}+h}(T) = v$.

Contrôlabilité de systèmes non-linéaires

I.7 On considère le système $\dot{x} = u^3$.

- (i) Montrer que $(0,0)$ est un point d'équilibre et étudier la contrôlabilité du système linéarisé associé. Que peut-on en déduire pour le système non-linéaire?
- (ii) Étudier la contrôlabilité locale du système non-linéaire autour de l'équilibre.

I.8 (*système proie-prédateur*) Soient α et β , deux réels strictement positifs. Considérons le système proies-prédateurs de Lotka-Volterra, donné par

$$\begin{cases} \dot{x}_1 = (1 - \alpha x_2)x_1 \\ \dot{x}_2 = (-1 + \beta x_1)x_2 + u \\ x_1(0) = x_1^0, \quad x_2(0) = 0. \end{cases}$$

où $x_1(t)$ désigne une proportion de proies dans un enclos à l'instant t , $x_2(t)$ une proportion de prédateurs dans ce même enclos et on relâche des prédateurs pour tenter de contrôler la population de proies.

- (i) Que se passe-t-il en l'absence de contrôle (i.e. $u = 0$)?
- (ii) Étudier la contrôlabilité du système linéarisé autour de chaque état d'équilibre du système sans contrôle et interpréter le résultat.

Principe du maximum de Pontryagin

Le principe du maximum de Pontryagin permet d'écrire de façon pratique les conditions d'optimalités associées à un problème de contrôle optimal faisant intervenir la solution d'une équation différentielle contrôlée. Si l'on cherche à les écrire naïvement sans recourir à cet outil, on aboutit en général à une formulation très difficilement exploitable, même dans des cas très simples. Considérons par exemple l'équation différentielle (linéaire) contrôlée

$$\begin{cases} -\ddot{x}(t) = u(t) & t \in]0,1[\\ x \in H_0^1(0,1) \end{cases}$$

ainsi que le problème de contrôle optimal

$$\inf_{v \in L^2(0,1)} J(v), \quad \text{avec } J(v) = \frac{1}{2} \int_0^1 (x_v(s) - 1)^2 ds.$$

où x_v désigne la solution du système différentiel ci-dessus associée à la commande v (l'existence s'obtient aisément à l'aide du théorème de Riesz par exemple). Il s'agit d'un problème de type moindre carrés et la fonction de contrôle est cherchée de sorte que la solution x_v soit la plus proche possible (au sens des moindres carrés) de la fonction constante égale à 1 au cours du temps. Écrivons formellement les conditions d'optimalité au premier ordre. Pour cela, admettons l'existence d'une solution u et considérons la commande $u + \varepsilon \delta u$, où $\delta u \in L^2(0,1)$. Alors,

$$\begin{aligned} \frac{J(u + \varepsilon \delta u) - J(u)}{\varepsilon} &= \frac{1}{2} \int_0^1 ((x_{u+\varepsilon \delta u}(s) - 1)^2 - (x_u(s) - 1)^2) ds \\ &= \int_0^1 x_{\delta u}(s) \left(x_u(s) - 1 + \frac{\varepsilon}{2} x_{\delta u}(s) - 1 \right) ds \end{aligned}$$

si bien qu'en faisant tendre ε vers 0, la différentielle de J (au sens de Gâteaux) en u dans la direction δu s'écrit

$$\langle dJ(u), \delta u \rangle = \int_0^1 (x_u(s) - 1) x_{\delta u}(s) ds,$$

où $x_{\delta u}(s)$ résout le problème

$$\begin{cases} -\ddot{x}_{\delta u}(t) = \delta u(t) & t \in]0,1[\\ x_{\delta u} \in H_0^1(0,1) \end{cases}$$

Les raisonnements habituels en optimisation (consistant à considérer une perturbation δu , puis $-\delta u$) conduisent à

$$\forall \delta u \in L^2(0,1), \langle dJ(u), \delta u \rangle = 0 \iff \forall \delta u \in L^2(0,1), \int_0^1 (x_u(s) - 1) x_{\delta u}(s) ds = 0.$$

Remarquons que cette dernière égalité est difficilement exploitable. Il serait beaucoup plus commode d'obtenir une condition d'optimalité de la forme

$$\forall \delta u \in L^2(0,1), \int_0^1 \delta u(s) F(s, u(s)) ds = 0,$$

où F est une fonction donnée, ce qui impliquerait que $F(s, u(s)) = 0$ p.p. dans $]0,1[$. Le principe du maximum de Pontryagin sert précisément à cela.

Débutons par un résultat fort utile dans le cas mono-entrée ($m = 1$), que nous admettrons (voir par exemple [3]).

Théorème II.0.1. Contrôlabilité sous contrainte

Soit $b \in \mathbb{R}^n$ et $U \subset \mathbb{R}$ un intervalle tel que $0 \in \text{int } U$. Considérons le système $\dot{x}(t) = Ax(t) + bu(t)$, avec $u(t) \in U$. Alors tout point de \mathbb{R}^n peut être conduit à l'origine en temps fini si et seulement si la paire (A, b) vérifie la condition de Kalman et la partie réelle de chaque valeur propre de A est inférieure ou égale à 0.

II.1 Commande en temps minimum de systèmes linéaires

Considérons le système linéaire

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) & t \in [0, T] \\ x(0) = x_0 \in \mathbb{R}^n \end{cases} \quad (\text{II.1})$$

où $A \in \mathbb{M}_n$, $B \in \mathbb{M}_{n,m}$ sont deux matrices et la commande $u \in L^\infty(0, T, U)$, où U est un sous-ensemble mesurable de \mathbb{R}^m . Nous cherchons à atteindre l'état final $x(T) = x_1$ en un temps minimum.

Intéressons-nous de prime abord à l'existence de contrôle réalisant cette propriété.

Théorème II.1.1. Existence de commandes temps optimales

On suppose que U est compact. Si le point x_1 est accessible depuis x_0 avec un contrôle u à valeurs dans U , alors il existe une trajectoire temps-minimale reliant x_0 à x_1 . De plus, x_1 est nécessairement extrémal, autrement dit $x_1 \in \partial \mathcal{A}(x_0, t^*)$.

Le théorème suivant caractérise la trajectoire et la commande optimale.

Théorème II.1.2. Condition d'optimalité et principe du maximum de Pontryagin

Soit $u \in L^\infty(0, T, U)$ une commande qui transfère le système (II.1) de $x(0) = x_0$ à $x(T) = x_T \in \mathbb{R}^n$. Si le temps T est minimum, alors il existe une fonction p non identiquement nulle solution de l'équation adjointe

$$\dot{p}(t) = -A^\top p(t), \quad t \in [0, T] \quad (\text{II.2})$$

telle que pour presque tous $s \in [0, T]$, $u(s)$ réalise instantanément le maximum de l'Hamiltonien $\mathcal{H} : U \ni v \mapsto \langle p(s), Bv \rangle_{\mathbb{R}^n}$.

Remarque II.1.3 condition de transversalité

Si u est un contrôle temps-minimal joignant en temps T une cible M_1 , où $M_1 \subset \mathbb{R}^n$ est convexe, alors on peut de plus choisir le vecteur adjoint pour que le vecteur $p(T)$ soit unitaire et normal à un hyperplan séparant (au sens large) $\mathcal{A}(x_0, T)$ et M_1 . C'est une condition dite de *transversalité*, obtenue facilement dans la preuve précédente.

Remarque II.1.4 contrôle bang-bang

Comme nous le verrons dans un exercice, une particularité du problème en temps minimum est que la commande optimale se trouve nécessairement sur le bord des contraintes U . Lorsque U est un intervalle de \mathbb{R} , la commande saute d'une extrémité à l'autre à des instants de commutations précis. On parlera de commande *bang-bang* pour illustrer le fait qu'on n'utilise que les valeurs maximales et minimales admissibles de la commande.

Exemple II.1.5 contrôle optimal d'un tram

Considérons le problème de contrôle optimal

$$\begin{cases} \dot{x}(t) = y(t) & t \in [0, T] \\ \dot{y}(t) = v(t) & t \in [0, T] \\ |v(t)| \leq 1 \end{cases} \quad (\text{II.3})$$

Supposons que la force est contrainte par $|v| \leq 1$. Ici l'objectif est de conduire le train le plus rapidement possible à la station d'arrivée avec une vitesse d'arrivée nulle (pour éviter un crash). On cherche donc une commande $u(\cdot)$ qui ramène le système ci-dessus de l'état initial (x_0, y_0) à l'état final 0 en un temps minimum T .

II.2 Problème LQ

II.2.1 Le principe du maximum de Pontryagin

Considérons un système dynamique linéaire autonome

$$\begin{cases} \dot{x}(t) = Ax(t) + Bv(t) & t \in [0, T], \\ x(0) = x_0 \in \mathbb{R}^n \end{cases} \quad (\text{II.4})$$

associé à la fonction coût quadratique J définie par

$$J(v) = \frac{1}{2} \int_0^T [(Qx(s), x(s)) + (Rv(s), v(s))] ds + \frac{1}{2} (Dx(T), x(T))$$

où R est une matrice de $\mathbb{M}_m(\mathbb{R})$ définie positive, $Q \in \mathbb{M}_n(\mathbb{R})$ et $D \in \mathbb{M}_n(\mathbb{R})$ sont supposées semi-définies positives.

Le problème du contrôle optimal LQ s'écrit :

trouver une commande $u \in L^2(0, T; \mathbb{R}^n)$ telle que la trajectoire $x(\cdot)$ associée au système (II.19) minimise J .

Remarquons que dans le modèle considéré, il n'y a pas de contrainte sur les commandes, i.e. $U = \mathbb{R}^m$, et il n'y a pas non plus de contrainte sur l'état final $x(T)$ qui se présente dans la fonction du coût J comme une pénalisation.

Théorème II.2.1. Existence et unicité

Le problème

$$\inf_{v \in L^2(0, T; \mathbb{R}^m)} J(v) \quad (\text{II.5})$$

a une solution unique.

Remarque II.2.2 extension du résultat précédent

Le théorème II.2.1 reste valable si $T = \infty$ avec $Q = 0$, pourvu que le système (II.19) soit contrôlable en un certain temps T fini. En effet, s'il existe une commande v sur $[0, T]$ qui ramène x_0 à 0, alors on peut prolonger la commande $v(t)$ par zéro pour $t \geq T$ de telle sorte que la trajectoire $x(t)$ est réduite à $\{0\}$ pour $t \geq T$. On a ainsi trouvé une trajectoire du coût fini et la borne inférieure de J est également finie. On peut alors effectuer le même procédé de minimisation comme dans le cas " $T < \infty$ " (Exercice).

Théorème II.2.3. Principe du maximum de Pontryagin

La trajectoire x , associée à la commande u , est optimale si et seulement s'il existe une variable d'état adjoint p définie par

$$\begin{cases} \dot{p}(t) + A^\top p(t) = -Qx(t) & t \in [0, T] \\ p(T) = Dx(T). \end{cases} \quad (\text{II.6})$$

De plus la commande optimale u est caractérisée par

$$u(t) = -R^{-1}B^\top p(t) \quad (\text{II.7})$$

Remarque II.2.4 le PMP en pratique

(i) Définissons le Hamiltonien \mathcal{H} par

$$H(x, v, p) = \frac{1}{2}[(Qx, x) + (Rv, v)] + (p, Ax + Bv).$$

Alors le vecteur $v(t)$ est solution du problème

$$\frac{\partial}{\partial v} H(x(t), v, p(t)) = 0.$$

Ainsi pour tout $0 \leq t \leq T$ fixé, la fonction $v \mapsto H(x(t), v, p(t))$ atteint son minimum en $u = -R^{-1}B^\top p(t)$. Autrement dit, la commande optimale u réalise le minimum instantané de cette fonction.

(ii) En reportant l'expression de u dans l'équation sur l'état, on obtient le système de $2n$ variables

$$\begin{cases} \dot{x}(t) = Ax(t) - BR^{-1}B^\top p(t) & t \in [0, T] \\ x(0) = x_0 \\ \dot{p}(t) = -A^\top p(t) - Qx(t) & t \in [0, T] \\ p(T) = Dx(T) \end{cases} \quad (\text{II.8})$$

qui caractérise parfaitement la trajectoire optimale. Attention, il ne s'agit pas d'un problème de Cauchy standard, puisque qu'il fait intervenir des conditions initiale et finale. Néanmoins, on va voir plus loin que la commande optimale u peut s'exprimer sous forme de boucle fermée grâce à l'équation de Riccati.

(iii) D'une manière plus commode, le système ci-dessus s'écrit

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p} = Ax + Bu & (\text{équation d'état}), \\ \dot{p} = -\frac{\partial H}{\partial x} = -(A^\top p + Qx) & (\text{équation adjointe}), \\ \frac{\partial H}{\partial u} = B^\top p + Ru = 0 & (\text{condition d'optimalité}) \end{cases}$$

(iv) Pour la commande et la trajectoire optimales, on peut montrer que $J(u) = \frac{1}{2}(p(0), x(0))$.

Exemple II.2.5

Résoudre le problème

$$\min_u \frac{1}{2} \int_0^T (x(t)^2 + u^2(t)) dt$$

sous la contrainte dynamique $\dot{x}(t) = u(t)$ et $x(0) = 1$.

II.2.2 Équation de Riccati et contrôle de boucle fermée

Théorème II.2.6. Équation de Riccati

Il existe une unique matrice symétrique $E \in C^1([0, T])$ de taille n telle que la trajectoire $x(\cdot)$ et l'état adjoint $p(\cdot)$ sont liés par la relation suivante

$$p(t) = E(t)x(t). \quad (\text{II.9})$$

La matrice E est solution de l'équation matricielle de Riccati

$$\begin{cases} \dot{E}(t) = -Q - A^\top E(t) - E(t)A + E(t)BR^{-1}B^\top E(t) & t \in [0, T] \\ E(T) = D. \end{cases}$$

Remarque II.2.7 Commande feedback

En utilisant (II.9), on obtient $u(t) = -R^{-1}B^\top p(t) = -R^{-1}B^\top E(t)x(t)$. La commande optimale u se met donc sous forme de boucle fermée. Ainsi la commande peut être calculée via la sortie $x(t)$ à tout instant t et le système dynamique devient

$$\begin{cases} \dot{x}(t) = Ax(t) - BR^{-1}B^\top E(t)x(t) & t \in [0, T] \\ x(0) = x_0. \end{cases}$$

Ce genre de commandes appelées *feedbacks* sont largement utilisées dans les problèmes de stabilisation.

Mentionnons une application fort intéressante des problèmes LQ : les filtres de Kalman. Le lecteur intéressé trouvera davantage de détails sur cette notion dans la section D.

II.3 Cas de systèmes non-linéaires

La version forte suivante du principe du maximum de Pontryagin (PMP) est très générale, puisqu'elle est valide pour des systèmes différentiels généraux et qu'elle prend en compte des contraintes sur le contrôle. On se référera par exemple à [3, 4] pour une démonstration.

On considère le système de contrôle dans \mathbb{R}^n

$$\dot{x}(t) = f(t, x(t), u(t)), \quad (\text{II.10})$$

où $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ est de classe C^1 et où les contrôles sont des applications mesurables et bornées définies sur un intervalle $[0, T_e(u)]$ de \mathbb{R}_+ et à valeurs dans $U \subset \mathbb{R}^m$.

Par ailleurs on définit la fonctionnelle de coût

$$C(T, u) = \int_0^T f^0(s, x(s), u(s)) ds + g(T, x(T))$$

où $f^0 : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ et $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ sont C^1 et $x(\cdot)$ est la trajectoire solution de (II.10) associée au contrôle u .

On considère le problème suivant :

Problème de contrôle optimal : déterminer une trajectoire reliant M_0 à M_1 et minimisant le coût. Le temps final peut être fixé ou non. Mathématiquement, ce problème se réécrit

$$\inf_{\substack{0 < T < T_e(u) \\ u \in \mathcal{U}}} C(T, u) \quad (\text{II.11})$$

où \mathcal{U} désigne l'ensemble des contrôles admissibles u tels que $u(t) \in U$ p.p., dont les trajectoires associées relient un point initial de M_0 à un point final de M_1 en temps $T(u) < T_e(u)$.

Définition II.3.1. Hamiltonien

On définit le Hamiltonien H du système de contrôle optimal (II.10)-(II.11) sur $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_- \times \mathbb{R}^m$ par

$$H(t, x, p, p^0, u) = \langle p, f(t, x, u) \rangle + p^0 f^0(t, x, u).$$

Théorème II.3.2. Principe du maximum de Pontryagin, version générale

Soient M_0 et M_1 deux sous-ensembles de \mathbb{R}^n .

Si le contrôle $u \in \mathcal{U}$ associé à la trajectoire $x(\cdot)$ est optimal sur $[0, T]$, alors il existe une application $p(\cdot) : [0, T] \rightarrow \mathbb{R}^n$ absolument continue appelée vecteur adjoint, et un réel $p^0 \leq 0$, tels que le couple $(p(\cdot), p^0)$ est non trivial, et tels que, pour presque tout $t \in [0, T]$,

$$\begin{aligned} \dot{x}(t) &= \frac{\partial H}{\partial p}(t, x(t), p(t), p^0, u(t)), \\ \dot{p}(t) &= -\frac{\partial H}{\partial x}(t, x(t), p(t), p^0, u(t)), \end{aligned} \quad (\text{II.12})$$

et on a la condition de maximisation presque partout sur $[0, T]$

$$H(t, x(t), p(t), p^0, u(t)) = \max_{v \in U} H(t, x(t), p(t), p^0, v). \quad (\text{II.13})$$

Si de plus le temps final pour joindre la cible M_1 n'est pas fixé, on a la condition au temps final^a T

$$\max_{v \in U} H(T, x(T), p(T), p^0, v) = -p^0 \frac{\partial g}{\partial t}(T, x(T)). \quad (\text{II.15})$$

^a. Si le contrôle u est continu au temps T , la condition (II.15) peut s'écrire

$$H(T, x(T), p(T), p^0, u(T)) = -p^0 \frac{\partial g}{\partial t}(T, x(T)). \quad (\text{II.14})$$

Remarque II.3.3 Espace tangent

Supposons que la variété M_1 est donnée par $M_1 = \{x \in \mathbb{R}^n \mid F(x) = 0\}$, où F est une fonction de classe C^1 de \mathbb{R}^n dans \mathbb{R}^p . En écrivant $F = (F_1, \dots, F_p)$ où les fonctions F_i sont à valeurs réelles, il vient $M_1 = \{x \in \mathbb{R}^n \mid F_1(x) = \dots = F_p(x) = 0\}$, et de plus l'espace tangent à M_1 en un point $x \in M_1$ est

$$T_x M_1 = \{v \in \mathbb{R}^n \mid \nabla F_i(x) \cdot v = 0, i = 1, \dots, p\}.$$

Remarque II.3.4 Conditions de transversalité

Si de plus M_0 et M_1 (ou juste l'un des deux ensembles) sont des variétés¹ de \mathbb{R}^n ayant des espaces tangents en $x(0) \in M_0$ et $x(T) \in M_1$, alors le vecteur adjoint peut être construit de manière à vérifier les conditions de transversalité aux deux extrémités (ou juste l'une des deux)

$$p(0) \perp T_{x(0)} M_0 \quad \text{et} \quad p(T) - p^0 \frac{\partial g}{\partial x}(T, x(T)) \perp T_{x(T)} M_1. \quad (\text{II.16})$$

De plus,

- si M_0 (resp. M_1) est réduit à un élément, cela n'impose aucune condition de transversalité sur $p(0)$ (resp. $p(T)$).
- si $M_0 = \mathbb{R}^n$ (resp. $M_1 = \mathbb{R}^n$), alors, cela impose la condition $p(0) = 0$ (resp. $p(T) - p^0 \frac{\partial g}{\partial x}(T, x(T)) = 0$).
- si la variété M_1 s'écrit sous la forme

$$M_1 = \{x \in \mathbb{R}^n \mid F_1(x) = \dots = F_p(x) = 0\},$$

où les F_i sont des fonctions de classe C^1 sur \mathbb{R}^n (indépendantes puisque M_1 est une variété), alors la deuxième condition de (II.16) portant sur $p(T)$ se met sous la forme

$$\exists \lambda_1, \dots, \lambda_p \in \mathbb{R} \mid p(T) = \sum_{i=1}^p \lambda_i \nabla F_i(x(T)) + p^0 \frac{\partial g}{\partial x}(T, x(T)). \quad (\text{II.17})$$

- Si l'extrémité est soumise à une contrainte d'inégalité, de la forme $\phi(x(T)) \leq 0$ où $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ est de classe C^1 , alors il existe $p^1 \geq 0$ tel que

$$p(T) - p^0 \frac{\partial g}{\partial x}(T, x(T)) = p^1 \frac{\partial \phi}{\partial x}(x(T)), \quad p^1 \phi(x(T)) = 0 \quad (\text{condition de complémentarité}).$$

Remarque II.3.5 Une conséquence supplémentaire

Dans les conditions du théorème, on a de plus pour presque tout $t \in [0, T]$

$$\frac{d}{dt} H(t, x(t), p(t), p^0, u(t)) = \frac{\partial H}{\partial t}(t, x(t), p(t), p^0, u(t)). \quad (\text{II.18})$$

1. Soit $M \subset \mathbb{R}^n$. On dit que M est une sous-variété de \mathbb{R}^n de dimension p et de classe C^k si pour tout x de M , il existe des voisinages respectifs U de x dans \mathbb{R}^n et V de 0 dans \mathbb{R}^n , ainsi qu'un C^k -difféomorphisme $f : U \rightarrow V$, envoyant x sur 0 et telle que $f(U \cap M) = V \cap (\mathbb{R}^p \times \{0\})$.

En effet, cela peut se voir formellement, dans le cas où aucune contrainte n'est imposée au contrôle, en appliquant la règle de différentiation des fonctions composées et en supposant u dérivable en temps. Omettons d'écrire la dépendance en $(t, x(t), p(t), p^0, u(t))$ pour plus de clarté :

$$\frac{d}{dt}H = \frac{\partial H}{\partial t} + \frac{\partial H}{\partial x}\dot{x} + \frac{\partial H}{\partial p}\dot{p} + \frac{\partial H}{\partial u}\dot{u} = \frac{\partial H}{\partial t},$$

d'après les équations fournies par le PMP.

En particulier si le système augmenté est *autonome*, i.e. si f et f^0 ne dépendent pas de t , alors H ne dépend pas de t , et on a

$$\forall t \in [0, T] \quad \max_{v \in \Omega} H(x(t), p(t), p^0, v) = \text{Cste.}$$

Notons que cette égalité est alors valable partout sur $[0, T]$ (en effet cette fonction de t est lipschitzienne).

Remarque II.3.6 analyse des problèmes LQ

Considérons un système dynamique linéaire autonome

$$\begin{cases} \dot{x}(t) = Ax(t) + Bv(t) & t \in [0, T], \\ x(0) = x_0 \in \mathbb{R}^n \end{cases} \quad (\text{II.19})$$

associé à la fonction coût quadratique :

$$J(v) = \frac{1}{2} \int_0^T [(Qx(s), x(s)) + (Rv(s), v(s))] ds + \frac{1}{2} (Dx(T), x(T))$$

où R est une matrice de $\mathbb{M}_m(\mathbb{R})$ définie positive et $Q \in \mathbb{M}_n(\mathbb{R})$ et $D \in \mathbb{M}_n(\mathbb{R})$ sont supposées semi-définies positives. Le Hamiltonien est défini par $H(t, x, p, p^0, v) = \langle p, Ax + Bv \rangle + p^0((Qx, x) + (Rv, v))$ et en appelant v^* le contrôle optimal minimisant J (on a montré son existence et son unicité en cours), la condition de maximisation presque partout fournit : $v^*(t)$ résout le problème $\max_{v \in \mathbb{R}^m} \langle p, Bv \rangle + p^0(Rv, v)$ et par conséquent, $v^*(t)$ satisfait la condition d'optimalité au premier ordre

$$p^0 R v^*(t) = -B^\top p(t), \quad t \in [0, T].$$

De plus, la condition de transversalité s'écrit : $p(T) = p^0 D x(T)$. Si $p^0 = 0$, alors $p(T) = 0$ et on en déduit que $p(\cdot) = 0$ par unicité des solutions du système de Cauchy. C'est absurde car (p^0, p) est non trivial. Par conséquent, $p^0 < 0$ et $p^0 R v^*(t) + B^\top p = 0$, où p désigne l'état adjoint solution du problème

$$-p' = A^\top p + p^0 Qx \quad \text{et} \quad p(T) = p^0 D x(T).$$

Posons $\tilde{p} = p/p^0$. Alors, la condition d'optimalité devient $v^*(t) = R^{-1} B^\top \tilde{p}$ et \tilde{p} résout

$$-\tilde{p}' = A^\top \tilde{p} + Qx \quad \text{et} \quad \tilde{p}(T) = D x(T).$$

Remarque II.3.7 principe du maximum faible

Dans le cas où $U = \mathbb{R}^m$, i.e. lorsqu'il n'y a pas de contrainte sur le contrôle, la condition de

maximum (II.13) devient $\frac{\partial H}{\partial u} = 0$.

Exemple II.3.8 Contrôle optimal d'un ressort non linéaire

Ce système est modélisé par le système de contrôle

$$\begin{cases} \dot{x}(t) = y(t), \\ \dot{y}(t) = -x(t) - 2x(t)^3 + u(t), \end{cases}$$

où on autorise comme contrôles toutes les fonctions $u(t)$ continues par morceaux telles que $|u(t)| \leq 1$. L'objectif est d'amener le ressort d'une position initiale quelconque $(x_0, y_0 = \dot{x}_0)$ à sa position d'équilibre $(0,0)$ en temps minimal t_* .

Application du Principe du Maximum. Le Hamiltonien du système précédent s'écrit

$$H(x, p, u) = p_x y + p_y (-x - 2x^3 + u) + p^0,$$

et si (x, p, u) est une extrémale alors on doit avoir

$$\dot{p}_x = -\frac{\partial H}{\partial x} = p_y(1 + 6x^2), \text{ et } \dot{p}_y = -\frac{\partial H}{\partial y} = -p_x.$$

Notons que puisque le vecteur adjoint (p_x, p_y) doit être non trivial, p_y ne peut s'annuler sur un intervalle (sinon on aurait également $p_x = -\dot{p}_y = 0$). Par ailleurs la condition de maximisation nous donne

$$p_y u = \max_{|v| \leq 1} p_y v.$$

Comme p_y ne s'annule sur aucun intervalle, on en déduit que, presque partout,

$$u(t) = \text{signe } p_y(t).$$

En particulier les contrôles optimaux sont successivement égaux à ± 1 , c'est le principe *bang-bang* (voir [3]). Plus précisément, le vecteur adjoint au temps final t_* étant défini à scalaire multiplicatif près, on peut affirmer

$$u(t) = \text{signe}(p_y(t)) \text{ où } p_y \text{ est la solution de } \begin{cases} \ddot{p}_y(t) + p_y(t)(1 + 6x(t)^2) = 0, \\ p_y(t_*) = \cos \alpha, \dot{p}_y(t_*) = -\sin \alpha, \end{cases}$$

le paramètre $\alpha \in [0, 2\pi[$ étant indéterminé.

En inversant le temps ($t \mapsto -t$), il est clair que notre problème est équivalent au problème du temps minimal pour le système

$$\begin{cases} \dot{x}(t) = -y(t) \\ \dot{y}(t) = x(t) + 2x(t)^3 - \text{signe}(p_y(t)) \\ \dot{p}_y(t) = p_x(t) \\ \dot{p}_x(t) = -p_y(t)(1 + 6x(t)^2) \end{cases} \quad (\text{II.20})$$

avec

$$x(0) = y(0) = 0, \quad x(t_*) = x_0, \quad y(t_*) = y_0, \quad p_y(0) = \cos \alpha, \quad p_x(0) = \sin \alpha,$$

où $\alpha \in [0, 2\pi[$ est à déterminer.

II.4 Exercices du chapitre

Problèmes temps optimal & LQ pour les systèmes linéaires

II.1 Minimiser la fonction

$$\int_0^1 u^2(t) dt + x^2(1)$$

sous la contrainte dynamique $x'(t) = x(t) + u(t)$ et $x(0) = 1$.

II.2 (*mouvement d'un point matériel*) Soit $T > 0$. Considérons, le système de contrôle

$$\begin{cases} \dot{x}(t) = u(t) & t \in [0, T] \\ x(0) = 0 \in \mathbb{R} \end{cases}$$

où $x(\cdot)$ et (\cdot) sont deux applications définies sur $[0, T]$ et à valeurs réelles.

(i) Dans cette question, on suppose que $T = 1$ et on considère le coût C_1 défini par

$$C_1(u) = \int_0^1 (x(t)^2 + (|u(t)| - 1)^2) dt,$$

On considère le contrôle u_n défini pour tout $n \in \mathbb{N}^*$ par

$$u_n = (-1)^k \quad \text{si } t \in \left[\frac{k}{2n}, \frac{k+1}{2n} \right], \quad k \in \llbracket 0, 2n-1 \rrbracket.$$

Calculer $C_1(u_n)$. Que peut-on en déduire?

(ii) Soit le coût C_2 défini par

$$C_2(u) = \int_0^T (x(t)^2 + u(t)^2) dt,$$

où x désigne la solution du système ci-dessus.

- Montrer l'existence d'une trajectoire optimale et la caractériser. Quelle est la valeur optimale du coût?
- Résoudre l'équation de Riccati, écrire le contrôle optimal comme un feedback et déterminer (à nouveau) la valeur minimale atteinte par le critère.

II.3 (*contrôle optimal d'un tram*) Soit $T > 0$. Considérons le problème du véhicule se déplaçant en ligne droite, modélisé par le système de contrôle

$$\begin{cases} \ddot{x}(t) = u(t) & t \in [0, T] \\ x(0) = \dot{x}(0) = 0. \end{cases}$$

On souhaite, pendant le temps T , maximiser la distance parcourue tout en minimisant l'énergie fournie. On choisit critère

$$C(u) = -x(T) + \int_0^T u(t)^2 dt.$$

Montrer l'existence d'une trajectoire optimale et la caractériser.

II.4 (trajectoires temps-optimales) Soit $T > 0$ et $U = [-1, 1]$. On considère le système contrôlé

$$\begin{cases} x^{(3)}(t) + 6x''(t) + 12x'(t) + 8x(t) = u(t) & t \in]0, T[\\ (x(0), x'(0), x''(0)) = (x_0, x_1, x_2) \in \mathbb{R}^3, \end{cases} \quad (2)$$

où $u \in L^2([0, T], \mathbb{R})$ est une fonction de contrôle de ce système.

- (i) Mettre le système (2) sous la forme usuelle $X' = AX + Bu$ et $X(0) = X_0$ où X est un vecteur, A et B sont deux matrices à préciser.
- (ii) Ce système est-il contrôlable?
- (iii) On suppose dorénavant que $u \in L^\infty([0, T[, U)$ et que T n'est plus fixé. On souhaite rejoindre l'origine avec un contrôle $u(\cdot)$ en temps T minimal.
 - (a) Déterminer les valeurs propres de A . On pourra par exemple remarquer que -2 est valeur propre évidente.
 - (b) Étudier l'existence d'une trajectoire temps-optimale.
 - (c) Démontrer que le contrôle optimal est bang-bang et qu'il a au plus deux commutations.

Principe du maximum de Pontryagin

II.5 (contrôle optimal en économie) Un consommateur a une durée de vie $T > 0$. Il gagne un salaire avec un taux $\alpha > 0$ constant (par unité de temps). Soit $x(t)$ son salaire accumulé et i le taux de rémunération de l'épargne ou le taux d'intérêt de la dette. La consommation au temps t est notée $u(t)$. La fonction "salaire $x(\cdot)$ du consommateur" est solution du système différentiel

$$x'(t) = \alpha + ix(t) - u(t),$$

et $x(0) = x(T) = 0$ (il n'y a ni héritage ni leg). La fonction $u(\cdot)$ agit ici comme un contrôle. On souhaite déterminer la consommation qui maximise la fonction d'utilité J définie par

$$J(u) = \int_0^T \ln u(t) e^{-\rho t} dt$$

où $\rho > 0$ est un paramètre tel que $i > \rho$.

- (i) Résoudre le problème de Cauchy $\begin{cases} x'(t) = \alpha + ix(t) - u(t) & t \geq 0 \\ x(0) = 0. \end{cases}$
- (ii) Définir le Hamiltonien associé à ce problème.
- (iii) On admet l'existence d'un contrôle optimal. Énoncer le principe du maximum de Pontryagin pour ce problème, en précisant les conditions de transversalité s'il y a lieu.
- (iv) Préciser la trajectoire optimale $x(\cdot)$ et démontrer que l'on a $u(t) = \alpha \frac{g(\rho)}{g(i)} e^{(i-\rho)t}$ où g est une fonction à préciser.
- (v) Comparer $u(t)$ et α lorsque t prend de grandes valeurs. Commenter.

II.6 (problème de la reine Didon) Soient ℓ et L positifs tels que $2\ell < L < \pi\ell$. On cherche à résoudre le problème :

$$\text{maximiser } \int_{-\ell}^{\ell} x(t) dt \text{ par rapport à } u \in C^0([-\ell, \ell]),$$

où (x, y) désigne la solution du système dynamique

$$\begin{cases} \dot{x}(t) = u(t), & t \geq 0 \\ \dot{y}(t) = \sqrt{1 + u(t)^2}, & t \geq 0 \\ x(-\ell) = x(\ell) = 0, \\ y(-\ell) = 0, y(\ell) = L, \end{cases}$$

On **admet** que ce problème possède une solution.

- (i) On définit la fonction \hat{h} sur \mathbb{R} par $\hat{h}(u) = p_1 u + p_2 \sqrt{1 + u^2}$ où $(p_1, p_2) \in \mathbb{R}^2$. On suppose que le problème $\sup_{\mathbb{R}} \hat{h}$ possède une solution. Montrer que, nécessairement, $p_2 < 0$.
- (ii) Caractériser le contrôle optimal. On montrera en particulier :
 - qu'il existe $a > 0$ et $b \in \mathbb{R}$ tels que le contrôle optimal u vérifie

$$\frac{u(t)}{\sqrt{1 + u(t)^2}} = -\frac{t - b}{a}.$$

- que $(t, x(t))$ se trouve sur une portion de cercle.

II.7 (*problème de temps minimal*) Considérons le système de contrôle

$$\begin{cases} \dot{x}(t) = y(t) + u(t) & t \geq 0 \\ \dot{y}(t) = -y(t) + u(t) & t \geq 0 \\ |u(\cdot)| \leq 1. \end{cases}$$

Le but est de joindre en temps minimal la droite $x = 0$, puis de rester sur cette droite.

On considère la cible $M_1 = \{(0, y) \mid |y| \leq 1\}$ et on s'intéresse à l'existence et la caractérisation de trajectoires temps optimales pour ce problème.

- (i) Montrer que si un tel contrôle existe, alors on a nécessairement $|y(t)| \leq 1$ lorsque $x(t) = 0$. Réciproquement, montrer que de tout point $(0, y) \in M_1$, part une trajectoire restant dans M_1 .
- (ii) Etudier l'existence d'une trajectoire temps optimale.
- (iii) On cherche à caractériser les trajectoires temps optimales. On choisit de raisonner "en temps inverse", en calculant les trajectoires joignant M_1 à tout point final.
 - (a) Écrire le Hamiltonien du problème de contrôle optimal et les équations adjointes. En déduire que le contrôle optimal est *bang-bang* avec au plus une commutation.
 - (b) En utilisant les conditions de transversalité, montrer que si les conditions initiales sont choisies dans $\{(0, y) \mid |y| < 1\}$, alors le contrôle optimal ne commute pas. Représenter quelques unes des trajectoires associées.

II.8 (*Contrôle optimal d'un procédé de fermentation*) Considérons le procédé de fermentation

$$\begin{cases} \dot{x}(t) = -x(t) + u(t)(1 - x(t)) & t \geq 0, \\ \dot{y}(t) = x(t) - u(t)y(t) & t \geq 0 \\ x(0) = x_0, y(0) = 0, \end{cases}$$

où $x(t)$ représente la concentration de sucre, $y(t)$ la concentration d'éthanol, et $u(t)$, le contrôle, est le taux d'évaporation. On suppose $0 \leq u(t) \leq M$, et $0 < x_0 < 1$. Soit y_1 tel que $y_1 > 1/M$ et $y_1 > x_0$; on veut résoudre le problème du temps minimal pour rejoindre $y(t_f) = y_1$.

- (i) Montrer que $x_0 e^{-t} \leq x(t) < 1$, pour tout $t \in [0, t_f]$.

Indication : pour montrer que $x(\cdot) \leq 1$, on suggère d'écrire l'équation satisfaite par $\tilde{x} : t \mapsto x(t) - 1$.

- (ii) On note les variables adjointes (p_x, p_y) et p^0 .

(a) Ecrire le Hamiltonien du problème de contrôle optimal et appliquer le PMP.

(b) Ecrire les conditions de transversalité sur le vecteur adjoint et sur le temps. Montrer que $p_y(t_f) \neq 0$.

- (iii) (a) Pour tout $t \in [0, t_f]$, soit $\varphi(t) = p_x(t)(1 - x(t)) - p_y(t)y(t)$. Calculer $\varphi'(t)$ et $\varphi''(t)$. Montrer que φ est strictement monotone.

(b) En déduire que les contrôles optimaux sont bang-bang avec au plus une commutation, et préciser leur expression.

(c) Montrer que $\dot{y}(t_f) \geq 0$.

(d) En déduire qu'il existe $\varepsilon > 0$ tel que $u(t) = 0$, pour presque tout $t \in [t_f - \varepsilon, t_f]$.

(e) Conclure sur la structure du contrôle optimal.

III. Méthodes d'adjoint pour les problèmes de contrôle optimal sous contrainte EDP

III.1 Optimisation dans des espaces de Hilbert

Soit H un espace de Hilbert muni de la norme $\|\cdot\|$ et du produit scalaire (\cdot, \cdot) . Soit J une fonction continûment dérivable de H dans \mathbb{R} et soit \mathcal{U}_{ad} un sous-ensemble convexe fermé de H . Nous considérons le problème de minimisation suivant :

$$\inf_{v \in \mathcal{U}_{ad}} J(v). \quad (\mathcal{P})$$

III.1.1 Rappels sur les fonctions convexes

Rappelons quelques propriétés fondamentales des fonctions convexes dont les démonstrations sont laissées au lecteur.

Définition III.1.1. Fonction convexe

Une fonction $J: H \rightarrow \mathbb{R}$ est convexe si

$$\forall u, v \in H, \forall \theta \in [0, 1], \quad J(\theta u + (1 - \theta)v) \leq \theta J(u) + (1 - \theta)J(v),$$

et strictement convexe si

$$\forall u \neq v \in H, \forall \theta \in]0, 1[, \quad J(\theta u + (1 - \theta)v) < \theta J(u) + (1 - \theta)J(v).$$

Proposition III.1.2. Fonctions convexes différentiables

Soit $J: H \rightarrow \mathbb{R}$ une fonction différentiable. Alors

(i) J est convexe si et seulement si

$$J(v) \geq J(u) + (\nabla J(u), v - u), \quad \forall u, v \in H.$$

(ii) J est strictement convexe si et seulement si

$$J(v) > J(u) + (\nabla J(u), v - u), \quad \forall u \neq v \in H.$$

Définition III.1.3. Fonctions fortement convexes

Une fonction $J : H \rightarrow \mathbb{R}$ est dite *fortement convexe* ou α -convexe si

$$J\left(\frac{u+v}{2}\right) \leq \frac{J(v) + J(u)}{2} - \frac{\alpha}{8} \|v - u\|^2, \quad \forall u, v \in H.$$

Proposition III.1.4. Caractérisation des fonctions fortement convexes

Soit $J : H \rightarrow \mathbb{R}$ une fonction différentiable. Alors les conditions suivantes sont équivalentes

- (i) J est α -convexe.
- (ii) J est minorée par une fonction quadratique

$$J(v) \geq J(u) + (\nabla J(u), v - u) + \frac{\alpha}{2} \|v - u\|^2, \quad \forall u, v \in H.$$

- (iii) J est α -elliptique

$$(\nabla J(v) - \nabla J(u), v - u) \geq \alpha \|v - u\|^2, \quad \forall u, v \in H.$$

III.1.2 Théorème d'existence et inéquation d'Euler

Théorème III.1.5. Existence d'un minimum de fonction α -convexe

Soient $J : H \rightarrow \mathbb{R}$ une fonction continue, α -convexe et \mathcal{U}_{ad} une partie non-vide convexe fermée. Alors le problème (\mathcal{P}) admet une unique solution.

Théorème III.1.6. C.N.S. d'optimalité

Supposons que $J : H \rightarrow \mathbb{R}$ est une fonction différentiable et convexe. Alors $u \in \mathcal{U}$ est solution du problème (\mathcal{P}) si et seulement s'il satisfait l'inéquation d'Euler

$$(\nabla J(u), v - u) \geq 0, \quad \forall v \in \mathcal{U}_{ad}.$$

Remarque III.1.7 Cas particuliers d'ensembles admissibles

- (i) Dans le cas sans contrainte ($\mathcal{U}_{ad} = H$), ou \mathcal{U}_{ad} contient un voisinage de u , l'inéquation d'Euler devient l'équation d'Euler $\nabla J(u) = 0$.
- (ii) On dit que \mathcal{U}_{ad} est un cône de sommet d'origine, si

$$\forall u \in \mathcal{U}_{ad}, \quad \lambda \geq 0 \implies \lambda \cdot u \in \mathcal{U}_{ad}.$$

Lorsque \mathcal{U}_{ad} est un cône convexe fermé de sommet d'origine, alors en prenant $v = \theta u \in \mathcal{U}_{ad}$ dans l'inéquation d'Euler, on obtient $(\theta - 1)(\nabla J(u), u) \geq 0$, pour tout $\theta \geq 0$, ce qui implique

$$(\nabla J(u), u) = 0 \quad \text{et} \quad (\nabla J(u), v) \geq 0 \quad \forall v \in \mathcal{U}_{ad}.$$

III.2 Contrôle optimal de problèmes elliptiques

III.2.1 Contrôle et observation distribués

Dans ce paragraphe, Ω désigne un ouvert connexe borné à bord régulier. On désigne par $\|\cdot\|$ la norme dans $L^2(\Omega)$ et par (\cdot, \cdot) le produit scalaire correspondant. On désigne par \mathcal{U}_{ad} un sous-ensemble convexe et fermé (non vide) de $L^2(\Omega)$.

Soient $f, z_d \in L^2(\Omega)$ deux fonctions connues et $\alpha > 0$ une constante positive. On considère l'équation elliptique de Dirichlet

$$\begin{cases} y - \Delta y = f + v, & \text{dans } \Omega, \\ y = 0 & \text{sur } \partial\Omega \end{cases} \quad (\text{III.1})$$

et la fonction quadratique

$$J(v) = \frac{1}{2} \|y - z_d\|^2 + \frac{\alpha}{2} \|v\|^2.$$

Le problème du contrôle optimal est de trouver un contrôle $u \in \mathcal{U}_{ad} \subset L^2(\Omega)$ tel que la solution y du problème elliptique ci-dessus avec $v = u$ minimise la fonctionnelle quadratique J sur l'ensemble des contrôles admissibles \mathcal{U}_{ad} :

$$\inf_{v \in \mathcal{U}_{ad}} J(v). \quad (\mathcal{P})$$

Théorème III.2.1. Conditions d'optimalité

Le problème du contrôle optimal (\mathcal{P}) admet une unique solution qui est caractérisée par le système suivant

$$\begin{cases} y - \Delta y = f + u, & \text{dans } \Omega, \\ p - \Delta p = y - z_d, & \text{dans } \Omega, \\ y = 0, \quad p = 0 & \text{sur } \partial\Omega, \\ (p + \alpha u, v - u) \geq 0, \quad \forall v \in \mathcal{U}_{ad}. \end{cases} \quad (\text{III.2})$$

La preuve repose de façon cruciale sur le résultat suivant :

Proposition III.2.2. Gradient de J

La fonctionnelle J est différentiable sur $L^2(\Omega)$ et son gradient est donné par

$$\nabla J(v) = p + \alpha v,$$

où p désigne l'unique solution du problème adjoint

$$\begin{cases} p - \Delta p = y - z_d, & \text{dans } \Omega, \\ p = 0 & \text{sur } \partial\Omega, \end{cases} \quad (\text{III.3})$$

Exemple III.2.3 cas sans contrainte

Si $\mathcal{U}_{ad} = L^2(\Omega)$, alors la dernière condition de (III.2) devient $u = -\frac{1}{\alpha}p$, si bien que y et p satisfont

$$\begin{cases} y - \Delta y + \frac{1}{\alpha}p = f, & \text{dans } \Omega, \\ p - \Delta p - y = -z_d, & \text{dans } \Omega, \\ y = 0, \quad p = 0 & \text{sur } \partial\Omega. \end{cases}$$

Par régularité elliptique (voir [1]) le contrôle u appartient à $H^2(\Omega)$.

Exemple III.2.4 Cas de contacts unilatéraux

Supposons que $\mathcal{U}_{ad} = \{v \in L^\infty(\Omega) \mid v \geq 0 \text{ p.p. dans } \Omega\}$. C'est un cône fermé convexe de sommet l'origine. Dans ce cas, le système (III.2) devient le système non-linéaire

$$\begin{cases} y - \Delta y + \frac{1}{\alpha} \min(0, p) = f, & \text{dans } \Omega, \\ p - \Delta p - y = -z_d, & \text{dans } \Omega, \\ y = 0, \quad p = 0, & \text{sur } \partial\Omega \end{cases}$$

De plus, on peut montrer en utilisant la théorie des inéquations elliptiques (Brézis-Stampacchia) que le contrôle u appartient à $H^1(\Omega)$.

Remarque III.2.5 Formulaire

- L'équation adjointe introduite permet d'écrire le gradient d'une fonctionnelle sous une forme exploitable, permettant en particulier d'écrire des conditions d'optimalité (inéquation d'Euler). Cet exemple fournit bel et bien une méthode de calcul du gradient J que nous pouvons généraliser sans difficulté essentielle à des formulations variationnelles plus générales. En effet, on se donne $f \in L^2(\Omega)$, $\sigma \in L^\infty(\Omega)$ une application telle que $\sigma(\cdot) \geq \sigma_0 > 0$ sur Ω . Soit $b \in \mathbb{R}^n$ et $c \in L^\infty(\Omega)$ telle que $c(\cdot) \geq 0$ dans Ω . On considère la solution $y \in H_0^1(\Omega)$ de l'équation

$$\begin{cases} -\operatorname{div}(\sigma \nabla y(x)) + \langle b, \nabla y(x) \rangle + c(x)y(x) = f + v & \text{dans } \Omega \\ y = 0 & \text{sur } \partial\Omega \end{cases}$$

ainsi que le critère J défini par

$$J(v) = \int_{\Omega} j(y(x), v(x)) dx$$

avec $j : \mathbb{R}^2 \rightarrow \mathbb{R}$, telle que pour tout $(x_1, x_2) \in \mathbb{R}^2$, $|j(x_1, x_2)| \leq C(|x_1| + |x_2|)$ pour $C > 0$. Alors, on peut montrer (exercice) en adaptant la démarche précédente que

$$(\nabla J(u), v - u) = \int_{\Omega} (v(x) - u(x)) \left(p + \frac{\partial j}{\partial x_2}(y(x), v(x)) \right) dx,$$

où p désigne l'état adjoint défini comme la solution de l'équation

$$\begin{cases} -\operatorname{div}(\sigma \nabla p(x)) - \langle b, \nabla p(x) \rangle + c(x)p(x) = \frac{\partial j}{\partial x_1}(y(x), v(x)) & \text{dans } \Omega \\ p = 0 & \text{sur } \partial\Omega. \end{cases}$$

Plutôt que de travailler dans un cadre abstrait, nous préférons traiter les problèmes cas par cas. On s'apercevra qu'à peu de chose près, les démarches prises dans le calcul du gradient de J sont essentiellement les mêmes. Cette remarque reste valable pour les problèmes paraboliques que nous considérerons ultérieurement.

- Le nom *d'état adjoint* provient du fait que la définition de p fait intervenir l'adjoint au sens des distributions de l'opérateur du problème considéré. Dans ce problème l'opérateur $-\Delta$ avec condition de Dirichlet est auto-adjoint. Donc, l'équation adjointe en p est la même que l'équation d'état en y .
- L'introduction *d'état adjoint* n'est pas toujours nécessaire lorsque le problème est "auto-adjoint" (voir la feuille de TD n° 1).

III.2.2 Algorithme de minimisation pour le problème (P)

À partir de la proposition III.2.2 et en particulier du calcul du gradient de J , il est aisé de déduire des résultats précédents un algorithme de minimisation. On désigne par $\Pi_{\mathcal{U}_{ad}}$ la projection sur le sous-ensemble convexe fermé non vide \mathcal{U}_{ad} .

Données: $v^0 \in \mathcal{U}_{ad}$, $\varepsilon > 0$, $\rho > 0$ et $\text{Res} = 1$

tant que $\text{Res} \geq \varepsilon$, **faire**

On détermine y^k et p^k , les solutions respectives des EDP

$$\begin{cases} y^k - \Delta y^k = f + v^k & \text{dans } \Omega \\ y^k = 0 & \text{sur } \partial\Omega, \end{cases}$$

et

$$\begin{cases} p^k - \Delta p^k = y^k - z_d & \text{dans } \Omega \\ p^k = 0 & \text{sur } \partial\Omega, \end{cases}$$

On pose $v^{k+1} = \Pi_{\mathcal{U}_{ad}}(v^k - \rho(p^k + \alpha v^k))$ et $\text{Res} = \|v^{k+1} - v^k\|$;

fin

Algorithme 1: Algorithme de résolution du problème (P) à l'aide d'une méthode de gradient projeté (à pas constant)

III.2.3 Contrôle frontière et observation distribuée

Dans ce paragraphe, Ω désigne un ouvert connexe borné à bord régulier. Soient $f, z_d \in L^2(\Omega)$ deux fonctions connues et $\alpha > 0$ une constante positive. On considère le problème elliptique de Dirichlet non-homogène

$$\begin{cases} y - \Delta y = f, & \text{dans } \Omega, \\ y = v & \text{sur } \partial\Omega \end{cases} \quad (\text{III.4})$$

associé à la fonction coût quadratique

$$J(v) = \frac{1}{2} \|y - z_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|v\|_{L^2(\partial\Omega)}^2.$$

On considère le problème du contrôle optimal frontière : trouver un contrôle $u \in \mathcal{U}_{ad} \subset L^2(\partial\Omega)$ (un convexe fermé non vide) tel que la solution y du problème ci-dessus (avec $v = u$)

minimise J sur \mathcal{U}_{ad} , i.e.

$$\inf_{v \in \mathcal{U}_{ad}} J(v).$$

Avant de parler du problème du contrôle, il faut donner un sens clair aux solutions du système non homogène (III.4) avec des données irrégulières $v \in L^2(\partial\Omega)$. En effet, grâce à la théorie des équations elliptiques, pour toutes données $v \in L^2(\partial\Omega)$ et $f \in L^2(\Omega)$ (largement suffisant), le système non homogène (III.4) admet une seule solution $y \in L^2(\Omega)$. En particulier, l'application affine

$$L^2(\partial\Omega) \ni v \rightarrow y(v) \in L^2(\Omega)$$

est continue pour les topologies correspondantes. Ainsi la fonction J est bien définie.

On a le résultat suivant caractérisant les solutions du problème ci-dessus.

Théorème III.2.6. Contrôle optimal frontière

Le problème du contrôle optimal (\mathcal{P}) admet une unique solution qui est caractérisée par le système suivant

$$\begin{cases} y - \Delta y = f, & \text{dans } \Omega, \\ p - \Delta p = y - z_d, & \text{dans } \Omega, \\ y = u, \quad p = 0, & \text{sur } \partial\Omega, \\ (\alpha u - \partial_\nu p, v - u) \geq 0, & \forall v \in \mathcal{U}_{ad}. \end{cases}$$

Remarque III.2.7 Vers un algorithme de minimisation

En notant que le gradient de J (qui est un élément de $L^2(\partial\Omega)$) est, dans ce cas, donné par

$$\nabla J(v) = \alpha v - \partial_\nu p,$$

il est aisé d'adapter l'approche développée dans la section III.2.2 pour obtenir un algorithme de minimisation par une méthode de gradient.

Exemple III.2.8 Cas sans contrainte

Si $\mathcal{U}_{ad} = L^2(\Omega)$, alors la condition “ $(\alpha u - \partial_\nu p, v - u) \geq 0, \forall v \in \mathcal{U}_{ad}$ ” devient $u = \frac{1}{\alpha} \partial_\nu p$ et le système d'optimalité devient

$$\begin{cases} y - \Delta y = f & \text{dans } \Omega, \\ p - \Delta p - y = -z_d & \text{dans } \Omega, \\ y = \frac{1}{\alpha} \partial_\nu p, \quad p = 0 & \text{sur } \partial\Omega. \end{cases}$$

D'après la théorie des équations elliptiques, le contrôle u appartient à $H^2(\Omega)$.

III.3 Contrôle optimal de l'équation de la chaleur

Dans cette section, Ω désignera un ouvert connexe borné de bord régulier, disons C^2 . Soit $T > 0$, un horizon de temps. Nous utiliserons régulièrement les notations :

$$Q =]0, T[\times \Omega \quad \text{et} \quad \Sigma =]0, T[\times \partial\Omega.$$

III.3.1 Préliminaires

Soit X un espace de Banach, on désigne par $C^m(0, T; X)$ l'espace des fonctions m fois continûment dérivables sur $[0, T]$ à valeurs dans X . Il s'agit bien d'un espace de Banach pour la norme

$$\|y\|_{C^m(0, T; X)} = \max_{0 \leq l \leq m} \sup_{0 \leq t \leq T} \|\partial_t^l y(t)\|_X.$$

D'autre part, on désigne par $L^2(0, T; X)$ l'espace des fonctions fortement mesurables sur $[0, T]$ (c'est à dire que $t \mapsto \|v(t)\|_X$ est mesurable pour la mesure dt) telles que

$$\|y\|_{L^2(0, T; X)}^2 = \int_0^T \|y(t)\|_X^2 dt < \infty.$$

Si X est un espace de Hilbert, l'espace $L^2(0, T; X)$ est aussi un espace de Hilbert pour le produit scalaire

$$(y, z)_{L^2(0, T; X)} = \int_0^T (y(t), z(t))_X dt.$$

Maintenant on se donne

- (i) deux espaces de Hilbert V, H muni de la norme $\|\cdot\|$ dans V et de la norme $|\cdot|$ et du produit scalaire (\cdot, \cdot) dans H tels que

$$V \subset H \quad \text{avec injection dense et compacte}$$

- (ii) une forme bilinéaire symétrique continue $a(\cdot, \cdot)$ qui est coercive au sens qu'il existe des constantes $\lambda \geq 0, \alpha > 0$ telles que

$$a(y, y) + \lambda |y|^2 \geq \alpha \|y\|^2, \quad \forall y \in V.$$

- (iii) deux fonctions $y_0 \in H$ et $f \in L^2(0, T; H)$.

On cherche une fonction y telle que

$$y \in C^0(0, T; H) \cap L^2(0, T; V) \tag{III.5}$$

et satisfaisant l'équation variationnelle

$$\frac{d}{dt}(y(t), \phi) + a(y(t), \phi) = (f(t), \phi), \quad y(0) = y_0 \tag{III.6}$$

au sens des distributions sur $]0, T[$ pour toute fonction test $\phi \in V$.

Une telle fonction y est appelée solution (faible) du problème variationnel (III.5)-(III.6).

Nous admettons le résultat suivant dont la démonstration repose (par exemple) sur la théorie de Hille-Yosida (voir par exemple [1]).

Théorème III.3.1. Existence, unicité, régularité

Sous les hypothèses (i)-(iii), le problème variationnel (III.5)-(III.6) admet une unique solution faible. De plus, l'application

$$H \times L^2(0, T; H) \ni (y_0, f) \mapsto y \in C^0(0, T; H) \cap L^2(0, T; V)$$

est continue pour les normes correspondantes.

Remarque III.3.2 Application à l'équation de la chaleur - conditions de Dirichlet ou Neumann

Considérons l'équation de la chaleur avec les conditions aux bords de Dirichlet

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = f & \text{dans } Q \\ y = 0 & \text{sur } \Sigma \\ y(0, \cdot) = y_0(\cdot) & \text{dans } \Omega. \end{cases}$$

On va chercher à appliquer le résultat précédent au système ci-dessus. A l'aide d'une intégration par parties, on trouve facilement

$$\frac{d}{dt} \int_{\Omega} y \phi dx + \int_{\Omega} \nabla y \nabla \phi dx = \int_{\Omega} f \phi dx, \quad \forall \phi \in H_0^1(\Omega).$$

Ainsi, choisissons $H = L^2(\Omega)$, $V = H_0^1(\Omega)$, il découle du théorème III.3.1 que si $y_0 \in L^2(\Omega)$ et $f \in L^2(0, T; L^2(\Omega))$, l'équation de la chaleur ci-dessus possède une unique solution faible.

De même, considérons l'équation de la chaleur ci-dessus avec des conditions aux bords de Neumann homogènes, i.e. $\frac{\partial y}{\partial \nu} = 0$ sur Σ . A l'aide d'une intégration par parties, on trouve

$$\frac{d}{dt} \int_{\Omega} y \phi dx + \int_{\Omega} \nabla y \nabla \phi dx = \int_{\Omega} \phi dx, \quad \forall \phi \in H^1(\Omega).$$

Le théorème abstrait s'applique à nouveau avec $V = H^1(\Omega)$, $H = L^2(\Omega)$ et fournit l'existence d'une unique solution faible si $y_0 \in L^2(\Omega)$ et $f \in L^2(0, T; L^2(\Omega))$.

III.3.2 Contrôle et observation distribués

Étant donnés $y_0 \in L^2(\Omega)$, $f \in L^2(0, T; L^2(\Omega))$ et un contrôle $v \in \mathcal{U}_{ad}$ où \mathcal{U}_{ad} est un convexe fermé (non vide) de $L^2(0, T; L^2(\Omega))$, on considère l'équation de la chaleur

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = f + v & \text{dans } Q, \\ y = 0 & \text{sur } \Sigma, \\ y(0, x) = y_0(x) & \text{dans } \Omega. \end{cases}$$

Problème de contrôle optimal : étant donnée une fonction de référence $z_d \in L^2(0, T; L^2(\Omega))$, on définit la fonction coût J par

$$J(v) = \frac{1}{2} \|y(v) - z_d\|_{L^2(0, T; L^2(\Omega))}^2 + \frac{\alpha}{2} \|v\|_{L^2(0, T; L^2(\Omega))}^2,$$

ainsi que le problème de contrôle optimal

$$\inf_{u \in \mathcal{U}_{ad}} J(u) \quad (\text{III.7})$$

Théorème III.3.3. Contrôle optimal distribué de la chaleur

Le problème (III.7) admet une unique solution u qui est caractérisée par le système de l'équation de la chaleur

$$\begin{cases} \frac{\partial y(u)}{\partial t} - \Delta y(u) = f + u & \text{dans } Q, \\ y(u) = 0 & \text{sur } \Sigma, \\ y(0) = y_0 & \text{dans } \Omega \end{cases} \quad (\text{III.8})$$

de l'équation adjointe rétrograde

$$\begin{cases} -\frac{\partial p(u)}{\partial t} - \Delta p(u) = y(u) - z_d & \text{dans } Q, \\ p(u) = 0 & \text{sur } \Sigma, \\ p(u)(T) = 0 & \text{dans } \Omega \end{cases} \quad (\text{III.9})$$

et de la condition d'optimalité

$$(p(u) + \alpha u, v - u)_{L^2(0,T;L^2(\Omega))} \geq 0, \quad \forall v \in \mathcal{U}_{ad}. \quad (\text{III.10})$$

La preuve repose de façon cruciale sur le résultat suivant :

Proposition III.3.4. Gradient de J

La fonctionnelle J est différentiable sur $L^2(\Omega)$ et son gradient est donné par

$$\nabla J(v) = p + \alpha v,$$

où p désigne l'unique solution du problème adjoint (III.9).

Remarque III.3.5 Extension.

Suivant la même démarche, on peut considérer de nombreux problèmes variants. En particulier le problème de Neumann (voir Exercice).

III.3.3 Algorithme de minimisation pour le problème (III.7)

À partir de la proposition III.2.2 et en particulier du calcul du gradient de J , il est aisé de déduire des résultats précédents un algorithme de minimisation. On désigne par $\Pi_{\mathcal{U}_{ad}}$ la projection sur le sous-ensemble convexe fermé non vide \mathcal{U}_{ad} .

Données: $v^0 \in \mathcal{U}_{ad}$, $\varepsilon > 0$, $\rho > 0$ et $\text{Res} = 1$

tant que $\text{Res} \geq \varepsilon$, **faire**

On détermine y^k et p^k , les solutions respectives des EDP

$$\begin{cases} \frac{\partial y^k}{\partial t} - \Delta y^k = f + v^k & \text{dans } Q, \\ y^k = 0 & \text{sur } \Sigma, \\ y^k(0) = y_0 & \text{dans } \Omega \end{cases}$$

et

$$\begin{cases} -\frac{\partial p^k}{\partial t} - \Delta p^k = y^k - z_d & \text{dans } Q \\ p^k = 0 & \text{sur } \Sigma, \\ p^k(T, \cdot) = 0 & \text{dans } \partial\Omega, \end{cases}$$

On pose $v^{k+1} = \Pi_{\mathcal{U}_{ad}}(v^k - \rho(p^k + \alpha v^k))$ et $\text{Res} = \|v^{k+1} - v^k\|$;

fin

Algorithme 2: Algorithme de résolution du problème (III.7) à l'aide d'une méthode de gradient projeté (à pas constant)

III.3.4 Contrôle frontière et observation distribuée.

Le théorème qui suit est un résultat d'existence pouvant s'obtenir par la méthode de transposition, impliquant le théorème de Riesz-Fréchet. Étant donné que l'établissement de ce type de résultat n'est pas l'objectif premier de ce cours, nous l'admettons pour toute la suite.

Avant d'énoncer le théorème d'existence, il convient de préciser la notion de solution faible que nous allons considérer. Pour cela, considérons ϕ , solution du problème adjoint homogène (rétrograde)

$$\begin{cases} \frac{\partial \phi}{\partial t} - \Delta \phi = h & \text{dans } Q \\ \phi = 0 & \text{sur } \Sigma \\ \phi(T, \cdot) = 0 & \text{dans } \Omega. \end{cases} \quad (\text{III.11})$$

Multilplions l'équation principale de (III.13) par ϕ et intégrons par parties. On obtient :

$$\int_0^T \int_{\Omega} y h \, dx \, dt = \int_{\Omega} y_0 \phi(0) \, dx + \int_0^T \int_{\Omega} f \phi \, dx \, dt - \int_0^T \int_{\partial\Omega} v \frac{\partial \phi}{\partial \nu} \, d\Gamma \, dt.$$

Nous dirons alors que la fonction y est une solution faible si elle satisfait

$$\begin{aligned} \|y\|_{L^2(0,T;L^2(\Omega))} &\leq C(\|y_0\|_{L^2(\Omega)} + \|v\|_{L^2(0,T;L^2(\partial\Omega))} + \|f\|_{L^2(0,T;L^2(\Omega))}), \\ \int_0^T \int_{\Omega} y h \, dx \, dt &= \int_{\Omega} y_0 \phi(0) \, dx + \int_0^T \int_{\Omega} f \phi \, dx \, dt - \int_0^T \int_{\partial\Omega} v \frac{\partial \phi}{\partial \nu} \, d\Gamma \, dt \end{aligned} \quad (\text{III.12})$$

pour toute fonction $h \in L^2(\Omega)$, ϕ désignant la solution du problème rétrograde (III.11).

Théorème III.3.6. Équation de la chaleur avec conditions non-homogènes

Soient $y_0 \in L^2(\Omega)$, $u \in L^2(0, T; L^2(\Sigma))$ et $f \in L^2(0, T; L^2(\Omega))$.

Alors l'équation de la chaleur avec conditions aux bords de Dirichlet non-homogènes

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = f & \text{dans } Q \\ y = v & \text{sur } \Sigma \\ y(0, \cdot) = y_0(\cdot) & \text{dans } \Omega. \end{cases} \quad (\text{III.13})$$

admet une unique solution faible $y \in L^2(0, T; H_0^1(\Omega))$. De façon analogue, l'équation de la chaleur avec les conditions aux bords de Neumann non-homogènes.

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = f & \text{dans } Q, \\ \frac{\partial y}{\partial \nu} = v & \text{sur } \Sigma, \\ y(0, \cdot) = y_0(\cdot) & \text{dans } \Omega. \end{cases}$$

admet une unique solution faible telle que $y \in L^2(0, T; H^1(\Omega))$.

Étant donnés $y_0 \in L^2(\Omega)$, $f \in L^2(0, T; L^2(\Omega))$ et un contrôle $v \in \mathcal{U}_{ad} \subset L^2(0, T; L^2(\partial\Omega))$, on considère l'équation de la chaleur avec conditions frontières de Dirichlet non-homogène, soit

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = f & \text{dans } Q, \\ y = v & \text{sur } \Sigma, \\ y(0, \cdot) = y_0(\cdot) & \text{dans } \Omega. \end{cases} \quad (\text{III.14})$$

D'après le théorème III.3.6, le problème (III.14) possède une unique solution faible

$$L^2(0, T; L^2(\partial\Omega)) \ni v \rightarrow y(v) \in L^2(0, T; L^2(\Omega))$$

qui est continue pour les normes correspondantes.

Problème de contrôle optimal : étant donnée une fonction de référence $z_d \in L^2(0, T; L^2(\Omega))$, on définit la fonction coût J par

$$J(v) = \frac{1}{2} \|y(v) - z_d\|_{L^2(0, T; L^2(\Omega))}^2 + \frac{\alpha}{2} \|v\|_{L^2(0, T; L^2(\partial\Omega))}^2.$$

ainsi que le problème de contrôle optimal

$$\inf_{v \in \mathcal{U}_{ad}} J(v). \quad (\text{III.15})$$

Théorème III.3.7.

Le problème (III.15) admet une unique solution u qui est caractérisée par le système de l'équation de la chaleur

$$\begin{cases} \frac{\partial y(u)}{\partial t} - \Delta y(u) = f & \text{dans } Q, \\ y(u) = u & \text{sur } \Sigma, \\ y(0) = y_0 & \text{dans } \Omega, \end{cases}$$

de l'équation adjointe rétrograde

$$\begin{cases} -\frac{\partial p(u)}{\partial t} - \Delta p(u) = y(u) - z_d & \text{dans } Q, \\ p(u) = 0 & \text{sur } \Sigma, \\ p(u)(T) = 0 & \text{dans } \Omega, \end{cases}$$

et de la condition d'optimalité

$$(\alpha u - p_v(u) + \alpha u, v - u)_{L^2(0,T;L^2(\partial\Omega))} \geq 0, \quad \forall v \in \mathcal{U}_{ad}. \quad (\text{III.16})$$

Cas sans contrainte. Lorsque $\mathcal{U}_{ad} = L^2(0, T; L^2(\partial\Omega))$, l'inéquation (III.16) implique

$$u = -\frac{1}{\alpha} \frac{\partial p(u)}{\partial \nu} \quad \text{sur } \Sigma.$$

Le système d'optimalité du théorème III.3.7 devient

$$\begin{cases} \frac{\partial y(u)}{\partial t} - \Delta y(u) = f & \text{dans } Q, \\ -\frac{\partial p(u)}{\partial t} - \Delta p(u) - y(u) = -z_d & \text{dans } Q, \\ y(u) = -\frac{1}{\alpha} \frac{\partial p(u)}{\partial \nu}, \quad p(u) = 0 & \text{sur } \Sigma, \\ y(0) = y_0, \quad p(u)(T) = 0 & \text{dans } \Omega. \end{cases}$$

Remarque III.3.8 Vers un algorithme de minimisation

En notant que le gradient de J (qui est un élément de $L^2((0,T;L^2(\partial\Omega)))$) est, dans ce cas, donné par

$$\nabla J(v) = \alpha v - \partial_\nu p,$$

il est aisé d'adapter l'approche développée dans la section III.3.3 pour obtenir un algorithme de minimisation par une méthode de gradient projeté.

III.4 Exercices du chapitre

III.1 (contrôle optimal sans contrainte) Soit Ω , un ouvert connexe borné de bord de bord C^2 . On considère le problème elliptique de Dirichlet

$$\begin{cases} -\Delta y + qy = v, & \text{dans } \Omega, \\ y = 0 & \text{sur } \partial\Omega \end{cases}$$

où q est une fonction bornée à valeurs positives donnée et $v \in L^2(\Omega)$.

- (i) Montrer que le problème ci-dessus possède une solution unique $y(v)$ et préciser sa régularité.
- (ii) Expliquer pourquoi l'application $L^2(\Omega) \ni v \mapsto y(v) \in H$ (où H est un espace de Hilbert à préciser) est différentiable. En déduire le gradient de l'application

$$G : L^2(\Omega) \ni v \mapsto \frac{1}{2} \int_{\Omega} |\nabla y(v)|^2 dx + \frac{1}{2} \int_{\Omega} (y(v) - 1)^2 dx.$$

- (iii) Etudier l'existence de solutions pour le problème

$$\inf \{ G(v) + \frac{\alpha}{2} \|v\|_{L^2(\Omega)}^2, v \in L^2(\Omega) \},$$

où $\alpha > 0$ est fixé.

On pourra par exemple utiliser la méthode directe du calcul des variations.

- (iv) Énoncer une condition nécessaire et suffisante d'optimalité pour ce problème.

III.2 (*contrôle frontière d'un problème de Neumann sous contrainte*) Soit Ω , un ouvert connexe borné de \mathbb{R}^n dont la frontière est supposée régulière et soit $f \in L^2(\Omega)$. On considère

$$\mathcal{U}_{ad} = \{v \in L^2(\partial\Omega) \mid v \geq 0 \text{ p.p. sur } \partial\Omega\}.$$

Le but de cet exercice est de résoudre le problème

$$\inf_{v \in \mathcal{U}_{ad}} J(v) \quad \text{où} \quad J(v) = \frac{1}{2} \|y(v) - z_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|v\|_{L^2(\partial\Omega)}^2$$

et $y(v)$ désigne la solution du problème de Neumann

$$\begin{cases} y(v) - \Delta y(v) = f & \text{dans } \Omega, \\ \partial_\nu y(v) = v & \text{sur } \partial\Omega \end{cases}$$

- (i) On se donne $u \in L^2(\partial\Omega)$. Montrer que le problème de Neumann admet une unique solution dans $H^1(\Omega)$ et qu'il existe $C > 0$ tel que

$$\forall u \in \mathcal{U}_{ad}, \quad \|y(u)\|_{H^1(\Omega)} \leq C(\|u\|_{L^2(\partial\Omega)} + \|f\|_{L^2(\Omega)}).$$

En déduire que $y'(u)(v-u) = y(v) - y(u)$, et $(\nabla J(u), v-u) = (y(u) - z_d, y(v) - y(u))_{L^2(\Omega)} + \alpha(u, v-u)_{L^2(\partial\Omega)}$.

- (ii) Montrer que J est α -convexe. En déduire qu'il existe un unique contrôle optimal $u \in \mathcal{U}_{ad}$.
- (iii) On introduit l'état adjoint $p(u)$ défini par

$$\begin{cases} p(u) - \Delta p(u) = y(u) - z_d & \text{dans } \Omega, \\ \partial_\nu p(u) = 0 & \text{sur } \partial\Omega. \end{cases}$$

Montrer que u est le contrôle optimal si et seulement si

$$(p(u) + \alpha u, v - u) \geq 0 \quad \forall v \in \mathcal{U}_{ad}.$$

- (iv) Proposer un algorithme numérique de résolution de ce problème.

- III.3** (*contrôle interne de l'équation de la chaleur*) Dans cet exercice, on considère un problème de contrôle optimal dont la fonction coût fait intervenir l'état final comme une pénalité, ce qui oblige l'état final à s'approcher "au mieux" l'état nul. Soit Ω , un ouvert connexe borné de \mathbb{R}^n de bord régulier, et posons $Q =]0, T[\times \Omega$, $\Sigma =]0, T[\times \partial\Omega$. Étant donnés $y_0 \in L^2(\Omega)$, $f \in L^2(0, T; L^2(\Omega))$ et un contrôle $v \in L^2(0, T; L^2(\Omega))$, on considère y , solution de l'équation de la chaleur

$$\begin{cases} \frac{\partial y}{\partial t} - \Delta y = f + v & \text{dans } Q, \\ y = 0 & \text{sur } \Sigma, \\ y(0, \cdot) = y_0(\cdot) & \text{dans } \Omega. \end{cases}$$

On se donne $\alpha > 0$ ainsi qu'une fonction de référence de l'état final $z_d \in L^2(\Omega)$. On définit la fonction coût

$$J : L^2(0, T; L^2(\Omega)) \ni v \mapsto \frac{1}{2} \|y(u)(T) - z_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|v\|_{L^2(0, T; L^2(\Omega))}^2,$$

où $y(u) \in C^0(0, T; L^2(\Omega)) \cap L^2(0, T; H_0^1(\Omega))$ désigne l'unique solution du problème ci-dessus. On cherche un contrôle optimal u qui minimise J sur un ensemble convexe fermé $\mathcal{U}_{ad} \subset L^2(0, T; L^2(\Omega))$.

- (i) Montrer que l'application affine

$$L^2(0, T; L^2(\Omega)) \ni v \mapsto y(v)(T) \in L^2(\Omega)$$

est continue pour les normes correspondantes.

- (ii) Montrer que J est α -convexe sur $L^2(0, T; L^2(\Omega))$. Que peut-on en déduire que le problème $\inf_{v \in \mathcal{U}_{ad}} J(v)$? Écrire l'inéquation d'Euler associée à ce problème.
- (iii) On définit l'état adjoint $p(u)$ comme la solution de l'équation

$$\begin{cases} -\frac{\partial p(u)}{\partial t} - \Delta p(u) = 0 & \text{dans } Q, \\ p(u) = 0 & \text{sur } \Sigma, \\ p(u)(T) = y(u)(T) - z_d & \text{dans } \Omega. \end{cases}$$

Montrer que

$$(p(u) + \alpha u, v - u)_{L^2(0, T; L^2(\Omega))} \geq 0 \quad \forall v \in \mathcal{U}_{ad}.$$

- (iv) Interpréter la condition d'optimalité dans le cas sans contrainte $\mathcal{U}_{ad} = L^2(0, T; L^2(\Omega))$
- (v) Proposer un algorithme numérique de résolution dans le cas de contraintes locales

$$\mathcal{U}_{ad} = \{v \in L^2(0, T; L^2(\Omega)) \mid a \leq v \leq b \text{ p.p. dans } Q\}.$$

L'exercice suivant est optionnel et est destiné à vous entraîner. En guise de correction, vous pourrez vous reporter au document sur la méthode HUM que vous trouverez dans la rubrique "documents divers", sur la page web du cours.

- III.4** (*la méthode HUM*) Considérons deux entiers positifs non nuls n et m , $A \in \mathbb{M}_n(\mathbb{R})$, $B \in \mathbb{M}_{n, m}(\mathbb{R})$. On considère le système linéaire $\dot{x} = Ax + Bu$ où x est l'état du système et u le

contrôle. On suppose que la paire (A, B) satisfait la condition de Kalman. Soit x_0, x_1 deux éléments de \mathbb{R}^n , H , l'espace (affine) défini par

$$H = \{(u, x) \in L^2(0, T) \times H^1(0, T) \mid (x(0), x(T)) = (x_0, x_1) \text{ où } x \text{ résout } \dot{x} = Ax + Bu\}$$

et J_ε , la fonctionnelle définie pour $\varepsilon > 0$ par

$$J_\varepsilon : H \ni (u, x) \mapsto \frac{1}{2} \int_0^T \|u(t)\|_{\mathbb{R}^m}^2 dt + \frac{1}{2\varepsilon} \int_0^T \|\dot{x}(t) - Ax(t) - Bu(t)\|_{\mathbb{R}^n}^2 dt$$

(i) Montrer que J_ε est strictement convexe, coercive et continue.

(ii) Montrer que le problème

$$\inf_{(u, x) \in H} J(u, x) \quad (\mathcal{P}_\varepsilon)$$

possède une unique solution.

(iii) Soit $(u_\varepsilon, x_\varepsilon)$, la solution de $(\mathcal{P}_\varepsilon)$.

(a) Calculer la différentielle $\langle dJ_\varepsilon(u_\varepsilon, x_\varepsilon), (\delta u, \delta x) \rangle$ de J_ε en $(u_\varepsilon, x_\varepsilon)$ dans la direction $(\delta u, \delta x) \in L^2(0, T) \times H_0^1(0, T)$. En posant $p_\varepsilon = \frac{1}{\varepsilon}(\dot{x}_\varepsilon - Ax_\varepsilon - Bu_\varepsilon)$, montrer que l'on a

$$\langle dJ_\varepsilon(u_\varepsilon, x_\varepsilon), (\delta u, \delta x) \rangle = \int_0^T \langle \delta x, \dot{p}_\varepsilon + A^\top p_\varepsilon \rangle + \int_0^T \langle \delta u, u_\varepsilon - B^\top p_\varepsilon \rangle$$

(b) Résoudre alors le problème $(\mathcal{P}_\varepsilon)$.

La méthode HUM consiste ensuite à faire tendre ε vers 0 et à étudier le problème limite.

Rappels sur l'exponentielle d'endomorphismes

A.1 Définition

Soit $(E, \|\cdot\|)$, un \mathbb{C} espace de Banach.

Définition A.1.1. Exponentielle d'endomorphisme

Soit $u \in E$. La série $\left(\sum \frac{u^n}{n!}\right)$ est absolument convergente donc convergente.
Sa somme est notée $\exp u$ ou e^u .

Proposition A.1.2.

Soient $(u, v) \in E^2$.

- $\|\exp u\| \leq \exp(\|u\|)$.
- Si u et v commutent dans E , alors $\exp(u + v) = \exp u \cdot \exp v$.
- Si $u \in E$, alors $\exp u$ est inversible et a pour inverse $\exp(-u)$.
- L'application $\Xi: \begin{matrix} E & \longrightarrow & E \\ u & \longmapsto & \exp u \end{matrix}$ est une application linéaire continue.

Remarque A.1.3

Si $u \in \mathcal{L}(E)$, avec E de dimension finie, alors en appelant \mathcal{B} , une base de E et M sa matrice dans la base \mathcal{B} , on montre que la matrice de l'endomorphisme $\exp u$ dans \mathcal{B} est $\exp M = \sum_{k=0}^{+\infty} \frac{M^k}{k!}$.

A.2 Méthodes pratiques de calcul d'exponentielles de matrices

Rappelons la **décomposition de Dunford**.

Définition A.2.1. Matrice nilpotente

Une matrice $N \in \mathbb{M}_n(\mathbb{C})$ est dite nilpotente s'il existe un entier naturel k tel que $N^k = 0$.

Théorème A.2.2. Décomposition de Dunford

Soit $M \in \mathbb{M}_n(\mathbb{C})$. Alors, il existe une matrice D , **diagonalisable** dans \mathbb{C} et une matrice N , **nilpotente**, telles que

$$M = D + N \quad \text{et} \quad DN = ND.$$

Par ailleurs cette décomposition est unique.

Si la décomposition de Dunford est connue. Alors, on peut écrire que : $e^M = e^{D+N} = e^D e^N$, car D et N commutent. Or, D étant diagonalisable, il existe $P \in GL_n(\mathbb{C})$ telle que $D = P^{-1} D' P$, où $D' = \text{diag}([\lambda_1, \dots, \lambda_n])$ avec $(\lambda_1, \dots, \lambda_n) \in \mathbb{C}^n$. Remarquons que, si $n \in \mathbb{N}^*$, on a

$$D^n = (P^{-1} D' P)^n = P^{-1} D'^n P.$$

On en déduit que :

$$\begin{aligned} \exp D &= \sum_{n=0}^{+\infty} \frac{P^{-1} D'^n P}{n!} = P^{-1} \left(\sum_{n=0}^{+\infty} \frac{D'^n}{n!} \right) P = P^{-1} \exp D' P \\ &= P^{-1} \text{diag}([e^{\lambda_1}, \dots, e^{\lambda_n}]) P. \end{aligned}$$

Pour calculer e^N , on utilise le fait que N est nilpotente. Ainsi, il existe $p > 0$ (indice de nilpotence) tel que $N^p = 0$ et il est immédiat que $N^q = 0$ pour tout $q \geq p$. Ainsi, on aura :

$$e^N = \sum_{n=0}^{+\infty} \frac{N^n}{n!} = \sum_{n=0}^{p-1} \frac{N^n}{n!}.$$

A.3 Systèmes différentiels linéaires à coefficients constants

Théorème A.3.1. Résolution d'un système homogène

Soit $A \in \mathbb{M}_n(\mathbb{C})$. Le système linéaire homogène

$$\begin{cases} \frac{dZ}{dt}(t) = AZ(t) & t \in \mathbb{R}_+ \\ Z(0) = Z_0 \in \mathbb{R}^n \end{cases} \quad (\mathcal{H})$$

a pour unique solution la fonction \hat{Z} définie sur \mathbb{R}_+ par $\hat{Z}(t) = e^{tA} Z_0$ pour tout $t \in \mathbb{R}_+$.

La démonstration repose sur le lemme suivant, attestant que l'on peut dériver terme à terme par rapport à t la série entière définissant e^{tA} .

Lemme A.3.2.

On confond le vecteur e^{tA} avec la fonction $t \mapsto e^{tA}$, par commodité de notations. Alors, on a

$$\frac{d}{dt}(e^{tA}) = \sum_{n=1}^{+\infty} n \frac{t^{n-1} A^n}{n!} = A \left(\sum_{n=1}^{+\infty} \frac{t^{n-1} A^{n-1}}{(n-1)!} \right) = A e^{tA} = e^{tA} A.$$

Passons à présent à la résolution générale d'un système différentiel avec second membre.

Théorème A.3.3. Formule de Duhamel

Soient I un sous-intervalle ouvert de \mathbb{R} , $t_0 \in I$, $A \in \mathbb{M}_n(\mathbb{C})$ et $b : I \subset \mathbb{R} \rightarrow \mathbb{R}^n$, une fonction continue. Le système

$$\begin{cases} \frac{dY}{dt} = AY + b \\ Y(t_0) = Y_0 \in \mathbb{R}^n. \end{cases} \quad (\mathcal{S})$$

a pour unique solution la fonction \hat{Y} donnée par la formule de Duhamel

$$\forall t \in I, \quad \hat{Y}(t) = e^{(t-t_0)A} Y_0 + \int_{t_0}^t e^{(t-s)A} b(s) ds.$$

Exemple A.3.4

On cherche à résoudre le système différentiel

$$\begin{cases} \dot{x}_1 = & + 3x_2 - 2x_3 + t \\ \dot{x}_2 = x_1 & + x_3 & 1 \\ \dot{x}_3 = 3x_1 - 5x_2 + 5x_3 \end{cases} \quad (\mathcal{S})$$

avec $(x_1(0), x_2(0), x_3(0)) = (0, 0, 0)$. Posons $X = [x_1, x_2, x_3]$. Écrivons ce système sous la forme $\dot{X} = AX + b$ avec

$$A = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & -1 & 1 \end{pmatrix} \quad \text{et} \quad b : \mathbb{R} \ni t \mapsto \begin{pmatrix} t \\ 1 \\ 0 \end{pmatrix} \in \mathbb{R}^3.$$

En trigonalisant A puis en utilisant la décomposition de Dunford, on montre que

$$\exp(tA) = P^{-1} \begin{pmatrix} e^t & 0 & 0 \\ 0 & e^{2t} & -te^{2t} \\ 0 & 0 & e^{2t} \end{pmatrix} P,$$

$$\text{avec } P^{-1} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & -1 & 1 \end{pmatrix} \text{ et } P = \begin{pmatrix} -1 & 2 & -1 \\ -1 & 3 & -2 \\ 1 & -1 & 1 \end{pmatrix}.$$

D'après la formule de Duhamel, on a

$$X(t) = P^{-1} \left(\int_0^t e^{(t-s)A} b(s) ds \right) P \quad \text{avec} \quad e^{(t-s)A} b(s) = \begin{pmatrix} te^{(t-s)} \\ e^{2(t-s)} \\ 0 \end{pmatrix}.$$

On en déduit après intégration que la solution X du système (\mathcal{S}) est donnée par

$$\forall t \in \mathbb{R}, \quad X(t) = P^{-1} \begin{pmatrix} t(e^t - 1) \\ \frac{1}{2}(e^{2t} - 1) \\ 0 \end{pmatrix} P.$$

Rappels sur la régularité des ouverts et les équations elliptiques

B.1 Espaces de Sobolev

Si Ω désigne un ouvert de \mathbb{R}^d et $p \in [1, +\infty]$, on utilise la notation $W^{1,p}(\Omega)$ pour désigner l'espace dit de Sobolev des fonctions $u \in L^p(\Omega)$ telles qu'il existe f_1, \dots, f_p dans $L^p(\Omega)$ telles que

$$\int_{\Omega} u \frac{\partial \varphi}{\partial x_i} = - \int_{\Omega} f_i \varphi, \quad \forall \varphi \in C_c^\infty(\Omega), \quad \forall i = 1, \dots, d.$$

Il est équipé de la norme définie par

$$\|u\|_{W^{1,p}(\Omega)} = \left(\|u\|_{L^p(\Omega)}^p + \sum_{i=1}^d \left\| \frac{\partial u}{\partial x_i} \right\|_{L^p(\Omega)}^p \right)^{1/p}.$$

On a en particulier $W^{1,2}(\Omega) = H^1(\Omega)$.

L'espace $W_0^{1,p}(\Omega)$ est alors défini comme l'adhérence de $C_c^\infty(\Omega)$ dans $W^{1,p}(\Omega)$ (pour la topologie forte de $W^{1,p}(\Omega)$ bien sûr). On a en particulier $W_0^{1,2}(\Omega) = H_0^1(\Omega)$.

Rappelons que

- $W^{1,p}(\Omega)$ et $W_0^{1,p}(\Omega)$ sont des espaces de Banach réflexifs si $1 < p < +\infty$ et séparables si $1 \leq p < +\infty$.
- $W^{1,p}(\mathbb{R}^d) = W_0^{1,p}(\mathbb{R}^d)$.
- L'application $W_0^{1,p}(\Omega) \ni u \mapsto \sum_{i=1}^d \left\| \frac{\partial u}{\partial x_i} \right\|_{L^p(\Omega)}$ définit une norme équivalente à la norme $\|\cdot\|_{W^{1,p}(\Omega)}$ dans l'espace $W_0^{1,p}(\Omega)$ (voir Section B.4).
- $W^{1,2}(\Omega)$ et $W_0^{1,2}(\Omega)$ sont des espaces de Hilbert pour le choix de produit scalaire

$$\langle u, v \rangle_{W^{1,2}(\Omega)} = \langle u, v \rangle_{L^2(\Omega)} + \sum_{i=1}^d \left\langle \frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i} \right\rangle_{L^2(\Omega)}.$$

- $W^{1,p}(\Omega)$ et $W_0^{1,p}(\Omega)$ ont de bonnes propriétés topologiques : ils sont réflexifs si $1 < p < +\infty$ et séparables si $1 \leq p < +\infty$.

B.2 Régularité des ouverts

Définition B.2.1. Ouvert lipschitzien, de classe \mathcal{C}^k

- (i) On dit qu'un ouvert Ω de \mathbb{R}^d est à bord lipschitzien si, pour tout $x_0 \in \partial\Omega$, il existe un cylindre $K = K' \times]-a, a[$ dans un repère orthonormé local d'origine $x_0 = 0$, avec K' une boule ouverte de \mathbb{R}^{d-1} de rayon r , et une fonction $\varphi : K' \rightarrow]-a, a[$ lipschitzienne telle que $\varphi(0) = 0$ tels que

$$\Omega \cap K = \{(x', y) \in K \mid y > \varphi(x')\} \quad \text{et} \quad \partial\Omega \cap K = \{(x', \varphi(x')) \mid x' \in K'\}.$$

- (ii) On définit un ouvert de classe \mathcal{C}^k en remplaçant le mot "lipschitzien" par \mathcal{C}^k dans la définition précédente.

B.3 Intégration par parties, formule de Green

Rappelons la formule d'intégration par parties en dimension d .

Théorème B.3.1. Intégration par parties

Soit Ω , un ouvert de \mathbb{R}^d de classe \mathcal{C}^1 par morceaux. Si u et v sont deux fonctions de $H^1(\Omega)$, on a

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v \, dx = - \int_{\Omega} \frac{\partial v}{\partial x_i} u \, dx + \int_{\partial\Omega} u v n_i \, d\sigma,$$

où n_i désigne la i -ème composante de la normale sortante au domaine Ω .

On notera que la formule d'intégration par parties utilise le fait que l'application trace $\gamma : \mathcal{D}(\overline{\Omega}) \ni u \mapsto u|_{\partial\Omega} \in \mathcal{C}^0(\Omega)$ se prolonge par continuité en une application linéaire de $H^1(\Omega)$ dans $L^2(\partial\Omega)$.

À partir de la formule d'intégration par parties, on déduit très facilement la formule de Green.

Théorème B.3.2. Formule de Green

Soit Ω , un ouvert de \mathbb{R}^d de classe \mathcal{C}^1 par morceaux. Soit $u \in H^2(\Omega)$ et $v \in H^1(\Omega)$. On a

$$\int_{\Omega} \Delta u v \, dx = - \int_{\Omega} \nabla u \cdot \nabla v \, dx + \int_{\partial\Omega} \frac{\partial u}{\partial n} v \, d\sigma,$$

où n désigne la normale sortante au domaine Ω .

B.4 Théorème de Rellich et inégalité de Poincaré

Énonçons de prime abord un résultat fondamental d'analyse fonctionnelle, sur la compacité de l'injection¹ des espaces de Sobolev dans les espaces L^p .

1. Rappelons que l'injection compacte de A dans B signifie que la boule unité de A vue dans B est relativement compacte.

Théorème B.4.1. Rellich-Kondrachov

Soit Ω , un ouvert borné quelconque de \mathbb{R}^d . Alors,

- si $p < d$, alors $W_0^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ avec injection compacte pour tout $q \in [1, p^*]$ avec $\frac{1}{p^*} + \frac{1}{d} = \frac{1}{p}$.
- si $p = d$, alors $W_0^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ avec injection compacte pour tout $q \in [p, +\infty[$.
- si $p > d$, alors $W_0^{1,p}(\Omega) \hookrightarrow \mathcal{C}(\overline{\Omega})$ avec injection compacte pour tout $q \in [1, +\infty[$.

De plus, si $\partial\Omega$ est de classe \mathcal{C}^1 , alors les injections précédentes restent vraies en remplaçant $W_0^{1,p}(\Omega)$ par $W^{1,p}(\Omega)$.

On trouvera une preuve de ce théorème dans de nombreux recueils d'analyse fonctionnelle, par exemple [1].

Corollaire B.4.2. Inégalité de Poincaré

Soit $p \in [1, +\infty[$. On suppose que Ω est un ouvert borné dans une direction de \mathbb{R}^d . Alors, il existe une constante $C_{d,\Omega}$ telle que

$$\|u\|_{L^p} \leq C \|\nabla u\|_{L^p}, \quad \forall u \in W_0^{1,p}(\Omega). \quad (\text{B.1})$$

De plus, dans le cas $p = 2$, la plus petite constante vérifiant cette inégalité est $C = \frac{1}{\lambda_1(\Omega)}$, où $\lambda_1(\Omega)$ est la plus petite valeur propre de l'opérateur de Laplace-Dirichlet.

Remarque B.4.3 Généralisation

L'inégalité de Poincaré énoncée dans le corollaire B.4.2 est en réalité valable si Ω est de mesure finie, ou encore si Ω est borné dans une direction. En revanche, la technique de preuve utilisée n'est plus la même.

Énonçons enfin une inégalité de Poincaré plus générale, car valable pour toute fonction de $H^1(\Omega)$, en imposant des conditions un peu plus restrictives sur Ω . La preuve du résultat qui suit est laissée en exercice au lecteur.

Proposition B.4.4. Inégalité de Poincaré-Wirtinger

Soit Ω , un ouvert connexe borné et Lipschitzien. Alors, il existe une constante $C_\Omega > 0$ telle que

$$\forall u \in H^1(\Omega), \quad \|u - \bar{u}\|_{L^2(\Omega)} \leq C_\Omega \|\nabla u\|_{L^2(\Omega)} \quad \text{avec} \quad \bar{u} = \frac{1}{|\Omega|} \int_{\Omega} u(x) dx.$$

ou de façon équivalente

$$\forall u \in L^2(\Omega), \quad \|u - \bar{u}\|_{H^{-1}(\Omega)} \leq C_\Omega \|\nabla u\|_{H^{-1}(\Omega)} \quad \text{avec} \quad \bar{u} = \frac{1}{|\Omega|} \int_{\Omega} u(x) dx.$$

B.5 Théorème de Lax-Milgram

L'existence et l'unicité des solutions d'équations aux dérivées partielles linéaires elliptiques peut, dans beaucoup de cas standards, être déduit du résultat abstrait suivant.

Théorème B.5.1. Théorème de Lax-Milgram

Soit $a(\cdot, \cdot)$, une forme bilinéaire continue (autrement dit, il existe $C > 0$ telle que $|a(u, v)| \leq C \|u\|_V \|v\|_V$ pour tout $(u, v) \in V^2$) et elliptique (autrement dit, il existe $c > 0$ telle que $a(u, u) \geq c \|u\|_V^2$ pour tout $u \in V$) sur un espace de Hilbert V . Soit F , une forme linéaire continue sur V . Alors, il existe un unique élément u de V tel que

$$a(u, v) = F(v) \quad \text{pour tout } v \in V. \quad (\text{B.2})$$

De plus, si a est symétrique, la solution u du problème ci-dessus est l'unique solution du problème

$$\inf_{v \in V} J(v) \quad \text{avec } J(v) = \frac{1}{2} a(v, v) - F(v). \quad (\text{B.3})$$

Exemple B.5.2 Équation de Dirichlet-Poisson

Soit Ω , un ouvert connexe borné et $f \in H^{-1}(\Omega)$. Intéressons-nous à l'existence et l'unicité de solutions pour le problème

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega. \end{cases}$$

Le lien entre le caractère "bien posé" de ce problème et le théorème de Lax-Milgram s'établit à l'aide de la remarque suivante : trouver $u \in H_0^1(\Omega)$ tel que $-\Delta u = f$ dans $\mathcal{D}'(\Omega)$ est équivalent à trouver $u \in H_0^1(\Omega)$ tel que

$$\forall v \in H_0^1(\Omega), \quad \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx. \quad (\text{B.4})$$

En effet, si $u \in H_0^1(\Omega)$ tel que $-\Delta u = f$ dans $\mathcal{D}'(\Omega)$, alors en multipliant l'équation par une fonction test $\varphi \in \mathcal{D}(\Omega)$ et en utilisant la formule de Green (Théorème B.3.2), on montre que pour toute fonction $\varphi \in \mathcal{D}(\Omega)$,

$$\int_{\Omega} \nabla u \cdot \nabla \varphi \, dx = \int_{\Omega} f \varphi \, dx.$$

Il suffit alors d'utiliser la densité de $\mathcal{D}(\Omega)$ dans $H_0^1(\Omega)$ pour conclure. Le sens réciproque se démontre exactement de la même façon.

À présent, montrons l'existence d'une solution pour le problème (B.4) écrit sous forme variationnelle. Soient u et v dans $H_0^1(\Omega)$. On introduit les notations

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx \quad \text{et} \quad F(v) = \int_{\Omega} f v \, dx.$$

D'après l'inégalité de Cauchy-Schwarz, on a

$$|a(u, v)| \leq \|u\|_{H_0^1(\Omega)} \|v\|_{H_0^1(\Omega)} \quad \text{et} \quad |F(v)| \leq \|f\|_{H^{-1}(\Omega)} \|v\|_{H_0^1(\Omega)}$$

ce qui montre que a et F sont continues, et

$$a(u, u) = \int_{\Omega} |\nabla u|^2 \, dx \geq C_{\Omega} \|u\|_{H_0^1(\Omega)}^2,$$

d'après l'inégalité de Poincaré (voir le corollaire B.4.2), ce qui montre que a est elliptique. On applique donc le théorème de Lax-Milgram et la conclusion s'ensuit.

B.6 Régularité elliptique

Sous réserve que le domaine Ω et les coefficients de l'équation aux dérivées partielles elliptique soient suffisamment réguliers, la solution existe en un sens plus fort que celui fourni par le théorème de Lax-Milgram.

Théorème B.6.1. Régularité elliptique

Soit $k \in \mathbb{N}$, Ω , un ouvert de classe $\mathcal{C}^{k+2}(\Omega)$, $a_{i,j} \in \mathcal{C}^{k,1}(\overline{\Omega})$ et $c \in \mathcal{C}^{k-1,1}(\overline{\Omega})$. Soit $f \in H^k(\Omega)$ et $g \in H^{k+2}(\Omega)$. Soit $u \in H^1(\Omega)$, une fonction telle que

$$Lu = f \text{ au sens de } \mathcal{D}'(\Omega); \quad u - g \in H_0^1(\Omega),$$

avec $Lu = -\operatorname{div}(A\nabla u) + cu$ et $A = (a_{i,j})_{1 \leq i,j \leq d}$.

Alors, $u \in H^{k+2}(\Omega)$ et on a l'estimation

$$\|u\|_{H^{k+2}(\Omega)} \leq C_k(\|f\|_{H^k(\Omega)} + \|g\|_{H^{k+2}(\Omega)}),$$

où C_k ne dépend pas de f et g .

Remarque B.6.2 |

Il est notable que la même conclusion peut être obtenue pour des problèmes de Neumann homogènes. On renvoie par exemple à [1, Theorem 9.26].

Application. Soit Ω , un ouvert connexe borné et $f \in H^{-1}(\Omega)$. Le théorème de Lax-Milgram assure que l'unique solution de l'équation

$$\begin{cases} -\Delta u = f & \text{dans } \Omega \\ u \in H_0^1(\Omega) \end{cases}$$

posée au sens faible, est dans $H_0^1(\Omega)$. D'après le théorème de régularité elliptique, si Ω est de classe \mathcal{C}^2 et si $f \in L^2(\Omega)$, alors la solution u du système ci-dessus appartient à $H^2(\Omega)$. En particulier, l'égalité $-\Delta u = f$ est satisfaite presque partout dans Ω .

Concluons cette partie par un résultat fondamental sur les équations elliptiques, appelé *principe du maximum*.

Théorème B.6.3. Principe du maximum

Soit Ω , un ouvert borné de \mathbb{R}^N . On se donne une matrice symétrique $A \in S_N(\mathbb{R})$ dont les composantes $a_{i,j}$ appartiennent à $\mathcal{C}^0(\overline{\Omega})$ et telle qu'il existe $\lambda > 0$ avec $\sum_{1 \leq i,j \leq N} a_{i,j}(x) \xi_i \xi_j \geq \lambda |\xi|^2$ pour tout $x \in \overline{\Omega}$ et tout $\xi \in \mathbb{R}^N$. Soit $b \in \mathcal{C}^0(\overline{\Omega}, \mathbb{R}^N)$ et $c \in \mathcal{C}^0(\overline{\Omega})$ telle que $c \geq 0$ dans Ω . Toute fonction $u \in \mathcal{C}^0(\overline{\Omega}) \cap \mathcal{C}^2(\Omega)$ qui satisfait

$$-\sum_{1 \leq i,j \leq N} a_{i,j}(x) \frac{\partial^2 u_{i,j}}{\partial x_i \partial x_j}(x) + \sum_{i=1}^N b_i(x) \frac{\partial u_i}{\partial x_i}(x) + c(x)u(x) \geq 0 \text{ dans } \Omega$$

et $u(x) \geq 0$ sur $\partial\Omega$

est positive ou nulle dans $\overline{\Omega}$.

Remarque B.6.4 Principe du maximum de Hopf

Si de plus Ω satisfait la condition de sphère intérieure (en tout point x de $\partial\Omega$, il existe une boule ouverte $B(y,R)$ incluse dans Ω telle que $x \in \overline{B}(y,R)$) et si u atteint un minimum local strict en un point x_0 de $\partial\Omega$ dans le cas où $c = 0$, ou bien un minimum local strictement négatif dans le cas où $c \geq 0$, alors $\frac{\partial u}{\partial n}(x_0) < 0$.

Exemple B.6.5 Fonctions surharmoniques

Soit Ω , un ouvert borné de classe \mathcal{C}^3 , et u telle que $-\Delta u = f$ et $u \in H_0^1(\Omega)$, avec $f \in H^1(\Omega)$ telle que $f \geq 0$. Alors, u est positive dans Ω .

Optimisation dans les espaces de Hilbert, topologie faible

C.1 Topologie faible dans les espaces de Hilbert

Soit H un espace de Hilbert de produit scalaire noté $\langle \cdot, \cdot \rangle_H$.

Définition C.1.1. convergence faible

On dit qu'une suite $(x_n)_{n \in \mathbb{N}}$ de H converge faiblement vers $x \in H$ et on note $x_n \rightharpoonup x$ si

$$\lim_{n \rightarrow +\infty} \langle x_n, y \rangle_H = \langle x, y \rangle_H \quad \forall y \in H.$$

Il est aisé de monter (à l'aide de l'inégalité de Cauchy-Schwarz) que si $(x_n)_{n \in \mathbb{N}}$ converge fortement vers x dans H , alors elle converge faiblement. La réciproque est fautive bien sûr.

Exemple C.1.2 cas de L^2

Dans $L^2([0,1])$, la suite de fonctions $(x \mapsto e^{inx})_{n \in \mathbb{N}}$ converge faiblement vers 0, autrement dit

$$\forall \varphi \in L^2([0,1]), \quad \lim_{n \rightarrow +\infty} \int_0^1 \varphi(t) e^{int} dt = 0.$$

La convergence n'est pas forte car $\|e^{inx}\|_{L^2([0,1])} = 1$. Ce résultat s'appelle le lemme de Riemann-Lebesgue.

Les théorèmes qui suivent fournissent des critères de fermeture/compacité assez simples. Il est à noter que les fermés faibles sont des fermés forts, mais la réciproque est fautive a priori. Par exemple, la sphère unité n'est pas faiblement fermée en dimension infinie, son adhérence pour la topologie faible étant la boule fermée. En revanche, les convexes fortement fermés sont faiblement fermés.

Théorème C.1.3. fermeture faible/compacité

- Soit K un sous-ensemble convexe fermé non-vidé de l'espace de Hilbert H . Alors, K est fermé pour la topologie faible, autrement dit si $(x_n)_{n \in \mathbb{N}}$ est une suite d'éléments de K convergeant faiblement vers x dans H , alors $x \in K$.
- Si $(x_n)_{n \in \mathbb{N}}$ est une suite bornée de H , alors on peut en extraire une sous-suite faiblement convergente dans H .

C.2 Semi-continuité inférieure

Il est utile de considérer une notion plus faible que la continuité habituelle des fonctions : la semi-continuité inférieure. Rappelons en préalable que si $(u_n)_{n \in \mathbb{N}}$ est une suite de réels, on définit

$$\liminf_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} \inf_{k \geq n} u_k.$$

Définition C.2.1. semi-continuité inférieure

Soit H , un espace de Hilbert et $f : H \rightarrow \mathbb{R}$. On dit que f est semi-continue inférieurement ou que f est s.c.i. si

$$\forall \alpha \in \mathbb{R}, \{f \leq \alpha\} \text{ est fermé dans } H.$$

On peut montrer que f est s.c.i. si, et seulement si

$$\forall x_0 \in H, \quad \liminf_{x \rightarrow x_0} f(x) \geq f(x_0).$$

On dit que la fonction f est faiblement semi-continue inférieurement lorsque la limite dans la définition de semi-continuité ci-dessus est prise au sens de la topologie faible.

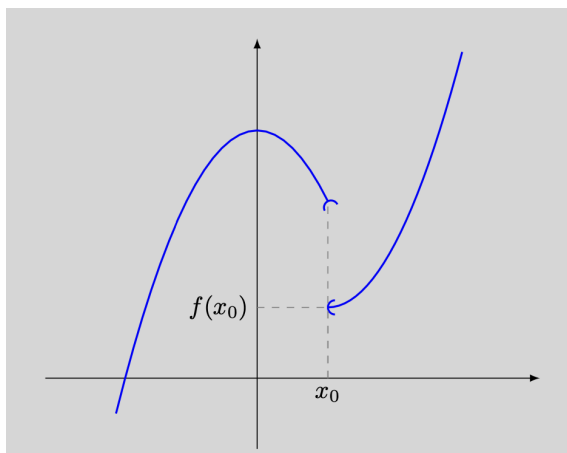


FIGURE C.1 : Un exemple de fonction s.c.i. en dimension 1

Remarque C.2.2

Il est souvent utile d'utiliser le résultat suivant : si $\varphi : H \rightarrow \mathbb{R}$ est convexe et s.c.i. pour la topologie forte, alors φ est également s.c.i. pour la topologie faible. Par exemple, il est clair que l'application $H \ni x \mapsto \|x\|_H$ est continue pour la topologie forte, donc en particulier s.c.i. pour cette topologie. On en déduit qu'elle est également s.c.i. pour la topologie faiblement autrement dit, si $x_n \rightharpoonup x$ dans H , alors

$$\liminf_{n \rightarrow +\infty} \|x_n\| \geq \|x\|.$$

Proposition C.2.3.

Soit K , un compact d'un espace de Hilbert H pour la topologie faible. Soit $J : K \rightarrow \mathbb{R}$, une application s.c.i. pour la topologie faible. alors le problème

$$\inf_{x \in K} J(x)$$

possède une solution.

Démonstration. Soit $(x_n)_{n \in \mathbb{N}}$, une suite minimisante d'éléments de K . Par compacité de K , il existe une suite extraite $(x_{\varphi(n)})_{n \in \mathbb{N}}$ qui converge faiblement vers un élément $x \in K$. En utilisant le fait que J est s.c.i. pour cette topologie, on en déduit :

$$J(x) \leq \liminf_{n \rightarrow +\infty} J(x_{\varphi(n)}) = \inf_K J.$$

La conclusion s'ensuit. □

C.3 Optimisation dans un espace de Hilbert

Énonçons à présent quelques résultats d'existence dans les espaces de Hilbert.

Théorème C.3.1. méthode variationnelle

Soit H , un espace de Hilbert, C un convexe fermé non vide de H , et $J : H \rightarrow \mathbb{R}$, une application différentiable et convexe. Si C n'est pas borné, on suppose J coercive, autrement dit $J(x) \rightarrow +\infty$ si $\|x\| \rightarrow +\infty$. Alors il existe $x^* \in C$ tel que $J(x^*) = \inf_{x \in C} J(x)$.

Démonstration. Soit $(x_n)_{n \in \mathbb{N}}$, une suite minimisante d'éléments de C . Notons que $(J(x_n))_{n \in \mathbb{N}}$ est bornée (si C est bornée, c'est évident car J est différentiable donc continue et si J est coercive, il existe une sous-suite $(x_{\varphi(n)})_{n \in \mathbb{N}}$ telle que $\|x_{\varphi(n)}\|_H \rightarrow +\infty$ quand $n \rightarrow +\infty$, ce qui implique que $\inf_C J = +\infty$ par coercivité. C'est absurde). D'après le théorème C.1.3, C est faiblement fermé et quitte à extraire, il existe $x^* \in C$ tel que $x_n \rightharpoonup x^*$. Enfin, puisque J est convexe et différentiable, sa courbe est située au dessus de tout hyperplan tangent, et il vient :

$$J(x_n) \geq J(x^*) + \langle \nabla J(x^*), x_n - x^* \rangle_H \xrightarrow{n \rightarrow +\infty} J(x^*)$$

et ainsi $\inf_C J = J(x^*)$ nécessairement. □

Exemple C.3.2 un classique

Soit Ω , un ouvert connexe borné de \mathbb{R}^n et $\sigma \in L^\infty(\Omega)$ telle que $\sigma(\cdot) \geq \sigma_0 > 0$ p.p. dans Ω . Soit $f \in L^2(\Omega)$. Alors le problème

$$\begin{cases} -\operatorname{div}(\sigma \nabla u) + u = f & \text{dans } \Omega \\ u = 0 & \text{sur } \partial\Omega \end{cases}$$

possède une unique solution dans $H_0^1(\Omega)$.

Pour le prouver, considérons l'espace de Hilbert $H = H_0^1(\Omega)$ et la fonctionnelle $J : H \rightarrow \mathbb{R}$ définie par

$$J(v) = \frac{1}{2} \int_{\Omega} \sigma |\nabla v|^2 + \frac{1}{2} \int_{\Omega} v^2 - \int_{\Omega} f v.$$

La fonctionnelle J est bien définie (exercice) et on montre de plus aisément que

- J est de classe C^∞ car somme d'une forme quadratique continue et d'une forme linéaire continue sur H , en utilisant l'équivalence des normes $\|\cdot\|_{H^1}$ et $\|\cdot\|_{H_0^1}$ dans H et l'inégalité de Cauchy-Schwarz.
- J est strictement convexe (calcul immédiat).
- J est coercive car

$$\forall v \in H, \quad J(v) \geq \frac{1}{2} \min\{\sigma_0, 1\} \|v\|_{H^1}^2 - \|f\|_{L^2} \|v\|_{H^1} \xrightarrow{\|v\|_{H^1} \rightarrow +\infty} +\infty.$$

On en déduit l'existence d'un unique $u \in H$ (l'unicité vient de la stricte convexité de J) tel que $\inf_H J = J(u)$. On en déduit que u est caractérisé de façon unique par la condition d'Euler $\langle dJ(u), v \rangle = 0$ pour tout $v \in H$, autrement dit

$$\forall v \in H, \quad \int_{\Omega} (\sigma \nabla u \cdot \nabla v + uv) = \int_{\Omega} f v.$$

Puisque la formulation variationnelle ci-dessus est l'équation d'Euler du problème $\inf_H J$ et que J est convexe, on sait que tout élément u la satisfaisant est un minimiseur du problème $\inf_H J$. Puisqu'on a par ailleurs montré l'unicité des solutions de ce problème, on conclut.

Le résultat qui suit nécessite moins de régularité que celui que nous venons d'énoncer. Attention au fait que ce résultat n'est pas immédiat car en dimension infinie, il ne suffit pas qu'un ensemble soit fermé et borné pour qu'il soit compact (penser par exemple à la sphère unité, dont l'adhérence pour la topologie faible est la boule unité).

Théorème C.3.3.

Soit H , un espace de Hilbert et $J: H \rightarrow \mathbb{R}$, une fonctionnelle convexe, coercive et s.c.i. pour la topologie forte. Alors, J est minorée sur H et le problème $\inf_H J$ a une solution.

Démonstration. Soit $x_0 \in H$. Par coercivité de J , il existe $\eta > 0$ tel que $\|x\| > \eta \Rightarrow J(x) > J(x_0)$. Par conséquent,

$$\inf_H J = \inf_{\overline{B_H(0, \eta)}} J,$$

où $\overline{B_H(0, \eta)}$ est la boule fermée de centre 0 et rayon η dans H . Soit $(x_n)_{n \in \mathbb{N}}$, une suite minimisante pour le problème $\inf_H J$. D'après le théorème C.1.3, on peut extraire de $(x_n)_{n \in \mathbb{N}}$ une sous-suite $(x_{\varphi(n)})_{n \in \mathbb{N}}$ convergente dans $\overline{B_H(0, \eta)}$ (qui est un fermé faible) pour la topologie faible. Notons x , sa limite. Enfin, J est s.c.i. pour la topologie faible en vertu de la remarque C.2.2 car J est convexe et s.c.i. pour la topologie forte. Finalement, le problème $\inf_{\overline{B_H(0, \eta)}} J$ satisfait les hypothèses de la propriété C.2.3, ce qui permet de conclure. \square

Filtre de Kalman déterministe

Ce célèbre problème consiste à estimer un signal bruité. Plus précisément, considérons un signal x solution du système

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (\text{D.1})$$

où la valeur initiale $x(0)$ est inconnue. Soit $\xi(t)$ une estimation de la variable d'observation

$$y(t) = Cx(t).$$

On cherche à estimer la valeur $x(T)$ à partir de l'observation ξ sans la connaissance de $x(0)$. Il paraît naturel de minimiser la fonction coût

$$J(u) = \frac{1}{2} \int_0^T [(Q(Cx(t) - \xi(t)), Cx(t) - \xi(t)) + (Ru(t), u(t))] dt + \frac{1}{2} (Dx(0), x(0)). \quad (\text{D.2})$$

dans $L^2(0, T)$.

Pour pouvoir appliquer la théorie LQ, on introduit

$$y = \begin{pmatrix} x \\ 1 \end{pmatrix}, \mathcal{A} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}, \mathcal{B} = \begin{pmatrix} B \\ 0 \end{pmatrix}, \mathcal{D} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \mathcal{Q} = \begin{pmatrix} C^\top QC & -C^\top Q\xi \\ -\xi^\top QC & \xi^\top Q\xi \end{pmatrix}.$$

Alors on montre que y est solution du système

$$\dot{y} = \mathcal{A}y + \mathcal{B}u$$

et que la fonctionnelle J se réécrit

$$J(u) = \frac{1}{2} \int_0^T [(\mathcal{Q}y(t), y(t)) + (Ru(t), u(t))] dt + \frac{1}{2} (\mathcal{D}y(0), y(0)).$$

En introduisant les nouvelles inconnues $\tilde{y}(t) = y(T-t)$ et $\tilde{u}(t) = u(T-t)$, on obtient le système rétrograde

$$\dot{\tilde{y}} = -\mathcal{A}\tilde{y} - \mathcal{B}\tilde{u}.$$

et

$$J(\tilde{u}) = \frac{1}{2} \int_0^T [(\mathcal{Q}\tilde{y}(t), \tilde{y}(t)) + (R\tilde{u}(t), \tilde{u}(t))] dt + \frac{1}{2} (\mathcal{D}\tilde{y}(T), \tilde{y}(T)).$$

D'après la théorie LQ, la commande optimale \tilde{u} est donnée par

$$\tilde{u}(t) = -R^{-1}\mathcal{B}^\top \mathcal{E}(t)\tilde{y}(t)$$

où \mathcal{E} est la solution de l'équation de Riccati

$$\begin{cases} \dot{\mathcal{E}}(t) = -\mathcal{Q} + \mathcal{A}^\top \mathcal{E}(t) + \mathcal{E}(t)\mathcal{A} + \mathcal{E}(t)\mathcal{B}R^{-1}\mathcal{B}^\top \mathcal{E}(t) & t \in [0, T] \\ \mathcal{E}(T) = \mathcal{D}. \end{cases}$$

En reportant l'expression

$$\mathcal{E}(t) = \begin{pmatrix} \tilde{E}(t) & \tilde{b}(t) \\ \tilde{b}(t)^\top & \tilde{a}(t) \end{pmatrix},$$

dans l'équation de Riccati, on trouve

$$\begin{cases} \dot{\tilde{E}} = -C^\top QC + A^\top \tilde{E} + \tilde{E}A + \tilde{E}BR^{-1}B^\top \tilde{E}, & \tilde{E}(T) = D, \\ \dot{\tilde{b}} = C^\top Q\tilde{\xi} + A^\top \tilde{b} + \tilde{E}BR^{-1}B^\top \tilde{b}, & \tilde{b}(T) = 0. \end{cases}$$

et

$$\tilde{u}(t) = R^{-1}B^\top \tilde{E}(t)\tilde{x}(t) + R^{-1}B^\top \tilde{b}(t).$$

D'autre part, d'après la remarque II.2.7, le coût minimal vaut

$$J(\tilde{u}) = \frac{1}{2}(\mathcal{E}(0)\tilde{y}(0), \tilde{y}(0)) = \frac{1}{2}(\tilde{E}(0)\tilde{x}(0), \tilde{x}(0)) + (\tilde{b}(0), \tilde{x}(0)) + \tilde{a}(0).$$

On cherche $\tilde{x}(0)$ qui minimise ce minimum (sachant que $\tilde{E}(0)$ est inversible puisque D est définie positive, voir (ii) dans la remarque II.2.4). Ainsi, on a

$$\tilde{x}(0) = -\tilde{E}^{-1}(0)\tilde{b}(0).$$

Maintenant revenons au temps progressif et résumons les résultats obtenus.

Théorème D.0.1. Filtre de Kalman

Il existe une unique trajectoire optimale $x(\cdot)$ du système (D.1)-(D.2) associée à la commande optimale u définie par

$$u(t) = R^{-1}B^\top E(t)x(t) + R^{-1}B^\top b(t) = K(t)x(t) + H(t)$$

et à l'état final

$$x(T) = -E^{-1}(T)b(T)$$

où les variables E, b sont déterminées par

$$\begin{cases} \dot{E} = C^\top QC - A^\top E - EA - EBR^{-1}B^\top E, & E(0) = D, \\ \dot{b} = -C^\top Q\xi - A^\top b - EBR^{-1}B^\top b, & b(0) = 0. \end{cases}$$

Remarque D.0.2

L'état final $x(T)$ dans le théorème ci-dessus est l'objectif principal du filtre de Kalman. Pour simplifier son calcul, nous posons

$$F(t) = E^{-1}(t), \quad z(t) = -F(t)b(t), \quad z(T) = x(T).$$

Alors un calcul direct donne

$$\begin{cases} \dot{F} = BR^{-1}B^\top + AF + FA^\top - FC^\top QCF, & F(0) = D^{-1}, \\ \dot{z} = (A - FC^\top QC)z + FC^\top Q\xi, & z(0) = 0. \end{cases}$$

Exemple D.0.3

On veut estimer $x(T)$ pour le système bruité

$$\dot{x}(t) = u(t)$$

d'après l'observation $\xi(t)$ de $y = x + v$. Construisons un filtre de Kalman avec

$$A = 0, \quad B = 1, \quad C = 1, \quad D = 1, \quad R = 1, \quad Q = q^2.$$

Puis, on résout le système différentiel (3.11) qui devient

$$\begin{cases} \dot{F} = 1 - q^2 F^2, & F(0) = 1, \\ \dot{z} = q^2 F(z - \xi), & z(0) = 0. \end{cases}$$

On trouve

$$F(t) = \frac{1}{q} + \frac{e^{-qt}(q-1)}{q}.$$

Lorsque le bruit ν est très petit, on choisit q assez grand de telle sorte que $qF(t) \sim 1$. On déduit de l'équation que $z(T) \sim \xi(T)$, soit $x(T) \sim \xi(T)$. Ceci est bien cohérent, puisque quand le bruit est petit, l'observation $\xi(T)$ n'est d'autre que l'état $x(T)$ à estimer. Dans le cas général, il faut intégrer numériquement la deuxième équation pour trouver $z(T) = x(T)$.

Bibliographie

- [1] H. Brezis, Analyse fonctionnelle, théorie et applications, Masson, Paris, 1983.
- [2] R. Glowinski, J.-L. Lions, J. He, Exact and approximate controllability for distributed parameter systems, Cambridge University Press, 2008.
- [3] E. B. Lee, L. Markus, Foundations of optimal control theory, John Wiley, New York, 1967.
- [4] L. Pontryagin, V. Boltyanski, R. Gamkrelidze, E. Michtchenko, Théorie mathématique des processus optimaux, Editions Mir, Moscou, 1974.
- [5] E. Trélat, Contrôle optimal : théorie & applications., Vuibert, Collection “Mathématiques Concrètes”, 2005.