Promotion X2017
Année 2
MAP434

Contrôle de modèles dynamiques

Alexandre Ern

version du 28 mai 2019

Édition 2019

Table des matières

Avant-propos

Ce cours est consacré à l'étude des systèmes commandés, c'est-à-dire des systèmes dynamiques sur lesquels on peut agir au moyen d'une commande ou d'un contrôle. Un premier objectif peut être d'amener le système d'un état initial donné à un état final (une cible), en respectant éventuellement certaines contraintes (par exemple, la valeur du contrôle ne peut être trop grande ou bien l'état du système doit respecter certaines contraintes). Il s'agit du problème de la **contrôlabilité**. Un deuxième objectif peut être celui de déterminer un contrôle optimal, c'est-à-dire minimisant un certain critère dépendant du contrôle et de la trajectoire résultant de ce contrôle. Il s'agit du problème de **contrôle optimal**. Nous aborderons ces deux problèmes dans ce cours. Le champ d'applications est très vaste. On rencontre des problèmes de contrôlabilité et de contrôle optimal dans des domaines très variés, comme l'aéronautique, l'électronique, le génie des procédés, la médecine, l'économie et la finance, internet et les communications, etc.

Ce cours se plaçant à un niveau introductif, nous nous restreindrons pour simplifier à des systèmes dynamiques dont l'état peut être décrit par un nombre **fini** de variables. De plus, nous considérerons uniquement des systèmes dépendant du temps et non pas du temps et de l'espace ; en d'autres termes, nous considérerons uniquement le contrôle de **systèmes différentiels** et non pas d'équations aux dérivées partielles. L'horizon temporel pourra être fixé ou non, mais il sera toujours fini. Un exemple important où cet horizon n'est pas fixé est celui de la temps-optimalité consistant à chercher un contrôle permettant d'atteindre une cible (atteignable) en temps minimum. Enfin, nous considérerons uniquement des systèmes **déterministes** et n'aborderons pas ici le cas (très important en pratique) des systèmes stochastiques comme les systèmes avec bruit.

Afin de fixer les idées, donnons un exemple simple de système de contrôle, celui du contrôle d'un aspirateur robot. On note $t \in [0,T]$ le temps où $T > 0$ est l'horizon temporel fixé. L'état du système est décrit par le triplet $(x, y, \theta) : [0,T] \to \mathbb{R}^3$. Le couple (x, y) repère la position de l'aspirateur dans le plan et θ l'angle des roues par rapport à l'axe des x. L'action sur le système s'exerce par le biais d'une fonction $u : [0,T] \to \mathbb{R}$ qui prescrit la vitesse angulaire de l'axe des roues. La dynamique du système est régie par le système différentiel suivant (qu'on appelle système de Dubbins) :

$$\begin{cases} \dot{x}(t) = v\cos(\theta(t)), \\ \dot{y}(t) = v\sin(\theta(t)), \\ \dot{\theta}(t) = u(t), & \leftarrow \text{action sur le système} \end{cases}$$

où v est la vitesse de l'aspirateur, supposée constante pour simplifier. De manière plus générale, nous considérerons des systèmes de contrôle sous la forme

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

où la fonction $x : [0,T] \to \mathbb{R}^d$, $d \geq 1$, décrit l'état du système, $u : [0,T] \to \mathbb{R}^k$, $k \geq 1$, est le contrôle, et $f : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d$ décrit la dynamique du système. En général, une condition initiale $x(0) = x_0 \in \mathbb{R}^d$ est également prescrite.

Ce cours est organisé en trois parties. La première, composée des chapitres 1 et 2, aborde le problème de la contrôlabilité. Le résultat phare est le **critère de Kalman** sur la contrôlabilité des systèmes linéaires autonomes et son extension à la contrôlabilité locale des systèmes non-linéaires. La deuxième partie, composée des chapitres 3 à 6, aborde le problème du contrôle optimal par le biais du **principe du minimum de Pontryaguine** (PMP). Dans les chapitres 3 et 4, nous commencerons par l'étude du système linéaire-quadratique (dit système LQ) qui consiste à minimiser un critère quadratique pour un système de contrôle linéaire. Le système LQ étant particulièrement simple, il nous sera possible de mener une analyse complète du problème. Celle-ci repose sur diverses idées importantes, comme la notion d'état adjoint, de Hamiltonien et de feedback grâce à l'équation de Riccati. Puis, dans les chapitres 5 et 6, nous aborderons le cas général du contrôle optimal de systèmes non-linéaires ; nous énoncerons le PMP, en esquisserons la preuve et en donnerons quelques exemples d'applications. Enfin, la troisième partie, composée des chapitres 7 et 8, est toujours consacrée aux problèmes de contrôle optimal, mais propose de les aborder sous un angle nouveau : celui de la **programmation dynamique**, d'abord en temps discret puis en temps continu. L'idée fondamentale est le **principe d'optimalité de Bellman**, conduisant à l'équation de Hamilton–Jacobi–Bellman.

Ce cours a été initié en 2014 par Pierre-Louis Lions [8], et la version actuelle du cours, même si elle a fait intégralement l'objet d'une nouvelle rédaction, lui doit énormément, tant sur le choix du périmètre conceptuel que sur l'exposition des principales notions mathématiques. Toutefois, la trame actuelle du cours a été revue, surtout pour les premiers chapitres, afin d'une part de faire émerger une première partie sur la contrôlabilité des systèmes linéaires et non-linéaires et d'autre part d'entrelacer la revue des principaux résultats sur l'optimisation dans les espaces de Hilbert (qui est aride mais incontournable !) avec l'étude du système LQ. En outre, plusieurs exemples ont été ajoutés pour illustrer le PMP, tout en insistant un peu moins sur certaines preuves. Par ailleurs, le contenu de ce cours s'est également inspiré, avec grand profit, du cours d'Emmanuel Trélat sur le contrôle optimal dispensé à l'Université Pierre et Marie Curie, et on ne saurait trop recommander la lecture de l'ouvrage [11] (rédigé en français). Le lecteur désireux d'aller encore plus loin pourra par exemple consulter des ouvrages plus spécialisés et exhaustifs (en anglais) comme ceux de Aubin [1], Bardi et Capuzzo-Dolcetta [2], Fletcher [4], Isidori [6], Lee et Markus [7], Rockafellar et Wets [9], Sontag [10] ou Vinter [12].

Alexandre Ern
Paris, janvier 2019

Chapitre 1

Contrôlabilité des systèmes linéaires

Ce chapitre est consacré à la contrôlabilité des systèmes linéaires. Le principal résultat est le **critère de Kalman** qui fournit une condition nécessaire et suffisante pour la contrôlabilité d'un système linéaire autonome. De manière tout à fait remarquable, ce critère se formule de manière purement algébrique, et la condition à vérifier est indépendante de la condition initiale et de l'horizon temporel. Dans un deuxième temps, nous considérons des systèmes de contrôle linéaires avec des bornes sur le contrôle. Cela nous conduit à introduire la notion importante d'**ensemble atteignable**.

1.1 Systèmes de contrôle linéaires

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

$$\dot{x}(t) = Ax(t) + Bu(t), \quad \forall t \in [0,T], \qquad x(0) = x_0 \in \mathbb{R}^d, \tag{1.1}$$

avec des matrices $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times k}$, où $d \geq 1$ et $k \geq 1$. La fonction temporelle

$$u : [0,T] \to \mathbb{R}^k \tag{1.2}$$

nous permet d'agir sur le système afin d'en modifier l'état. On dit que u est le **contrôle**. Une fois le contrôle u fixé, (1.1) est un **problème de Cauchy**. Afin d'expliciter le fait que la trajectoire x, solution de (1.1), dépend du contrôle u, nous la noterons souvent x_u, et nous écrirons (1.1) sous la forme

$$\dot{x}_u(t) = Ax_u(t) + Bu(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0 \in \mathbb{R}^d. \tag{1.3}$$

Par la suite, nous supposerons que

$$u \in L^1([0,T];\mathbb{R}^k), \tag{1.4}$$

et nous serons parfois amenés à faire des hypothèses un peu plus fortes sur le contrôle, comme par exemple que u prend ses valeurs dans un sous-ensemble fermé non-vide de $\mathbb{R}^k$, ce que

nous noterons $u \in L^1([0,T];U)$; nous ferons parfois des hypothèses d'intégrabilité plus forte en temps, comme par exemple $L^2([0,T];U)$ ou $L^\infty([0,T];U)$. Rappelons à toutes fins utiles que l'espace $L^1([0,T];\mathbb{R}^k)$ est équipé de la norme

$$\|u\|_{L^1([0,T];\mathbb{R}^k)} = \int_0^T |u(s)|_{\mathbb{R}^k} \, \mathrm{d}s, \tag{1.5}$$

où $|\cdot|_{\mathbb{R}^k}$ désigne la norme euclidienne sur $\mathbb{R}^k$. (On peut remplacer la norme euclidienne par toute autre norme sur $\mathbb{R}^k$.) Dans ce cours, on utilisera la notation † pour désigner la transposition des vecteurs ou des matrices ; on écrira donc $x^\dagger y$ pour le produit scalaire entre deux vecteurs et $Z^\dagger$ pour la transposée de la matrice Z.

Définition 1.1 (Systèmes de contrôle linéaires)**.** *On dit que* (1.3) *est un* ***système de contrôle linéaire****. On dit que ce système est* ***autonome*** *(ou* ***stationnaire****) lorsque les matrices A et B ne dépendent pas du temps. Plus généralement, on dit que le système de contrôle linéaire est* ***instationnaire*** *lorsqu'il s'écrit sous la forme*

$$\dot{x}_u(t) = A(t)x_u(t) + B(t)u(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0, \tag{1.6}$$

avec $A \in L^1([0,T];\mathbb{R}^{d\times d})$ et $B \in L^1([0,T];\mathbb{R}^{d\times k})$. Enfin, on dit que le système de contrôle linéaire a un ***terme de dérive*** *lorsqu'il s'écrit sous la forme*

$$\dot{x}_u(t) = Ax_u(t) + Bu(t) + f(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0, \tag{1.7}$$

avec $f \in L^1([0,T];\mathbb{R}^d)$, les matrices A et B pouvant ou non dépendre du temps.

Dans cette section, nous allons considérer le système de contrôle linéaire autonome (1.1). La première question à se poser est si, pour tout contrôle $u \in L^1([0,T];\mathbb{R}^k)$ fixé, il existe une unique trajectoire $x : [0,T] \to \mathbb{R}^d$ associée à ce contrôle, solution du problème de Cauchy (1.1). Comme le contrôle u n'est *a priori* pas une fonction continue du temps, on ne peut pas chercher une trajectoire de classe $C^1([0,T];\mathbb{R}^d)$. Un bon cadre fonctionnel pour la trajectoire est celui des fonctions absolument continues sur $[0,T]$, dont on rappelle la définition.

Définition 1.2 (Fonction absolument continue)**.** *On dit qu'une fonction $F : [0,T] \to \mathbb{R}^d$ est absolument continue sur $[0,T]$ et on écrit $F \in AC([0,T];\mathbb{R}^d)$ s'il existe $f \in L^1([0,T];\mathbb{R}^d)$ telle que*

$$F(t) - F(0) = \int_0^t f(s) \, \mathrm{d}s, \quad \forall t \in [0,T]. \tag{1.8}$$

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.

Proposition 1.3 (Formule de Duhamel)**.** *Pour tout contrôle $u \in L^1([0,T];\mathbb{R}^k)$, il existe une unique trajectoire*

$$x_u \in AC([0,T];\mathbb{R}^d) \tag{1.9}$$

solution de (1.1) *au sens où cette trajectoire vérifie la condition initiale* $x_u(0) = 0$ *et le système différentiel* $\dot{x}_u(t) = Ax_u(t)+Bu(t)$ *presque partout (p.p.) sur* $[0,T]$. *Cette trajectoire est donnée par la formule de Duhamel*

$$x_u(t) = e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu(s)\,\mathrm{d}s, \quad \forall t \in [0,T]. \tag{1.10}$$

On notera que cette expression a bien un sens pour $u \in L^1([0,T];\mathbb{R}^k)$ *car la fonction* $s \mapsto e^{(t-s)A}$ *est bornée sur* $[0,T]$.

Remarque 1.4. [Exponentielle de matrice] On rappelle que $e^A = \sum_{n\geq 0} \frac{1}{n!}A^n$, $\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A$, et que si A_1, A_2 commutent ($A_1A_2 - A_2A_1 = 0$), alors $e^{A_1}e^{A_2} = e^{A_2}e^{A_1} = e^{A_1+A_2}$. □

Remarque 1.5. [Fonction dérivable presque partout] Attention, si une fonction $F : [0,T] \to \mathbb{R}$ est continue sur $[0,T]$ et dérivable p.p. sur $[0,T]$, elle peut ne pas être égale à l'intégrale de sa dérivée (même si celle-ci est L^1). Un contre-exemple est fourni par l'escalier de Cantor (ou escalier du diable) illustré à la figure 1.1 ; cette fonction n'est donc pas absolument continue. □

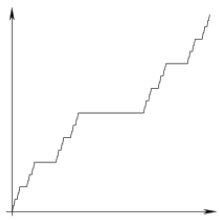

FIGURE 1.1 – L'escalier de Cantor : fonction continue et dérivable presque partout qui n'est pas égale à l'intégrale de sa dérivée.

1.2 Cas sans contraintes : critère de Kalman

On considère le système de contrôle linéaire autonome

$$\dot{x}_u(t) = Ax_u(t) + Bu(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0 \in \mathbb{R}^d. \tag{1.11}$$

Définition 1.6 (Contrôlabilité)**.** *On dit que le système* (1.11) *est contrôlable en temps* T *à partir de* x_0 *si*

$$\forall x_1 \in \mathbb{R}^d, \quad \exists u \in L^\infty([0,T];\mathbb{R}^k), \quad x_u(T) = x_1. \tag{1.12}$$

On cherche donc à atteindre la cible x_1 *au temps* T *à partir de* x_0.

Remarque 1.7. [Intégrabilité] On pourrait aussi chercher $u \in L^1([0,T];\mathbb{R}^k)$. □

En posant $x_2 = x_1 - e^{TA}x_0$, la contrôlabilité en T à partir de x_0 équivaut à

$$\forall x_2 \in \mathbb{R}^d, \quad \exists u \in L^\infty([0,T];\mathbb{R}^k), \quad x_2 = \int_0^T e^{(T-s)A}Bu(s)\,\mathrm{d}s, \tag{1.13}$$

i.e., à la **surjectivité** de l'application

$$\Phi : L^\infty([0,T];\mathbb{R}^k) \to \mathbb{R}^d, \quad \Phi(u) = \int_0^T e^{(T-s)A}Bu(s)\,\mathrm{d}s. \tag{1.14}$$

Un résultat remarquable, dû à Kalman, permet de caractériser la surjectivité de cette application à partir d'une condition **purement algébrique** ne faisant intervenir que les matrices A et B. On introduit la matrice de Kalman $C \in \mathbb{R}^{d\times dk}$ telle que

$$C = \left(B, AB, \cdots, A^{d-1}B\right). \tag{1.15}$$

Théorème 1.8 (Critère de Kalman)**.** *Le système linéaire autonome $\dot{x}_u(t) = Ax_u(t) + Bu(t)$ est contrôlable pour tout $T > 0$ et pour tout $x_0 \in \mathbb{R}^d$* ***si et seulement si***

$$\mathrm{rang}(C) = d, \tag{1.16}$$

ce qui signifie que la matrice C est de rang maximal.

Remarque 1.9. [Condition (1.16)] La condition de Kalman (1.16) est **indépendante** de l'horizon temporel $T > 0$ et de la donnée initiale $x_0 \in \mathbb{R}^d$. La contrôlabilité d'un système linéaire autonome est donc indépendante de ces deux paramètres. Cela signifie en particulier que lorsqu'un système de contrôle linéaire autonome est contrôlable, on peut atteindre à partir d'une donnée initiale toute cible, même très lointaine, en un horizon temporel même très court. Ce n'est pas très surprenant dans la mesure où on ne s'est pas imposé de bornes sur la valeur du contrôle ; celui-ci peut donc prendre des valeurs très grandes si nécessaire. □

Remarque 1.10. [Changement de base] On vérifie facilement que la condition de Kalman est invariante par changement de base. En effet, soit $P \in \mathbb{R}^{d\times d}$ une matrice inversible de changement de base. On considère le système linéaire autonome $\dot{x}(t) = Ax(t) + Bu(t)$. Dans la nouvelle base, ce système s'écrit

$$\dot{y}(t) = \widetilde{A}y(t) + \widetilde{B}u(t),$$

avec $y(t) = P^{-1}x(t)$, $\widetilde{A} = P^{-1}AP$, $\widetilde{B} = P^{-1}B$, si bien que

$$\widetilde{C} = \left(\widetilde{B}, \widetilde{A}\widetilde{B}, \cdots, \widetilde{A}^{d-1}\widetilde{B}\right) = P^{-1}C.$$

Par conséquent, $\mathrm{rang}(C) = \mathrm{rang}(\widetilde{C})$. □

Démonstration. (1) Supposons d'abord que $\mathrm{rang}(C) < d$. Il existe donc un vecteur $\Psi \in \mathbb{R}^d$, $\Psi \neq 0$, tel que

$$\Psi^\dagger B = \Psi^\dagger AB = \cdots = \Psi^\dagger A^{d-1}B = 0\,(\in \mathbb{R}^k),$$

où $\Psi^\dagger$ désigne le transposé de Ψ ($\Psi^\dagger$ est un vecteur ligne). D'après le théorème d'Hamilton–Cayley, il existe des réels $s_0, \cdots, s_{d-1}$ tels que

$$A^d = s_0 I_d + \cdots + s_{d-1}A^{d-1},$$

où I_d est la matrice identité dans $\mathbb{R}^{d\times d}$. On en déduit par récurrence que $\Psi^\dagger A^k B = 0$ pour tout $k \in \mathbb{N}$, puis que $\Psi^\dagger e^{tA}B = 0$ pour tout $t \in [0,T]$. Par conséquent, $\Psi^\dagger \Phi(u) = 0$ pour tout contrôle u, i.e., l'application Φ ne peut être surjective.
(2) Réciproquement, si l'application Φ n'est pas surjective, il existe un vecteur $\Psi \in \mathbb{R}^d$, $\Psi \neq 0$, tel que

$$\Psi^\dagger \int_0^T e^{(T-s)A}Bu(s)\,\mathrm{d}s = 0, \quad \forall u \in L^\infty([0,T];\mathbb{R}^k).$$

En choisissant le contrôle $u(s) = B^\dagger e^{(T-s)A^\dagger}\Psi$, qui est bien dans $L^\infty([0,T];\mathbb{R}^k)$, on en déduit que

$$\Psi^\dagger e^{tA}B = 0\ (\in \mathbb{R}^k), \quad \forall t \in [0,T].$$

En $t = 0$, il vient $\Psi^\dagger B = 0$, puis en dérivant par rapport à t, il vient $\Psi^\dagger AB = 0$ et ainsi de suite ; d'où

$$\Psi^\dagger B = \Psi^\dagger AB = \cdots = \Psi^\dagger A^{d-1}B = 0\,(\in \mathbb{R}^k).$$

La matrice C ne peut donc être de rang maximal. □

Exemple 1.11. [Contrôle d'un tram] L'état du tram (supposé de masse unité) est décrit par sa position $x(t)$ et sa vitesse $v(t)$ le long d'un axe unidirectionnel et on contrôle l'accélération du tram sous la forme

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

Cette équation différentielle du second ordre en temps se récrit comme un système d'ordre un en temps (avec $d = 2$, $k = 1$) :

$$\dot{X}(t) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} X(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \qquad X(t) = \begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix}.$$

La matrice de Kalman $C \in \mathbb{R}^{2\times 2}$ est

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathrm{rang}(C) = 2.$$

Le tram est donc contrôlable en tout temps T à partir de tout $X_0 = (x_0, v_0)^\dagger$ (position et vitesse initiales) : cela signifie que quel que soit $X_1 = (x_1, v_1)^\dagger$ (position et vitesse cibles en T), il existe un contrôle $u \in L^\infty([0,T];\mathbb{R})$ amenant le tram de X_0 en X_1 au temps T. □

Exemple 1.12. [Circuit RLC] Considérons maintenant un exemple issu de l'électronique : le circuit RLC. Ici, x (l'état) représente la charge du circuit et u (le contrôle) la tension appliquée

$$u(t) = L\ddot{x}(t) + R\dot{x}(t) + C^{-1}x(t),$$

ou encore $\ddot{x}(t) = -\frac{R}{L}\dot{x}(t) - \frac{1}{LC}x(t) + \frac{1}{L}u(t)$ On obtient le système de contrôle linéaire (avec $d = 2$, $k = 1$) sous la forme

$$\dot{X}(t) = \begin{pmatrix} 0 & 1 \\ \frac{-1}{LC} & \frac{-R}{L} \end{pmatrix} X(t) + \begin{pmatrix} 0 \\ \frac{1}{L} \end{pmatrix} u(t), \qquad X(t) = \begin{pmatrix} x(t) \\ \dot{x}(t) \end{pmatrix}.$$

La matrice de Kalman $C \in \mathbb{R}^{2\times 2}$ est

$$C = \begin{pmatrix} 0 & \frac{1}{L} \\ \frac{1}{L} & \frac{-R}{L^2} \end{pmatrix}, \qquad \operatorname{rang}(C) = 2,$$

ce qui montre que le circuit RLC est contrôlable. □

Il est intéressant de considérer une reformulation du critère de Kalman. On introduit la matrice $G_T \in \mathbb{R}^{d\times d}$ telle que

$$G_T = \int_0^T e^{(T-s)A} B B^\dagger e^{(T-s)A^\dagger} \, \mathrm{d}s. \tag{1.17}$$

Il est clair que la matrice G_T est symétrique, et on vérifie facilement qu'elle est semi-définie positive car $y^\dagger G_T y = \int_0^T |B^\dagger e^{(T-s)A^\dagger} y|^2_{\mathbb{R}^k} \, \mathrm{d}s \geq 0$ pour tout vecteur $y \in \mathbb{R}^d$.

Lemme 1.13 (Reformulation du critère de Kalman)**.** *Le système linéaire autonome $\dot{x}(t) = Ax(t) + Bu(t)$ est contrôlable pour tout $T > 0$ et pour tout $x_0 \in \mathbb{R}^d$* ***si et seulement si*** *la matrice G_T est inversible.*

Démonstration. (1) Soit $x_1 \in \mathbb{R}^d$. Supposons la matrice G_T inversible et posons

$$\overline{u}(t) = B^\dagger e^{(T-s)A^\dagger} y \quad \text{où} \quad y = G_T^{-1}(x_1 - e^{TA}x_0).$$

Par la formule de Duhamel, on voit que

$$x_{\overline{u}}(T) = e^{TA}x_0 + \int_0^T e^{(T-s)A} B\overline{u}(s) \, \mathrm{d}s = e^{TA}x_0 + G_T y = x_1.$$

Ceci montre que le système est contrôlable.
(2) Supposons qu'il existe $\Psi \in \mathbb{R}^d$, $\Psi \neq 0$, dans $\ker(G_T)$. Il vient

$$0 = \Psi^\dagger G_T \Psi = \int_0^T |B^\dagger e^{(T-s)A^\dagger} \Psi|^2_{\mathbb{R}^k} \, \mathrm{d}s,$$

si bien que $\Psi^\dagger e^{(T-s)A} B = 0$ pour tout $s \in [0,T]$. Par la formule de Duhamel, on obtient $\Psi^\dagger(x_u(T) - e^{TA}x_0) = 0$, ce qui montre que $x_u(T)$ est dans un hyperplan affine. Par conséquent, le système n'est pas contrôlable. □

Remarque 1.14. [Matrice G_T] Le critère de Kalman $\operatorname{rang}(C) = d$ étant indépendant de T, on en déduit que l'inversibilité de la matrice G_T est donc, elle aussi, indépendante de T. Dans le cas des systèmes de contrôle linéaires autonomes, le critère de Kalman est plus simple à vérifier que l'inversibilité de G_T. Toutefois, la matrice G_T nous sera utile dans le chapitre 3 lorsque nous étudierons la synthèse d'un contrôle optimal pour la minimisation d'un critère quadratique. □

Concluons cette section par une extension du critère de Kalman au cas de la contrôlabilité des systèmes linéaires instationnaires, i.e., de la forme

$$\dot{x}_u(t) = A(t)x_u(t) + B(t)u(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0, \tag{1.18}$$

avec $A \in L^1([0,T];\mathbb{R}^{d\times d})$ et $B \in L^1([0,T];\mathbb{R}^{d\times k})$. Pour de tels systèmes, la formule de Duhamel n'est plus valable. On utilise la notion de **résolvante** $R : [0,T] \to \mathbb{R}^{d\times d}$ telle que

$$\dot{R}(t) = A(t)R(t), \quad R(0) = I, \tag{1.19}$$

où I est la matrice identité de $\mathbb{R}^{d\times d}$. On notera que

$$A \in L^1([0,T];\mathbb{R}^{d\times d}) \Longrightarrow R \in AC([0,T];\mathbb{R}^{d\times d}), \tag{1.20a}$$
$$A \in C^0([0,T];\mathbb{R}^{d\times d}) \Longrightarrow R \in C^1([0,T];\mathbb{R}^{d\times d}). \tag{1.20b}$$

Comme $\frac{d}{dt}\det(R(t)) = \operatorname{tr}(A(t))\det(R(t))$ et $\det(R(0)) = 1$, la matrice $R(t)$ est inversible à tout temps (la quantité $\det(R(t))$ s'appelle le Wronskien au temps t). On notera également que, dans le cas autonome où $A(t) = A$, on a $R(t) = e^{tA}$. On vérifie sans peine que la solution du système différentiel instationnaire (1.18) est

$$x_u(t) = R(t)x_0 + R(t)\int_0^t R(s)^{-1}B(s)u(s)\,\mathrm{d}s, \quad \forall t \in [0,T]. \tag{1.21}$$

Lemme 1.15 (Critère de contrôlabilité, cas instationnaire)**.** *Le système instationnaire* (1.18) *est contrôlable en temps T à partir de x_0* ***si et seulement si*** *la matrice de contrôlabilité*

$$K_T := \int_0^T R(s)^{-1}B(s)B(s)^\dagger(R(s)^{-1})^\dagger\,\mathrm{d}s \in \mathbb{R}^{d\times d} \tag{1.22}$$

est inversible.

Démonstration. Identique au cas autonome. □

Remarque 1.16. [Matrice K_T] La condition (1.22) dépend de T, mais pas de x_0. Ainsi, la contrôlabilité en temps T à partir de x_0 implique la contrôlabilité en temps T à partir de tout point ; en revanche, on ne peut s'affranchir de la dépendance en T. On notera également que dans le cas autonome, on a $R(s) = e^{sA}$ et $B(s) = B$, si bien que

$$K_T = e^{-TA}\left(\int_0^T e^{(T-s)A}BB^\dagger e^{(T-s)A^\dagger}\,\mathrm{d}s\right)e^{-TA^\dagger} = e^{-TA}G_T e^{-TA^\dagger}$$

On retrouve donc le critère du lemme 1.13 sur la matrice G_T. □

Contre-exemple 1.17. [Non-contrôlabilité] On considère le système de contrôle linéaire instationnaire

$$\dot{X}_u(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X_u(t) + \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} u(t). \tag{1.23}$$

On vérifie facilement que $R(s) = e^{sA} = \begin{pmatrix} \cos(s) & -\sin(s) \\ \sin(s) & \cos(s) \end{pmatrix}$, d'où

$$R(s)^{-1}B(s) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \implies K_T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

La matrice K_T n'est donc pas inversible, si bien que le système (1.23) n'est pas contrôlable. Le problème vient du fait que la matrice $R(s)^{-1}B(s)$ est indépendante de s. En revanche, si le vecteur B était constant (et non-nul), le système serait contrôlable car B et AB seraient alors des vecteurs orthogonaux non-nuls, si bien que la matrice de Kalman $C = (B, AB)$ serait de rang plein. □

1.3 Cas avec contraintes : ensemble atteignable

On considère le système de contrôle linéaire autonome

$$\dot{x}_u(t) = Ax_u(t) + Bu(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0. \tag{1.24}$$

Comme ci-dessus, l'horizon temporel $T > 0$ et la condition initiale $x_0 \in \mathbb{R}^d$ sont fixés. Les résultats de cette section s'étendent au cas instationnaire avec terme de dérive, mais pour simplifier, nous ne traiterons pas ce cas plus général.

Dans cette section, nous suppose le contrôle à valeurs dans un sous-ensemble **compact non-vide**

$$U \subset \mathbb{R}^k. \tag{1.25}$$

En particulier, le contrôle $u(t)$ est borné pour tout $t \in [0,T]$. On a donc $u \in L^\infty([0,T]; U)$. (On notera que $L^1([0,T]; U) = L^\infty([0,T]; U)$ lorsque l'ensemble U est borné.)

Définition 1.18 (Ensemble atteignable)**.** *Pour tout $t \in [0,T]$ et tout $x_0 \in \mathbb{R}^d$, l'**ensemble atteignable** en temps t à partir de x_0 est défini comme suit :*

$$\mathcal{A}(t, x_0) = \{x_1 \in \mathbb{R}^d \mid \exists u \in L^\infty([0,t]; U) \text{ tel que } x_u(t) = x_1\}. \tag{1.26}$$

Théorème 1.19 (Propriétés de l'ensemble atteignable)**.** *Pour tout $t \in [0,T]$, l'ensemble atteignable $\mathcal{A}(t, x_0)$ est **compact**, **convexe**, et varie **continûment** en t. La continuité en temps est uniforme, i.e., pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que*

$$\forall t_1, t_2 \in [0,T], \quad |t_1 - t_2| \leq \delta \implies d(\mathcal{A}(t_1, x_0), \mathcal{A}(t_2, x_0)) \leq \epsilon, \tag{1.27}$$

où la distance de Hausdorff entre deux sous-ensembles $\mathcal{A}_1$ et $\mathcal{A}_2$ de $\mathbb{R}^d$ est définie comme suit (cf. la figure 1.2) :

$$\begin{aligned} d(\mathcal{A}_1, \mathcal{A}_2) &:= \max\left(\sup_{x_1 \in \mathcal{A}_1} d(x_1, \mathcal{A}_2), \sup_{x_2 \in \mathcal{A}_2} d(x_2, \mathcal{A}_1) \right) \\ &= \max\left(\sup_{x_1 \in \mathcal{A}_1} \inf_{y_2 \in \mathcal{A}_2} |x_1 - y_2|_{\mathbb{R}^d}, \sup_{x_2 \in \mathcal{A}_2} \inf_{y_1 \in \mathcal{A}_1} |x_2 - y_1|_{\mathbb{R}^d} \right). \end{aligned} \tag{1.28}$$

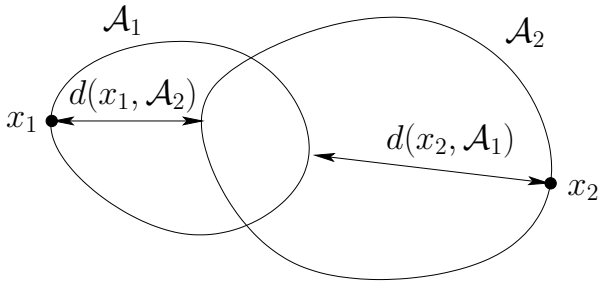

FIGURE 1.2 – Distance de Hausdorff entre deux sous-ensembles $\mathcal{A}_1$ et $\mathcal{A}_2$ de $\mathbb{R}^d$.

Démonstration. Nous verrons les preuves de variation continue en temps et de compacité au chapitre 2 dans le cas plus général des systèmes de contrôle non-linéaires. Nous nous contentons ici de prouver la convexité de l'ensemble atteignable $\mathcal{A}(t, x_0)$, propriété qui est, quant à elle, spécifique au cas linéaire.
(1) Cas où le sous-ensemble U est convexe. Dans ce cas, la preuve de convexité de l'ensemble atteignable $\mathcal{A}(t, x_0)$ est élémentaire. Soit $x_1, x_2 \in \mathcal{A}(t, x_0)$, soit $\theta \in [0, 1]$ et montrons que $\theta x_1 + (1-\theta)x_2 \in \mathcal{A}(t, x_0)$. Par définition, il existe des contrôles $u_i \in L^\infty([0,t]; U)$, $i \in \{1, 2\}$, tels que

$$x_i = e^{tA}x_0 + \int_0^t e^{(t-s)A} B u_i(s)\, \mathrm{d}s,$$

où x_i est la trajectoire associée au contrôle u_i, $i \in \{1, 2\}$. Posons $u(s) = \theta u_1(s) + (1-\theta)u_2(s)$, pour tout $s \in [0,t]$. La fonction u est mesurable et cette fonction est à valeurs dans U grâce à la convexité du sous-ensemble U. De plus, par linéarité, la trajectoire x_u associée au contrôle u vérifie

$$\begin{aligned} x_u(t) &= e^{tA}x_0 + \int_0^t e^{(t-s)A} B u(s)\, \mathrm{d}s \\ &= e^{tA}x_0 + \theta \int_0^t e^{(t-s)A} B u_1(s)\, \mathrm{d}s + (1-\theta) \int_0^t e^{(t-s)A} B u_2(s)\, \mathrm{d}s \\ &= \theta x_1 + (1-\theta)x_2, \end{aligned}$$

ce qui montre que $\theta x_1 + (1-\theta)x_2 \in \mathcal{A}(t, x_0)$.
(2) Cas général pour U. Dans ce cas, on invoque le Lemme de Lyapunov 1.20 rappelé ci-dessous (pour la preuve, voir par exemple la référence [5]). Soit $x_1, x_2 \in \mathcal{A}(t, x_0)$, soit $\theta \in [0, 1]$

et montrons à nouveau que $\theta x_1 + (1-\theta)x_2 = x(t) \in \mathcal{A}(t, x_0)$. Par définition, il existe des contrôles $u_i \in L^\infty([0,t]; U)$, $i \in \{1,2\}$, tels que $x_i = e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu_i(s)\,\mathrm{d}s$. Posons $y_i = x_i - e^{tA}x_0$ et considérons la fonction $f \in L^1([0,t]; \mathbb{R}^{2d})$ telle que

$$f(s) = \begin{pmatrix} e^{(t-s)A}Bu_1(s) \\ e^{(t-s)A}Bu_2(s) \end{pmatrix} \in \mathbb{R}^{2d}.$$

On a $\int_{\{0\}} f(s)\,\mathrm{d}s = (0,0)^\dagger$ et $\int_{[0,t]} f(s)\,\mathrm{d}s = (y_1, y_2)^\dagger$. En invoquant le lemme de Lyapunov, on en déduit qu'il existe un sous-ensemble mesurable $E \subset [0,t]$ tel que

$$\int_E f(s)\,\mathrm{d}s = \begin{pmatrix} \theta y_1 \\ \theta y_2 \end{pmatrix}.$$

En notant E^c le complémentaire de E dans $[0,t]$, on a

$$\int_{E^c} f(s)\,\mathrm{d}s = \int_{[0,t]} f(s)\,\mathrm{d}s - \int_E f(s)\,\mathrm{d}s = \begin{pmatrix} (1-\theta)y_1 \\ (1-\theta)y_2 \end{pmatrix}.$$

Finalement, on pose

$$u(s) = \begin{cases} u_1(s) & \text{si } s \in E, \\ u_2(s) & \text{si } s \in E^c. \end{cases}$$

Le contrôle ainsi défini est bien une fonction mesurable de $[0,t]$ dans U car les ensembles E et E^c sont mesurables. De plus, la trajectoire x_u associée à ce contrôle satisfait

$$\begin{aligned} x_u(t) - e^{tA}x_0 &= \int_{[0,t]} e^{(t-s)A}Bu(s)\,\mathrm{d}s \\ &= \int_E e^{(t-s)A}Bu_1(s)\,\mathrm{d}s + \int_{E^c} e^{(t-s)A}Bu_2(s)\,\mathrm{d}s = \theta y_1 + (1-\theta)y_2, \end{aligned}$$

ce qui montre que $\theta x_1 + (1-\theta)x_2 = x_u(t) \in \mathcal{A}(t, x_0)$. □

Lemme 1.20 (Lyapunov)**.** *Soit $t > 0$ et un entier $n \geq 1$. Soit une fonction $f \in L^1([0,t]; \mathbb{R}^n)$. Alors, le sous-ensemble*

$$\left\{ \int_E f(s)\,\mathrm{d}s \mid E \subset [0,t] \text{ mesurable} \right\} \tag{1.29}$$

est un sous-ensemble ***convexe*** *de $\mathbb{R}^n$.*

Remarque 1.21. [Atteignabilité avec U et $\mathrm{conv}(U)$] On peut montrer que l'ensemble atteignable pour des contrôles à valeurs dans U est le même que pour des contrôles à valeurs dans $\mathrm{conv}(U)$ (l'enveloppe convexe de U). □

Exemple 1.22. [Mouvement d'un point matériel] On considère un point matériel en mouvement rectiligne. On contrôle la vitesse de ce point par un contrôle à valeurs dans l'intervalle borné $U := [-1,1]$:

$$\dot{x}(t) = u(t),\ \forall t \in [0,T], \quad x(0) = 0, \quad u(t) \in U = [-1,1],$$

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). On constate qu'on obtient le même ensemble atteignable en se restreignant à des contrôles à valeurs dans $\partial U = \{-1,1\}$. De tels contrôles sont appelés des **contrôles bang-bang** car ils ne prennent que des valeurs extrémales dans ∂U. Une illustration est présentée à la figure 1.3. □

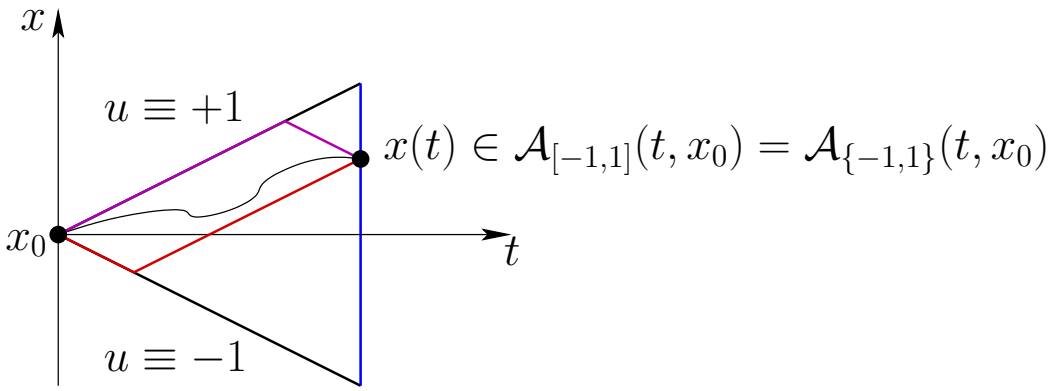

FIGURE 1.3 – Ensemble atteignable par un point matériel dont on contrôle la vitesse dans $U = [-1,1]$.

Chapitre 2

Contrôlabilité des systèmes non-linéaires

Ce chapitre est consacré à la contrôlabilité des systèmes de contrôle non-linéaires. Comme au chapitre précédent, la notion d'**ensemble atteignable** joue un rôle important. Le résultat principal de ce chapitre est un critère de **contrôlabilité locale** au voisinage d'une cible située dans l'ensemble atteignable, ce critère se formulant à l'aide de la contrôlabilité du système linéarisé. Afin d'établir ce résultat, nous montrerons que, sous certaines hypothèses, la différentielle de l'application entrée-sortie (qui à un contrôle associe l'état du système au temps final) est différentiable et que sa différentielle est l'application entrée-sortie du système linéarisé. Ce chapitre sera aussi l'occasion de voir ou revoir certains outils mathématiques importants : théorème de Cauchy–Lipschitz pour les systèmes différentiels avec fonctions mesurables, topologie faible dans les espaces de Hilbert et différentielle de Fréchet.

2.1 Théorème de Cauchy–Lipschitz

On fixe un horizon temporel $T > 0$ et une condition initiale $x_0 \in \mathbb{R}^d$. On considère le **problème de Cauchy** qui consiste à chercher une fonction $x : [0,T] \to \mathbb{R}^d$ telle que

$$\dot{x}(t) = F(t, x(t)), \quad \forall t \in [0,T], \qquad x(0) = x_0, \tag{2.1}$$

pour une application donnée $F : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$. Commençons par rappeler un résultat bien connu.

Théorème 2.1 (Cauchy–Lipschitz, cas continu et Lipschitz global)**.** *On suppose que :*
(i) *L'application F est* ***continue*** *en t et en x, i.e., $F \in C^0([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$;*
(ii) *L'application F est* ***globalement lipschitzienne*** *en x, i.e.,*

$$\exists C_0 \in \mathbb{R}_+, \quad \forall t \in [0,T],\ \forall x_1, x_2 \in \mathbb{R}^d, \quad |F(t,x_1) - F(t,x_2)|_{\mathbb{R}^d} \leq C_0 |x_1 - x_2|_{\mathbb{R}^d}. \tag{2.2}$$

Alors, il existe une unique solution au problème de Cauchy telle que

$$x \in C^1([0,T]; \mathbb{R}^d). \tag{2.3}$$

Cette solution satisfait donc le système différentiel (2.1) *pour tout $t \in [0,T]$.*

Démonstration. Le principe de la preuve consiste à observer que x est solution du problème de Cauchy (2.1) si et seulement si

$$x(t) = x_0 + \int_0^t F(s, x(s))\,\mathrm{d}s, \quad \forall t \in [0,T].$$

On introduit l'espace $Y = C^0([0,T];\mathbb{R}^d)$; il s'agit d'un espace de Banach (espace vectoriel normé complet) équipé de la norme de la convergence uniforme $\|y\|_Y = \sup_{t\in[0,T]} |y(t)|_{\mathbb{R}^d}$ pour tout $y \in Y$. Résoudre le problème de Cauchy revient à chercher un point fixe de l'application $\Phi : Y \to Y$ où pour tout $y \in Y$, $\Phi(y)$ est tel que

$$\Phi(y)(t) = x_0 + \int_0^t F(s, y(s))\,\mathrm{d}s, \quad \forall t \in [0,T].$$

Montrons que l'application Φ est strictement contractante de Y dans Y. On considère la norme $\|y\|_{Y*} = \sup_{t\in[0,T]}(e^{-C_0 t}|y(t)|_{\mathbb{R}^d})$ où C_0 est la constante intervenant dans la propriété de Lipschitz globale de l'application F. Il est clair que la norme $\|\cdot\|_{Y*}$ est équivalente à la norme $\|\cdot\|_Y$ sur Y. On constate que pour tout $y_1, y_2 \in Y$, on a

$$\begin{aligned}
\|\Phi(y_1) - \Phi(y_2)\|_{Y*} &= \sup_{t\in[0,T]} \left(e^{-C_0 t} |\Phi(y_1)(t) - \Phi(y_2)(t)|_{\mathbb{R}^d} \right) \\
&\leq \sup_{t\in[0,T]} \left(e^{-C_0 t} \int_0^t |F(s, y_1(s)) - F(s, y_2(s))|_{\mathbb{R}^d}\,\mathrm{d}s \right) \\
&\leq \sup_{t\in[0,T]} \left(e^{-C_0 t} C_0 \int_0^t |y_1(s) - y_2(s)|_{\mathbb{R}^d}\,\mathrm{d}s \right) \\
&= \sup_{t\in[0,T]} \left(e^{-C_0 t} C_0 \int_0^t e^{C_0 s} e^{-C_0 s} |y_1(s) - y_2(s)|_{\mathbb{R}^d}\,\mathrm{d}s \right) \\
&\leq \left(\sup_{t\in[0,T]} e^{-C_0 t} C_0 \int_0^t e^{C_0 s}\,\mathrm{d}s \right) \|y_1 - y_2\|_{Y*} \\
&= \left(\sup_{t\in[0,T]} 1 - e^{-C_0 t} \right) \|y_1 - y_2\|_{Y*} = \left(1 - e^{-C_0 T}\right) \|y_1 - y_2\|_{Y*},
\end{aligned}$$

où on a utilisé le caractère globalement lipschitzien en x de l'application F pour passer de la deuxième à la troisième ligne du calcul. L'application Φ est donc bien strictement contractante de Y dans Y. On conclut par le théorème du point fixe de Picard. □

L'hypothèse de continuité en t de l'application F faite au théorème 2.1 n'est pas vraiment satisfaisante pour l'étude des systèmes de contrôle. En effet, ces systèmes s'écrivent sous la forme

$$\dot{x}(t) = f(t, x(t), u(t)), \quad \forall t \in [0,T], \qquad x(0) = x_0, \tag{2.4}$$

où $u \in L^1([0,T];\mathbb{R}^k)$ et $f : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d$. L'étude du système différentiel (2.4) se ramène à celle du problème de Cauchy (2.1) en posant

$$F(t,x) = f(t, x, u(t)), \quad \forall (t,x) \in [0,T] \times \mathbb{R}^d. \tag{2.5}$$

On voit donc que même si l'application f est régulière en u, le fait que le contrôle ne dépende pas continûment du temps fait que l'application F ne sera pas nécessairement continue en t. Afin de traiter cette situation, on dispose de la variante suivante du théorème 2.1 (la preuve utilise des arguments analogues à ceux évoqués ci-dessus). On renvoie le lecteur à la définition 1.2 pour la notion de fonction absolument continue.

Théorème 2.2 (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 $x \in \mathbb{R}^d$, 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.,*

$$\forall x \in \mathbb{R}^d, \quad \exists \beta \in L^1([0,T];\mathbb{R}_+), \quad \forall t \in [0,T], \quad |F(t,x)|_{\mathbb{R}^d} \leq \beta(t); \tag{2.6}$$

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

$$\begin{aligned} &\exists C_0 \in L^1([0,T];\mathbb{R}_+), \\ &\text{p.p. } t \in [0,T],\ \forall x_1, x_2 \in \mathbb{R}^d, \quad |F(t,x_1) - F(t,x_2)|_{\mathbb{R}^d} \leq C_0(t)|x_1 - x_2|_{\mathbb{R}^d}. \end{aligned} \tag{2.7}$$

Alors, il existe une ***unique solution*** *au problème de Cauchy telle que*

$$x \in AC([0,T];\mathbb{R}^d). \tag{2.8}$$

Cette solution, qui est dérivable p.p. sur $[0,T]$, satisfait le système différentiel (2.1) *pour presque tout $t \in [0,T]$; elle vérifie également*

$$x(t) = x_0 + \int_0^t F(s,x(s))\,\mathrm{d}s, \quad \forall t \in [0,T]. \tag{2.9}$$

Remarque 2.3. [Intégrabilité] Grâce à la propriété (iii) du théorème 2.2, il suffit, afin d'établir la propriété (ii), de montrer que $F(t,0) \in L^1([0,T];\mathbb{R}^d)$. □

Un cas d'application du théorème 2.2 est le cas linéaire (éventuellement avec un terme de dérive) où on a $F(t,x) = A(t)x + r(t)$ avec $A \in L^1([0,T];\mathbb{R}^{d\times d})$ et $r \in L^1([0,T];\mathbb{R}^d)$; l'application F est alors globalement lipschitzienne de constante $C_0(t) = |A(t)|_{\mathbb{R}^{d\times d}}$ (où $|\cdot|_{\mathbb{R}^{d\times d}}$ désigne la norme matricielle subordonnée à la norme euclidienne). Lorsque l'application F est non-linéaire en x, la propriété d'être globalement lipschitzienne est en général perdue. Dans ce cas, il est bien connu que la solution x du problème de Cauchy (2.1) peut exploser en temps fini.

Exemple 2.4. [Explosion en temps fini] Donnons un exemple simple d'explosion en temps fini. On se place dans $\mathbb{R}$ ($d = 1$) et on considère l'application $F(t,x) = 1 - x^2$ (qui ne dépend que de x). Le problème de Cauchy est donc $\dot{x}(t) = 1 - x(t)^2$ avec $x(0) = x_0 \in \mathbb{R}$. Si $|x_0| \leq 1$, il vient $x(t) = \tanh(t + t_0)$ avec $\tanh(t_0) = x_0$ et $\lim_{t\to\infty} x(t) = 1$; on a donc existence globale en temps de la solution. En revanche, si $|x_0| > 1$, il vient $x(t) = \coth(t + t_0)$ avec $\coth(t_0) = x_0$ et deux situations peuvent se produire : (i) si $x_0 > 1$, alors $t_0 > 0$ et on a $\lim_{t\to\infty} x(t) = 1$, i.e., on a encore existence globale en temps de la solution ; (ii) si $x_0 < -1$, alors $t_0 < 0$ et dans ces conditions, $\lim_{t\uparrow t_0} |x(t)| = +\infty$; on a donc explosion en temps fini. □

Remarque 2.5. [Non-unicité] Lorsque l'application F est uniquement continue en x, on peut ne pas avoir unicité de la solution du problème de Cauchy. Par exemple, pour le problème de Cauchy $\dot{x}(t) = \sqrt{|x(t)|}$ avec $x(0) = 0$ (i.e., pour $F(t,x) = \sqrt{|x|}$), $x(t) \equiv 0$ est solution, et il en est de même de $x(t) = \frac{1}{4}t^2$ et de $x(t) = \frac{1}{4}\max(t - t_0, 0)^2$ pour tout $t_0 \in \mathbb{R}_+$. □

Afin de traiter le cas de dynamiques non-linéaires, on dispose de l'extension suivante du théorème 2.2, où la propriété de Lipschitz globale est remplacée par une propriété locale (pour la preuve, voir par exemple l'annexe C de la référence [10]).

Théorème 2.6 (Cauchy–Lipschitz, cas mesurable et Lipschitz local)**.** *On suppose que :*
(i) *L'application F est mesurable en t et continue en x ;*
(ii) *L'application F est intégrable en t, i.e.,*

$$\forall x \in \mathbb{R}^d, \quad \exists \beta \in L^1([0,T];\mathbb{R}_+), \quad \forall t \in [0,T], \quad |F(t,x)|_{\mathbb{R}^d} \leq \beta(t); \tag{2.10}$$

(iii) *L'application F est* ***localement lipschitzienne*** *en x, i.e.,*

$$\begin{aligned} &\forall x \in \mathbb{R}^d, \quad \exists r > 0, \quad \exists C_0 \in L^1([0,T];\mathbb{R}_+), \\ &\text{p.p. } t \in [0,T],\ \forall x_1, x_2 \in B(x,r), \quad |F(t,x_1) - F(t,x_2)|_{\mathbb{R}^d} \leq C_0(t)|x_1 - x_2|_{\mathbb{R}^d}, \end{aligned} \tag{2.11}$$

où $B(x,r)$ désigne la boule ouverte de centre x et de rayon r.
Alors, il existe une ***unique solution maximale*** *au problème de Cauchy* (2.1)*. Cette solution est définie sur l'intervalle $J \subseteq [0,T]$ et on a soit $J = [0,T]$ soit $J = [0,T_*[$ avec $T_* < T$ et $\lim_{t \uparrow T_*} |x(t)|_{\mathbb{R}^d} = +\infty$. La solution maximale x est dans $AC(J;\mathbb{R}^d)$, elle satisfait le système différentiel* (2.1) *pour presque tout $t \in J$ et elle vérifie* (2.9) *pour tout $t \in J$.*

Exemple 2.7. [Explosion pour un système de contrôle] On se place dans $\mathbb{R}$ ($d = 1$) et on considère le système de contrôle (2.4) avec un contrôle à valeurs scalaires ($k = 1$) 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 $T_* = \frac{\pi}{2\sqrt{u_0}}$ qui dépend de la valeur (constante) prise par le contrôle. □

2.2 Ensemble atteignable

On fixe un horizon temporel $T > 0$ et une condition initiale $x_0 \in \mathbb{R}^d$. On considère le système de contrôle non-linéaire

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in [0,T], \qquad x_u(0) = x_0. \tag{2.12}$$

Soit $U \subset \mathbb{R}^k$ un sous-ensemble compact non-vide de $\mathbb{R}^k$. La définition de l'ensemble atteignable (en temps $t \in [0,T]$ à partir de x_0) est identique à celle que nous avons introduite dans le cas linéaire (cf. la définition 1.18).

Définition 2.8 (Ensemble atteignable)**.** *Pour tout $t \in [0,T]$, l'**ensemble atteignable** en temps t à partir de x_0 est défini comme suit :*

$$\mathcal{A}(t, x_0) = \{x_1 \in \mathbb{R}^d \mid \exists u \in L^\infty([0,t]; U) \text{ tel que } x_u(t) = x_1\}. \tag{2.13}$$

Nous allons établir deux propriétés importantes et utiles de l'ensemble atteignable : sa variation continue en temps et sa compacité.

Lemme 2.9 (Variation continue en temps)**.** *On suppose que*
(i) *f est de classe C^0 sur $\mathbb{R} \times \mathbb{R}^d \times U$;*
(ii) *U est un sous-ensemble compact non-vide de $\mathbb{R}^k$;*
(iii) *les trajectoires sont uniformément bornées, i.e.,*

$$\exists M > 0, \quad \forall u \in L^\infty([0,T]; U), \quad \sup_{t \in [0,T]} |x_u(t)|_{\mathbb{R}^d} \leq M. \tag{2.14}$$

Alors, l'ensemble $\mathcal{A}(t, x_0)$ varie continûment en temps, et ce de manière uniforme, i.e., pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que

$$\forall t_1, t_2 \in [0,T], \quad |t_1 - t_2| \leq \delta \implies d(\mathcal{A}(t_1, x_0), \mathcal{A}(t_2, x_0)) \leq \epsilon, \tag{2.15}$$

où la distance de Hausdorff entre deux sous-ensembles est définie en (1.28) *(cf. la figure 2.1).*

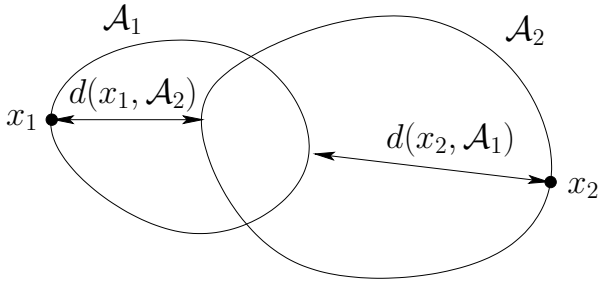

FIGURE 2.1 – Distance de Hausdorff entre deux sous-ensembles $\mathcal{A}_1$ et $\mathcal{A}_2$.

Remarque 2.10. [Cas linéaire] Les hypothèses du lemme 2.9 sont bien vérifiées dans le cas linéaire. L'ensemble atteignable varie donc continûment en temps dans ce cas. □

Démonstration. Soit $\epsilon > 0$. On va montrer qu'il existe $\delta > 0$ tel que

$$\forall t_1, t_2 \in [0,T], \quad |t_1 - t_2| \leq \delta \implies d(\mathcal{A}_1, \mathcal{A}_2) \leq \epsilon,$$

où $\mathcal{A}_1 = \mathcal{A}(t_1, x_0)$ et $\mathcal{A}_2 = \mathcal{A}(t_2, x_0)$. Supposons pour fixer les idées que $t_2 > t_1$. Soit $x_2 \in \mathcal{A}_2$. Il existe donc un contrôle $u \in L^\infty([0,t_2]; U)$ tel que

$$x_2 = x_0 + \int_0^{t_2} f(s, x(s), u(s))\, \mathrm{d}s.$$

Avec ce même contrôle, on pose

$$x_1 = x_0 + \int_0^{t_1} f(s, x(s), u(s))\,\mathrm{d}s \in \mathcal{A}(t_1, x_0).$$

D'après les hypothèses sur f, x et u, on a

$$|x_2 - x_1|_{\mathbb{R}^d} \leq \int_{t_1}^{t_2} |f(s, x(s), u(s))|_{\mathbb{R}^d}\,\mathrm{d}s \leq C|t_2 - t_1|.$$

Ceci montre que $d(x_2, \mathcal{A}_1) \leq |x_2 - x_1|_{\mathbb{R}^d} \leq C|t_2 - t_1|$. On raisonne de même pour $x_1 \in \mathcal{A}_1$, ce qui conclut la preuve. □

Lemme 2.11 (Compacité)**.** *On suppose que*
(i) *f est de classe C^0 sur $\mathbb{R} \times \mathbb{R}^d \times U$ et de classe C^1 en x ;*
(ii) *U est un sous-ensemble compact non-vide de $\mathbb{R}^k$;*
(iii) *les trajectoires sont uniformément bornées, i.e.,*

$$\exists M > 0, \quad \forall u \in L^\infty([0,T]; U), \quad \sup_{t\in[0,T]} |x_u(t)|_{\mathbb{R}^d} \leq M; \tag{2.16}$$

(iv) *pour tout $(t, x) \in [0,T] \times \mathbb{R}^d$, l'ensemble des vecteurs vitesse $K(t, x) := \{f(t, x, u) \mid u \in U\}$ est un sous-ensemble **convexe** de $\mathbb{R}^d$.*
Alors, pour tout $t \in [0,T]$, l'ensemble atteignable $\mathcal{A}(t, x_0)$ est un sous-ensemble compact de $\mathbb{R}^d$.

Remarque 2.12. [Cas linéaire] Les hypothèses du lemme 2.11 sont bien vérifiées dans le cas linéaire avec U convexe. L'ensemble atteignable est donc compact dans ce cas. □

Démonstration. On se place dans l'espace de Hilbert $V = L^2([0,T]; \mathbb{R}^d)$ et on va montrer la compacité de l'ensemble atteignable $\mathcal{A}(T, x_0)$. La preuve utilise des notions de topologie faible dans les espaces de Hilbert ; les quelques notions qui nous seront utiles dans cette preuve sont rappelées à la sous-section 2.4.1 ci-dessous.
(1) Soit $(y_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{A}(T, x_0) \subset \mathbb{R}^d$. Soit $(u_n)_{n\in\mathbb{N}}$ une suite de contrôles dans $L^\infty([0,T]; U)$ et $(x_n)_{n\in\mathbb{N}}$ la suite de trajectoires correspondantes dans $AC([0,T]; \mathbb{R}^d)$ menant de x_0 à y_n. Posons $g_n(s) = f(s, x_n(s), u_n(s))\,\mathrm{d}s$ pour tout $n \in \mathbb{N}$ et $s \in [0,T]$. On a

$$x_n(t) = x_0 + \int_0^t g_n(s)\,\mathrm{d}s, \ \forall t \in [0,T] \quad \text{et} \quad y_n = x_n(T).$$

D'après les hypothèses, la suite $(g_n)_{n\in\mathbb{N}}$ est bornée dans V. En invoquant le théorème 2.23 sur la compacité faible dans les espaces de Hilbert, on en déduit qu'à une sous-suite près, la suite $(g_n)_{n\in\mathbb{N}}$ converge vers une fonction $g \in V$ pour la topologie faible. On définit la trajectoire $x \in AC([0,T]; \mathbb{R}^d)$ en posant

$$x(t) = x_0 + \int_0^t g(s)\,\mathrm{d}s, \ \forall t \in [0,T].$$

Par convergence faible, on a $\int_0^t g_n(s)\,\mathrm{d}s = (g_n, 1_{[0,t]})_V \to (g, 1_{[0,t]})_V = \int_0^t g(s)\,\mathrm{d}s$, i.e.,

$$\lim_{n\to+\infty} x_n(t) = x(t), \quad \forall t \in [0,T].$$

En particulier, on a donc

$$\lim_{n\to+\infty} y_n = x(T).$$

Il reste à montrer que la trajectoire $x(t)$ peut bien être engendrée par un contrôle $u \in L^\infty([0,T];U)$.
(2) Posons $\theta_n(s) = f(s, x(s), u_n(s))$ et introduisons l'ensemble

$$\Theta = \{\theta \in V \mid \theta(s) \in K(s, x(s)),\ \forall s \in [0,T]\},$$

de sorte que $(\theta_n)_{n\in\mathbb{N}}$ est une suite de Θ. Par hypothèse, $K(s, x(s))$ est un sous-ensemble convexe de $\mathbb{R}^d$ pour tout $s \in [0,T]$. On en déduit que Θ est un sous-ensemble convexe de V. De plus, Θ est fermé dans V car la convergence dans V implique la convergence p.p. d'une sous-suite, et $K(s, x(s))$ est fermé dans $\mathbb{R}^d$. Grâce au théorème 2.24 sur la fermeture faible des convexes dans les espaces de Hilbert, on en déduit que Θ est faiblement fermé dans V. De plus, comme la suite $(\theta_n)_{n\in\mathbb{N}}$ est bornée dans V, on déduit du théorème 2.23 qu'elle converge faiblement, à une sous-suite près, vers une fonction $\theta \in \Theta$. Il existe donc une fonction $u : [0,T] \to U$ telle que $\theta(s) = f(s, x(s), u(s))$ p.p. dans $[0,T]$, et la fonction u peut être choisie mesurable (cf. la sous-section 2.4.3 pour plus de précisions sur ce point). Pour tout $\varphi \in V$, on a

$$\int_0^T g_n(s)\varphi(s)\,\mathrm{d}s = \int_0^T \theta_n(s)\varphi(s)\,\mathrm{d}s + \int_0^T (f(s, x_n(s), u_n(s)) - f(s, x(s), u_n(s))\varphi(s)\,\mathrm{d}s. \quad (2.17)$$

Comme $|f(s, x_n(s), u_n(s)) - f(s, x(s), u_n(s))|_{\mathbb{R}^d} \le C|x_n(s) - x(s)|_{\mathbb{R}^d}$ et $|x_n(s) - x(s)|_{\mathbb{R}^d}$ tend vers zéro p.p. dans $[0,T]$, le deuxième terme au membre de droite de (2.17) tend vers zéro (invoquer le théorème de convergence dominée de Lebesgue). En outre, par convergence faible, on a $\int_0^T g(s)\varphi(s)\,\mathrm{d}s = \int_0^T \theta(s)\varphi(s)\,\mathrm{d}s$, i.e., $g(s) = \theta(s)$ p.p. dans $[0,T]$. En conclusion, on a bien $g(s) = f(s, x(s), u(s))$ p.p. sur $[0,T]$. □

2.3 Contrôlabilité locale des systèmes non-linéaires

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

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in [0,T], \qquad x_u(0) = x_0. \quad (2.18)$$

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

Définition 2.13 (Application entrée-sortie)**.** *L'**application entrée-sortie** en temps T à partir de x_0 est l'application*

$$E_{T,x_0} : \mathcal{U}_{T,x_0} \to \mathcal{A}(T, x_0), \quad E_{T,x_0}(u) = x_u(T), \quad (2.19)$$

où $\mathcal{U}_{T,x_0} \subset L^\infty([0,T];U)$, U étant un sous-ensemble fermé non-vide de $\mathbb{R}^k$, est le domaine de E_{T,x_0}, i.e., l'ensemble des contrôles tels que la trajectoire associée x_u est bien définie sur $[0,T]$. L'ensemble atteignable $\mathcal{A}(T, x_0)$ est l'image de l'application entrée-sortie E_{T,x_0}.

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 $y \in \mathcal{A}(T, x_0)$.

Définition 2.14 (Contrôlabilité locale)**.** *On dit que le système de contrôle non-linéaire* (2.18) *est contrôlable localement en un point $y \in \mathcal{A}(T, x_0)$ s'il existe un voisinage V_y de y dans $\mathbb{R}^d$ tel que $V_y \subset \mathcal{A}(T, x_0)$, i.e., pour tout $y' \in V_y$, il existe un contrôle $u_{y'} \in \mathcal{U}_{T,x_0}$ amenant l'état de x_0 à y' en temps T.*

Afin d'étudier la contrôlabilité locale du système de contrôle non-linéaire (2.18), nous allons considérer la différentielle (de Fréchet) de l'application entrée-sortie E_{T,x_0}. On renvoie le lecteur à la sous-section 2.4.2 ci-dessous pour quelques rappels sur la notion de différentielle de Fréchet dans les espaces de Banach. Pour simplifier, on se place pour le reste de cette section dans le cas **sans contrainte**, i.e., on suppose que $U = \mathbb{R}^k$ si bien que l'on a $\mathcal{U}_{T,x_0} \subset L^\infty([0,T];\mathbb{R}^k)$. Par des arguments de dépendance de la solution d'un système différentiel en des paramètres, on vérifie facilement que $\mathcal{U}_{T,x_0}$ est un sous-ensemble ouvert de $L^\infty([0,T];\mathbb{R}^k)$. On est donc dans la situation où

$$E_{T,x_0} : \mathcal{U}_{T,x_0} \subset L^\infty([0,T];\mathbb{R}^k) \to \mathcal{A}(T, x_0) \subset \mathbb{R}^d. \tag{2.20}$$

Soit $u \in \mathcal{U}_{T,x_0}$ et $x_u \in AC([0,T];\mathbb{R}^d)$ la trajectoire associée. Soit

$$\delta u \in L^\infty([0,T];\mathbb{R}^k), \tag{2.21}$$

une perturbation du contrôle ; on suppose cette perturbation suffisamment petite pour que $u + \delta u \in \mathcal{U}_{T,x_0}$ (ceci est possible puisque $\mathcal{U}_{T,x_0}$ est un sous-ensemble ouvert de $L^\infty([0,T];\mathbb{R}^k)$). On considère le système différentiel linéarisé le long de la trajectoire x_u, i.e.,

$$\dot{\delta x}(t) = A_u(t)\delta x(t) + B_u(t)\delta u(t), \quad \forall t \in [0,T], \qquad \delta x(0) = 0, \tag{2.22}$$

où pour tout $t \in [0,T]$,

$$A_u(t) = \frac{\partial f}{\partial x}(t, x_u(t), u(t)) \in \mathbb{R}^{d\times d}, \qquad B_u(t) = \frac{\partial f}{\partial u}(t, x_u(t), u(t)) \in \mathbb{R}^{d\times k}. \tag{2.23}$$

Lemme 2.15 (Différentiabilité)**.** *L'application entrée-sortie E_{T,x_0} est **différentiable** (au sens de Fréchet) en tout $u \in \mathcal{U}_{T,x_0}$ et sa différentielle $E'_{T,x_0}(u) : L^\infty([0,T];\mathbb{R}^k) \to \mathbb{R}^d$ est l'application entrée-sortie du système linéarisé le long de la trajectoire x_u ; plus explicitement, pour tout $\delta u \in L^\infty([0,T];\mathbb{R}^k)$, on a*

$$\langle E'_{T,x_0}(u), \delta u\rangle = \delta x(T), \tag{2.24}$$

où δx est solution du système différentiel linéarisé (2.22)*.*

Remarque 2.16. [Continuité] La différentielle $E'_{T,x_0}(u)$ est bien une forme linéaire continue en δu car on a

$$\langle E'_{T,x_0}(u), \delta u\rangle = R(T)\int_0^T R(s)^{-1}B_u(s)\delta u(s)\,\mathrm{d}s,$$

où $R(t)$ est la résolvante du système linéarisé, i.e., la solution matricielle dans $\mathbb{R}^{d\times d}$ de $\dot{R}(t) = A_u(t)R(t)$, pour tout $t \in [0,T]$, et $R(0) = I_d$. On a donc bien $|\langle E'_{T,x_0}(u), \delta u\rangle| \leq C\|\delta u\|_{L^\infty([0,T];\mathbb{R}^k)}$. En outre, $E'_{T,x_0}(u)$ dépend continûment de u. □

Démonstration. Nous nous contentons d'esquisser la preuve. Soit $\delta u \in V = L^\infty([0,T];\mathbb{R}^k)$ tel que $u+\delta u \in \mathcal{U}_{T,x_0}$ (qui est ouvert dans V). On note $x_{u+\delta u}$ la trajectoire associée à $u+\delta u$ issue de x_0. En effectuant des développements de Taylor sur f, il vient

$$\begin{aligned}\dot{x}_{u+\delta u}(t) - \dot{x}_u(t) &= f(t, x_{u+\delta u}(t), u(t)+\delta u(t)) - f(t, x_u(t), u(t)) \\ &= \tfrac{\partial f}{\partial x}(t, x_u(t), u(t))(x_{u+\delta u}(t) - x_u(t)) + \tfrac{\partial f}{\partial u}(t, x_u(t), u(t))\delta u(t) + o(\delta u) \\ &= A_u(t)(x_{u+\delta u}(t) - x_u(t)) + B_u(t)\delta u(t) + o(\delta u),\end{aligned}$$

car $x_{u+\delta u} - x_u = O(\delta u)$ (dépendance continue en un paramètre de la solution d'un système différentiel). En posant $\epsilon(t) = x_{u+\delta u}(t) - x_u(t) - \delta x(t)$, on en déduit que $\epsilon(0) = 0$ et que

$$\begin{aligned}\dot{\epsilon}(t) &= \dot{x}_{u+\delta u}(t) - \dot{x}_u(t) - \dot{\delta x}(t) \\ &= A_u(t)(x_{u+\delta u}(t) - x_u(t) - \delta x(t)) + o(\delta u) = A_u(t)\epsilon(t) + o(\delta u).\end{aligned}$$

Par des arguments de stabilité, on montre que $\epsilon = o(\delta u)$. En conclusion, on obtient

$$\begin{aligned}E_{T,x_0}(u+\delta u) - E_{T,x_0}(u) &= x_{u+\delta u}(T) - x_u(T) \\ &= \delta x(T) + \epsilon(T) = \delta x(T) + o(\delta u),\end{aligned}$$

et on a vu que $\delta u \mapsto \delta x(T)$ définit une forme linéaire continue sur δu pour la topologie de V. Ceci conclut la preuve. □

Théorème 2.17 (Contrôlabilité locale)**.** *Si le système différentiel linéarisé le long de la trajectoire x_u est **contrôlable** (en temps T), alors le système différentiel non-linéaire est **localement contrôlable** (en temps T à partir de x_0).*

Démonstration. Si le système différentiel linéarisé est contrôlable, alors la différentielle de l'application entrée-sortie E'_{T,x_0} est surjective. On conclut par le théorème de la submersion rappelé ci-dessous (qui est une variante du théorème des fonctions implicites, voir par exemple la référence [7]). □

Théorème 2.18 (Submersion)**.** *Soit V et W deux espaces de Banach, et $F : V \to W$ une application continûment différentiable. Soit $v \in V$. Si l'application différentielle $F'(v) : V \to W$ est surjective, alors F est localement surjective au voisinage de $F(v) \in W$.*

Remarque 2.19. [Point d'équilibre] On considère le cas particulier d'un point d'équilibre d'un système différentiel autonome, i.e., un couple (x_0, u_0) tel que $f(x_0, u_0) = 0$. Noter que $x_0 \in \mathcal{A}(t, x_0)$ en utilisant le contrôle constant égal à u_0. Le critère de contrôlabilité locale en x_0 consiste à vérifier que les matrices $A = \frac{\partial f}{\partial x}(x_0, u_0)$ et $B = \frac{\partial f}{\partial u}(x_0, u_0)$ vérifient la condition de Kalman. En effet, comme $f(x_0, u_0) = 0$, la trajectoire de référence est réduite à un point, si bien que le système linéarisé est également autonome, et on peut appliquer la condition de Kalman pour en vérifier la contrôlabilité. □

Remarque 2.20. [Inversion du temps] En cas de contrôlabilité locale et lorsque la dynamique est autonome et de la forme $f(x,u) = ug(x)$ (en supposant pour simplifier u à valeurs scalaires),

on déduit par inversion du temps que pour tout $y \in \mathcal{A}(T, x_0)$ tel que $V_y \subset \mathcal{A}(T, x_0)$, on peut ramener tout point $y' \in V_y$ à x_0. En effet, en notant u' le contrôle amenant x_0 en y' en temps T, on pose $\tilde{u}'(t) = -u'(T-t)$ et on vérifie que $\tilde{x}(t) = x_{u'}(T-t)$ vérifie bien $\tilde{x}(0) = y'$, $\tilde{x}(T) = x_0$ et $\frac{d}{dt}\tilde{x}(t) = -\frac{d}{dt}x_{u'}(T-t) = -u'(T-t)g(x_{u'}(T-t)) = \tilde{u}'(t)g(\tilde{x}(t))$, ce qui montre que $\tilde{x}$ est bien la trajectoire associée au contrôle $\tilde{u}'$. □

Exemple 2.21. [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 ($m = 1$, $l = 1$). 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

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

En posant $x = (x_1, x_2) = (\theta, \dot{\theta}) \in \mathbb{R}^2$, on se ramène à un système d'ordre un :

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

On calcule

$$\frac{\partial f}{\partial x}(x, u) = \begin{pmatrix} 0 & 1 \\ \cos(x_1) + u\sin(x_1) & 0 \end{pmatrix}, \qquad \frac{\partial f}{\partial u}(x, u) = \begin{pmatrix} 0 \\ -\cos(x_1) \end{pmatrix}.$$

On considère le point d'équilibre instable $(x_0, u_0) = ((0,0)^\dagger, 0)$. Le système linéarisé autour de ce point s'écrit sous la forme $\dot{\delta x}(t) = A\delta x(t) + B\delta u(t)$ avec

$$A = \frac{\partial f}{\partial x}(x_0, u_0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = \frac{\partial f}{\partial u}(x_0, u_0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

La condition de Kalman est bien satisfaite car

$$C = (B, AB) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

On a donc montré que le pendule inversé est **localement contrôlable** autour de son point d'équilibre instable $(x_0, u_0) = ((0,0)^\dagger, 0)$. Enfin, en adaptant le raisonnement présenté à la remarque 2.20, on montre qu'on peut ramener tout point au voisinage du point d'équilibre instable vers ce point. □

2.4 Rappels/compléments : topologie faible, différentielle, sélection mesurable

L'objectif de cette section est de rappeler quelques notions utiles sur la topologie faible dans les espaces de Hilbert et la différentielle de Fréchet dans les espaces de Banach d'une part et d'apporter quelques compléments sur les résultats de sélection mesurable qui sont parfois invoqués dans ce cours d'autre part.

2.4.1 Topologie faible dans les espaces de Hilbert

Soit V un espace de Hilbert de produit scalaire noté $(\cdot,\cdot)_V$. On se contente ici de rappeler les définitions et résultats qui nous seront utiles ; pour des compléments, le lecteur pourra consulter les chapitres 3 et 5 de la référence [3].

Définition 2.22 (Convergence faible)**.** *On dit qu'une suite $(v_n)_{n\in\mathbb{N}}$ de V* ***converge faiblement*** *vers $v \in V$ si*

$$\lim_{n\to+\infty} (v_n, \varphi)_V = (v, \varphi)_V \text{ (dans } \mathbb{R}) \qquad \forall \varphi \in V. \tag{2.25}$$

L'inégalité de Cauchy–Schwarz montre que si la suite $(v_n)_{n\in\mathbb{N}}$ converge fortement vers v (i.e., $\lim_{n\to+\infty} \|v_n - v\|_V = 0$), alors la suite $(v_n)_{n\in\mathbb{N}}$ converge faiblement vers v.

Théorème 2.23 (Compacité faible)**.** *Si $(v_n)_{n\in\mathbb{N}}$ est une suite* ***bornée*** *dans V, on peut en extraire une sous-suite* ***faiblement convergente****.*

Théorème 2.24 (Fermeture faible des convexes)**.** *Soit K un sous-ensemble fermé non-vide de l'espace de Hilbert V. On suppose que K est* ***convexe****. Alors, K est* ***fermé pour la topologie faible****. En d'autres termes, si $(v_n)_{n\in\mathbb{N}}$ est une suite de K qui converge faiblement vers v dans V, alors $v \in K$.*

2.4.2 Différentielle de Fréchet

Soit V un espace de Banach de norme $\|\cdot\|_V$. On rappelle que l'espace dual V' est composé des formes linéaires **continues** sur V, i.e., $\phi \in V'$ est une application linéaire $\phi : V \to \mathbb{R}$ telle que

$$\exists C > 0, \quad |\langle \phi, v\rangle| \le C \|v\|_V, \quad \forall v \in V. \tag{2.26}$$

Définition 2.25 (Différentielle de Fréchet)**.** *Soit V un espace de Banach et $J : V \to \mathbb{R}$ une application. On dit que J est* ***différentiable (au sens de Fréchet)*** *en $v \in V$ s'il existe une forme linéaire continue*

$$J'(v) \in V' \tag{2.27}$$

telle que

$$J(v + \delta v) = J(v) + \langle J'(v), \delta v\rangle + o(\delta v), \qquad \forall \delta v \in V, \tag{2.28}$$

où la notation $o(\delta v)$ signifie que $\lim_{\delta v\to 0} \frac{o(\delta v)}{\|\delta v\|_V} = 0$.

Dans le cas où V est un espace de Hilbert, on peut utiliser le théorème de représentation de Riesz pour identifier la forme linéaire continue $J'(v) \in V'$ avec son représentant

$$\nabla J(v) \in V. \tag{2.29}$$

En notant $(\cdot,\cdot)_V$ le produit scalaire dans V, on a

$$\langle J'(v), \delta v\rangle = (\nabla J(v), \delta v)_V, \qquad \forall \delta v \in V. \tag{2.30}$$

En dimension finie avec $V = \mathbb{R}^d$, on écrit $J(v_1, \dots, x_d) \in \mathbb{R}$ ainsi que

$$J'(v) = \left(\frac{\partial J}{\partial v_1}, \dots, \frac{\partial J}{\partial v_d}\right), \qquad \nabla J(v) = \begin{pmatrix} \frac{\partial J}{\partial v_1} \\ \vdots \\ \frac{\partial J}{\partial v_d} \end{pmatrix}. \tag{2.31}$$

Exemple 2.26. [Fonctionnelle quadratique] On considère $V = L^2(\Omega)$ et $J(v) = \frac{1}{2}\int_\Omega v(x)^2 \,\mathrm{d}x$. On a

$$\begin{aligned} J(v+\delta v) &= \frac{1}{2}\int_\Omega v(x)^2 \,\mathrm{d}x + \int_\Omega v(x)\delta v(x)\,\mathrm{d}x + \frac{1}{2}\int_\Omega \delta v(x)^2 \\ &= J(v) + \int_\Omega v(x)\delta v(x)\,\mathrm{d}x + o(\delta v). \end{aligned}$$

D'où $\langle J'(v), \delta v\rangle = \int_\Omega v(x)\delta v(x)\,\mathrm{d}x$ et $\nabla J(v) = v$. □

Plus généralement, soit V, W deux espaces de Banach et $J : V \to W$ une application. On dit que J est **différentiable (au sens de Fréchet)** en $v \in V$ s'il existe une application linéaire continue $J'(v) : V \to W$ telle que

$$J(v+\delta v) = J(v) + J'(v)(\delta v) + o(\delta v)\ (\in W), \qquad \forall \delta v \in V, \tag{2.32}$$

avec $\lim_{\delta v \to 0} \frac{\|o(\delta v)\|_W}{\|\delta v\|_V} = 0$.

2.4.3 Sélection mesurable

Les résultats de sélection mesurable qui sont brièvement présentés dans cette sous-section jouent un rôle important dans la justification mathématique rigoureuse de divers résultats de contrôle optimal. Ces résultats font appel à des notions relativement fines de théorie de la mesure, et ne seront donc qu'esquissés ici. Une présentation complète peut être trouvée dans le chapitre 14 du livre [9]. Le contenu de cette sous-section est inspiré de ce chapitre.

Commençons par présenter la problématique. On pose $I = [0,T]$. On considère une application $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}} = [-\infty, +\infty]$. Pour tout $t \in I$, on considère le sous-ensemble

$$\overline{U}(t) = \arg\min_{u \in \mathbb{R}^k} \Phi(t,u) \subset \mathbb{R}^k, \tag{2.33}$$

et on pose $J = \{t \in I \mid \overline{U}(t) \neq \emptyset\}$. On souhaite savoir s'il existe une application $\overline{u} : J \to \mathbb{R}^k$ qui soit **mesurable** et telle que $\overline{u}(t) \in \overline{U}(t)$ pour tout $t \in J$. Une telle application est appelée une **sélection mesurable**. Un résultat simple et utile est que si l'application Φ est **mesurable** par rapport à t (à u fixé) et si elle est **convexe et continue** par rapport à u (à t fixé), alors il existe une telle sélection mesurable.

Le reste de cette sous-section a pour objectif d'apporter une réponse mathématique un peu plus complète au problème de la sélection mesurable. Dans un premier temps, on considère des applications définies sur I à valeurs dans les sous-ensembles de $\mathbb{R}^k$. On note $S : I \rightrightarrows \mathbb{R}^k$ une telle application (le symbole $\rightrightarrows$ est là pour nous rappeler que $S(t)$ est un sous-ensemble de $\mathbb{R}^k$ qui n'est pas forcément réduit à un point). On équipe I d'une σ-algébre notée $\mathcal{A}$ (par exemple, la tribu borélienne de $\mathbb{R}$ restreinte à I).

Définition 2.27 (Mesurabilité)**.** *On dit que l'application $S : I \rightrightarrows \mathbb{R}^k$ est mesurable si pour tout ouvert $O \subset \mathbb{R}^k$, l'image réciproque*

$$S^{-1}(O) = \bigcup_{u \in O} S^{-1}(u) = \{t \in I \mid S(t) \cap O \neq \emptyset\} \tag{2.34}$$

est mesurable, i.e., si $S^{-1}(O) \in \mathcal{A}$. En particulier, le domaine de S, $\operatorname{dom} S = S^{-1}(\mathbb{R}^k)$, est donc mesurable (on notera que si $S(t) = \emptyset$, alors $t \notin \operatorname{dom} S$).

Si l'application S ne prend comme valeurs que des singletons, on retrouve la définition usuelle de la mesurabilité d'une application de I dans $\mathbb{R}^k$.

Théorème 2.28 (Représentation de Castaing)**.** *La mesurabilité d'une application $S : I \rightrightarrows \mathbb{R}^k$ à **valeurs fermées** (cela signifie que pour tout $t \in I$, $S(t)$ est un fermé) est équivalente à l'existence d'une représentation de Castaing, i.e., à l'existence d'une famille dénombrable de fonctions mesurables $s_n : \operatorname{dom} S \to \mathbb{R}^k$, $\forall n \in \mathbb{N}$, telles que pour tout $t \in \operatorname{dom} S$, $S(t) = \overline{\{s_n(t)\}_{n\in\mathbb{N}}}$.*

Corollaire 2.29 (Sélection mesurable)**.** *Une application $S : I \rightrightarrows \mathbb{R}^k$ mesurable à valeurs fermées admet une sélection mesurable, i.e., il existe une application mesurable $s : \operatorname{dom} S \to \mathbb{R}^k$ telle que $s(t) \in S(t)$ pour tout $t \in \operatorname{dom} S$.*

Considérons à nouveau une application $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}}$. L'application-épigraphe $\mathcal{E}_\Phi : I \rightrightarrows \mathbb{R}^k \times \mathbb{R}$ et l'application-domaine $\mathcal{D}_\Phi : I \rightrightarrows \mathbb{R}^k$, associées à Φ, sont telles que, pour tout $t \in I$,

$$\mathcal{E}_\Phi(t) = \{(u, \alpha) \in \mathbb{R}^k \times \mathbb{R} \mid \Phi(t, u) \leq \alpha\}, \tag{2.35a}$$

$$\mathcal{D}_\Phi(t) = \{u \in \mathbb{R}^k \mid \Phi(t, u) < +\infty\}. \tag{2.35b}$$

Définition 2.30 (Intégrande normal)**.** *On dit que l'application $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}}$ est un intégrande normal si son application-épigraphe $\mathcal{E}_\Phi : I \rightrightarrows \mathbb{R}^k \times \mathbb{R}$ est mesurable à valeurs fermées.*

Proposition 2.31 (Ensembles de niveau)**.** *L'application $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}}$ est un intégrande normal si et seulement si pour tout $\alpha \in \overline{\mathbb{R}}$, l'application ensemble de niveau $N_\alpha : I \rightrightarrows \mathbb{R}^k$ telle que $N_\alpha(t) = \{u \in \mathbb{R}^k \mid \Phi(t, u) \leq \alpha\}$ est mesurable à valeurs fermées.*

On rappelle qu'une fonction $f : \mathbb{R}^k \to \overline{\mathbb{R}}$ est semi-continue inférieurement (sci en abrégé) si son épigraphe $\{(u, \alpha) \in \mathbb{R}^k \times \mathbb{R} \mid f(u) \leq \alpha\}$ est fermé ; de manière équivalente, pour tout $u \in \mathbb{R}^k$ et tout $\epsilon > 0$, il existe un voisinage U de u tel que pour tout $v \in U$, on a $f(v) \geq f(u) - \epsilon$.

Proposition 2.32 (Conséquences de la normalité d'un intégrande)**.** *On suppose que l'application $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}}$ est un intégrande normal. Alors,*
(i) *l'application-domaine $\mathcal{D}_\Phi : I \rightrightarrows \mathbb{R}^k$ est mesurable ;*
(ii) *pour toute fonction mesurable $I \ni t \mapsto u(t) \in \mathbb{R}^k$, la fonction $t \mapsto \Phi(t, u(t))$ est mesurable ;*
(iii) *l'application Φ est mesurable par rapport à t (à u fixé) et elle est sci par rapport à u (à t fixé) ; en revanche, toute application qui est mesurable par rapport à t et sci par rapport u n'est pas nécessairement un intégrande normal.*

Proposition 2.33 (Fonction de Carathéodory)**.** *Toute fonction de Carathéodory, i.e., toute fonction qui est mesurable par rapport à t (à u fixé) et continue par rapport à u (à t fixé) est un intégrande normal.*

Exemple 2.34. [Indicatrice] On suppose que l'application $S : I \rightrightarrows \mathbb{R}^k$ est mesurable et à valeurs fermées. Alors, la fonction indicatrice $\delta_S : I \times \mathbb{R}^k \to \overline{\mathbb{R}}$ telle que

$$\delta_S(t,u) = \begin{cases} 0 & \text{si } u \in S(t), \\ +\infty & \text{sinon,} \end{cases}$$

est un intégrande normal. □

Venons-en au résultat principal lié à la notion d'intégrande normal.

Théorème 2.35 (Mesurabilité de minimiseurs et du minimum)**.** *On suppose que l'application $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}}$ est un intégrande normal. On pose pour tout $t \in I$,*

$$\varphi(t) = \inf_{u \in \mathbb{R}^k} \Phi(t,u), \quad \overline{U}(t) = \arg\min_{u \in \mathbb{R}^k} \Phi(t,u). \tag{2.36}$$

Alors, l'application $\varphi : I \to \overline{\mathbb{R}}$ est mesurable et l'application $\overline{U} : I \rightrightarrows \mathbb{R}^k$ est mesurable à valeurs fermées. Par conséquent, le sous-ensemble $J = \{t \in I \mid \overline{U}(t) \neq \emptyset\} \subset I$ est mesurable et pour tout $t \in J$, on peut choisir un minimiseur $\overline{u}(t)$ dans $\overline{U}(t)$ de sorte que l'application $t \mapsto \overline{u}(t)$ soit mesurable.

Proposition 2.36 (Convexité)**.** *Soit $\Phi : [0,T] \times \mathbb{R}^k \to \overline{\mathbb{R}}$ une application mesurable par rapport à t et sci par rapport à u. Alors, si Φ est convexe par rapport à u (à t fixé), Φ est un intégrande normal.*

Chapitre 3

Optimisation dans les espaces de Hilbert

Ce chapitre est consacré à l'optimisation de fonctionnelles, éventuellement sous contraintes, dans les espaces de Hilbert. Afin de motiver cette problématique, nous commencerons par étudier un problème de contrôle optimal très simple où la dynamique est linéaire (et autonome) et le critère à minimiser est quadratique en le contrôle ; nous verrons que dans ce cas, il est relativement aisé de produire un contrôle optimal. Le cœur de ce chapitre contient divers résultats abstraits d'optimisation qui serviront à plusieurs reprises dans ce cours et où la notion de **convexité** joue un rôle central. Il s'agit d'une part de résultats nous permettant d'affirmer l'**existence**, voire l'**unicité**, d'un minimiseur et d'autre part de **conditions nécessaires**, voire **suffisantes**, d'optimalité faisant intervenir la notion de différentielle. Enfin, nous donnerons un exemple d'application important de ces résultats abstraits en traitant le problème de **temps-optimalité** pour un système de contrôle linéaire ; ce problème consiste à trouver un contrôle permettant d'atteindre une cible atteignable donnée en temps minimum.

3.1 Contrôle optimal sous critère quadratique

Le but de cette section est de présenter un exemple relativement simple de problème de contrôle optimal afin de motiver les résultats qui suivront sur l'optimisation dans les espaces de Hilbert.

On considère le système de contrôle linéaire autonome (1.1), à savoir

$$\dot{x}_u(t) = Ax_u(t) + Bu(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0 \in \mathbb{R}^d, \tag{3.1}$$

avec des matrices $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times k}$, où $d \geq 1$ et $k \geq 1$. L'état du système est décrit par la fonction $x_u : [0,T] \to \mathbb{R}^d$ et on considère des contrôles $u : [0,T] \to \mathbb{R}^k$ dans l'espace de Hilbert

$$U = L^2([0,T];\mathbb{R}^k). \tag{3.2}$$

On se donne une cible $x_1 \in \mathbb{R}^d$ et on suppose que les matrices A et B vérifient la condition de Kalman (cf. le théorème 1.8) si bien que le système de contrôle linéaire (3.1) est contrôlable.

En d'autres termes, il existe des contrôles $u \in U$ tels que $x_u(T) = x_1$. On va chercher parmi tous ces contrôles permettant d'atteindre la cible x_1 (en temps T à partir de x_0) celui (ou ceux) qui minimise(nt) le critère quadratique suivant :

$$J(u) = \int_0^T |u(t)|^2_{\mathbb{R}^k} \, \mathrm{d}t. \tag{3.3}$$

On cherche donc un **contrôle optimal** $\overline{u} \in U$ amenant l'état en x_1 (en temps T en partant de x_0), i.e., tel que pour tout autre contrôle $u \in U$ ayant les mêmes propriétés, on a $J(u) \geq J(\overline{u})$. Afin de formaliser ce problème de contrôle optimal, on introduit le sous-ensemble $K \subset U$ tel que

$$K = \left\{ u \in U \mid \int_0^T e^{(T-s)A} B u(s) \, \mathrm{d}s = x_1 - e^{TA} x_0 \right\}, \tag{3.4}$$

et la **fonctionnelle** $J : U \to \mathbb{R}$ définie en (3.3). On pourra noter au passage que la fonctionnelle J est ici particulièrement simple puisque l'on a $J(u) = \|u\|^2_V$. Le problème de contrôle optimal s'écrit sous la forme suivante :

$$\text{Chercher } \overline{u} \in K \text{ tel que } J(\overline{u}) = \inf_{u \in K} J(u). \tag{3.5}$$

Nous serons amenés à nous poser les questions suivantes concernant le problème de contrôle optimal (3.5), et plus généralement les problèmes de contrôle optimal rencontrés dans ce cours :

(Q1) existe-t-il une solution, i.e., un contrôle optimal ?

(Q2) cette solution est-elle unique ?

(Q3) peut-on formuler une condition **suffisante** d'optimalité, i.e., nous permettant d'affirmer que si un contrôle $\overline{u} \in K$ vérifie cette condition, alors $\overline{u}$ est un contrôle optimal ?

Pour des problèmes de contrôle optimal relativement simples, comme celui considéré ci-dessus, nous serons en mesure d'apporter une réponse complète à ces questions. Pour des problèmes plus compliqués, nous devrons souvent nous contenter de traiter la question suivante :

(Q4) peut-on formuler une condition **nécessaire** d'optimalité, i.e., nous permettant d'affirmer que si $\overline{u} \in K$ est un contrôle optimal, alors il vérifie cette condition.

L'intérêt pratique d'une condition nécessaire d'optimalité est qu'elle nous permet d'effectuer un premier tri parmi les contrôles dans K. Dans le cas favorable où ce premier tri nous permet d'identifier un nombre relativement restreint de contrôles candidats à l'optimalité, on pourra ensuite vérifier la valeur du critère pour chacun d'entre eux et ainsi trouver un contrôle optimal.

Concluons cette section en exhibant un contrôle optimal dans K pour le problème (3.5). Nous montrerons ultérieurement, grâce aux résultats abstraits de la section suivante, l'unicité du contrôle optimal pour le problème (3.5). Cela nous permettra alors d'affirmer que nous avons trouvé **le** contrôle optimal pour le problème (3.5). On considère la matrice $G_T \in \mathbb{R}^{d \times d}$ définie en (1.17), i.e.,

$$G_T = \int_0^T e^{(T-s)A} B B^\dagger e^{(T-s)A^\dagger} \, \mathrm{d}s. \tag{3.6}$$

Comme le système de contrôle linéaire (3.1) est contrôlable par hypothèse, la matrice G_T est **inversible** (cf. le lemme 1.13). On pose

$$\overline{u}(s) = B^\dagger e^{(T-s)A^\dagger} y, \qquad y = G_T^{-1}(x_1 - e^{TA}x_0). \tag{3.7}$$

Lemme 3.1 (Synthèse d'un contrôle optimal)**.** *Le contrôle $\overline{u}$ défini par* (3.7) *est solution de* (3.5).

Démonstration. Nous devons vérifier que $\overline{u} \in K$ et que $J(\overline{u}) \leq J(u)$ pour tout $u \in K$.
(1) On a bien $\overline{u} \in U = L^2([0,T];\mathbb{R}^k)$ et en utilisant la formule de Duhamel, il vient

$$x_{\overline{u}}(T) = e^{TA}x_0 + \int_0^T e^{(T-s)A} B\overline{u}(s)\,\mathrm{d}s = e^{TA}x_0 + G_T y = x_1. \tag{3.8}$$

Ceci montre que $\overline{u} \in K$.
(2) Soit $u \in K$, i.e., $u \in L^2([0,T];\mathbb{R}^k)$ et $x_u(T) = x_1$. En posant $\delta = u - \overline{u}$, on constate par linéarité que le contrôle δ amène la condition initiale 0 à la cible 0 en temps T, i.e., on a $\int_0^T e^{(T-s)A} B\delta(s)\,\mathrm{d}s = 0$. En développant, il vient

$$\begin{aligned}
J(u) = J(\overline{u}+\delta) &= \int_0^T |\overline{u}(s) + \delta(s)|^2_{\mathbb{R}^k}\,\mathrm{d}s \\
&= J(\overline{u}) + 2\int_0^T \delta(s)^\dagger \overline{u}(s)\,\mathrm{d}s + J(\delta) \\
&= J(\overline{u}) + 2\int_0^T \delta(s)^\dagger (B^\dagger e^{(T-s)A^\dagger} y)\,\mathrm{d}s + J(\delta) \\
&= J(\overline{u}) + 2\int_0^T (e^{(T-s)A} B\delta(s))^\dagger y\,\mathrm{d}s + J(\delta) \\
&= J(\overline{u}) + 2\left(\int_0^T e^{(T-s)A} B\delta(s)\,\mathrm{d}s\right)^\dagger y + J(\delta) = J(\overline{u}) + J(\delta) \geq J(\overline{u}),
\end{aligned}$$

car $J(\delta) \geq 0$. Ceci montre que $J(\overline{u}) \leq J(u)$ pour tout $u \in K$. □

3.2 Minimisation de fonctionnelles

Soit V un **espace de Hilbert**, soit $K \subset V$ un sous-ensemble **fermé non-vide** de V. On désigne par $(\cdot,\cdot)_V$ le produit scalaire dans V et $\|\cdot\|_V$ la norme associée. Soit $J : K \to \mathbb{R}$ une **fonctionnelle** (i.e., une application de K dans $\mathbb{R}$; on utilise le terme fonctionnelle car bien souvent les éléments de V sont des fonctions, par exemple du temps). On considère le problème de minimiser la fonctionnelle J sur K, i.e.,

$$\text{Chercher } \overline{v} \in K \text{ tel que } J(\overline{v}) = \inf_{v \in K} J(v). \tag{3.9}$$

Il s'agit d'un problème de minimisation sous contraintes car on se restreint au sous-ensemble $K \subset V$. Lorsque $K = V$, on parle de problème de minimisation libre (ou sans contraintes), i.e., on considère le problème suivant :

$$\text{Chercher } \overline{v} \in V \text{ tel que } J(\overline{v}) = \inf_{v \in V} J(v). \tag{3.10}$$

3.2.1 Un premier exemple : projection sur un convexe

Définition 3.2 (Sous-ensemble convexe)**.** *Un sous-ensemble K d'un espace vectoriel V est dit* ***convexe*** *si*

$$\theta u + (1-\theta)v \in K, \quad \forall u, v \in K,\ \forall \theta \in [0,1]. \tag{3.11}$$

Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V. Pour tout $v \in V$, on cherche le point de K le plus proche de v. Ce point (nous verrons qu'il est unique) est appelé la projection de v sur K et est noté $\Pi_K(v)$. Le problème de la projection d'un point sur un convexe rentre dans le cadre du problème (3.9) en introduisant la fonctionnelle $J(w) = \|v - w\|_V$.

Proposition 3.3 (Projection sur un sous-ensemble convexe)**.** *Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V. Alors, pour tout $v \in V$, il existe un unique élément de K, noté $\Pi_K(v)$, tel que*

$$\|v - \Pi_K(v)\|_V = \inf_{w \in K} \|v - w\|_V, \quad \forall w \in V. \tag{3.12}$$

De plus, $\Pi_K(v)$ est l'unique point de K tel que

$$(v - \Pi_K(v), w - \Pi_K(v))_V \leq 0, \quad \forall w \in K. \tag{3.13}$$

On notera au passage que $d(v,K) = \inf_{w\in K} \|v - w\|_V = \|v - \Pi_K(v)\|_V$.

Démonstration. (1) Existence et unicité. On utilise l'identité suivante (dite formule de la médiane) :

$$\left\| \frac{y+z}{2} \right\|_V^2 = \frac{\|y\|_V^2 + \|z\|_V^2}{2} - \frac{1}{4}\|y - z\|_V^2, \quad \forall y, z \in V.$$

Soit $(w_n)_{n\in\mathbb{N}}$ une suite minimisante dans K, i.e., telle que $\|v - w_n\|_V \to \inf_{w\in K} \|v - w\|_V$ quand $n \to +\infty$. En appliquant la formule de la médiane à $y = v - w_n$ et $z = v - w_p$ pour tout $n, p \in \mathbb{N}$, on montre que la suite $(w_n)_{n\in\mathbb{N}}$ est de Cauchy dans V. Elle converge donc vers une limite $\ell \in V$. Comme K est fermé, $\ell \in K$. Ceci montre l'existence de la projection de v sur K. L'unicité résulte à nouveau de la formule de la médiane.
(2) Montrons l'identité (3.13). Pour tout $w \in K$ et $\theta \in]0,1]$, comme $\Pi_K(v) + \theta(w - \Pi_K(v)) \in K$ par convexité de K, on obtient

$$\begin{aligned} \|v - \Pi_K(v)\|_V^2 &\leq \|v - (\Pi_K(v) + \theta(w - \Pi_K(v)))\|_V^2 \\ &= \|v - \Pi_K(v)\|_V^2 - 2\theta(v - \Pi_K(v), w - \Pi_K(v))_V + \theta^2 \|w - \Pi_K(v)\|_V^2. \end{aligned}$$

En simplifiant par $\|v - \Pi_K(v)\|_V^2$ puis en divisant par θ, il vient

$$2(v - \Pi_K(v), w - \Pi_K(v))_V \leq \theta \|w - \Pi_K(v)\|_V^2,$$

et on conclut en faisant tendre θ vers zéro par valeurs positives. □

Corollaire 3.4 (Séparation d'un point et d'un convexe par un hyperplan)**.** *Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V. Soit $v \in V$ tel que $v \notin K$. Alors, il existe un hyperplan affine $H = \{w \in V \mid L(w) = \alpha\}$ où $L \in V'$ et $\alpha \in \mathbb{R}$ séparant v de K, i.e., tel que*

$$L(v) > \alpha, \quad K \subset \{w \in V \mid L(w) \leq \alpha\}. \tag{3.14}$$

Démonstration. Il suffit de considérer la forme linéaire continue $L(w) = (v - \Pi_K(v), w)_V$ et poser $\alpha = (v - \Pi_K(v), \Pi_K(v))_V$. La condition (3.13) signifie que pour tout $w \in K$, on a $L(w) \leq \alpha$. De plus, on a $L(v) - \alpha = \|v - \Pi_K(v)\|_V^2 > 0$ car $v \notin K$. □

3.2.2 Minimisation de fonctionnelles convexes sur des convexes

Définition 3.5 (Convexité, stricte convexité, forte convextié)**.** *Soit K un sous-ensemble* ***convexe*** *d'un espace de Hilbert V et soit $J : K \to \mathbb{R}$.*
(i) *On dit que J est* ***convexe*** *sur K si*

$$J(\theta u + (1-\theta)v) \leq \theta J(u) + (1-\theta)J(v), \quad \forall u, v \in K,\ \forall \theta \in [0,1]. \tag{3.15}$$

(ii) *On dit que J est* ***strictement convexe*** *sur K si l'inégalité* (3.15) *est stricte pour tout $u \neq v$ et tout $\theta \in]0,1[$.*
(iii) *On dit que J est* ***fortement convexe*** *ou α-**convexe** sur K s'il existe un réel $\alpha > 0$ tel que*

$$J\left(\frac{u+v}{2}\right) \leq \frac{J(u) + J(v)}{2} - \frac{\alpha}{8}\|u - v\|_V^2, \quad \forall u, v \in K. \tag{3.16}$$

Lorsque la fonctionnelle J est également continue, on a plus généralement

$$J(\theta u + (1-\theta)v) \leq \theta J(u) + (1-\theta)J(v) - \frac{\alpha}{2}\theta(1-\theta)\|u - v\|_V^2, \quad \forall u, v \in K,\ \forall \theta \in [0,1]. \tag{3.17}$$

Bien entendu, on a les implications suivantes :

$$\text{forte convexité} \implies \text{stricte convexité} \implies \text{convexité}. \tag{3.18}$$

Proposition 3.6 (Minoration de fonctionnelles convexes)**.** *Soit K un sous-ensemble fermé non-vide d'un espace de Hilbert V et soit $J : K \to \mathbb{R}$. On suppose que l'ensemble K est convexe et que la fonctionnelle J est continue.*
(i) *Si J est* ***convexe*** *sur K, il existe une forme linéaire continue $L \in V'$ et une constante $\delta \in \mathbb{R}$ telles que*

$$J(v) \geq L(v) + \delta, \quad \forall v \in K. \tag{3.19}$$

(ii) *Si J est* ***fortement convexe*** *sur K, il existe deux constantes $\gamma > 0$ et $\delta' \in \mathbb{R}$ telles que*

$$J(v) \geq \gamma \|v\|_V^2 + \delta', \quad \forall v \in K. \tag{3.20}$$

Démonstration. (1) Preuve de (3.19). Comme J est convexe et continue, son épigraphe, qui est défini comme l'ensemble

$$\mathcal{E} = \{(\lambda, v) \in \mathbb{R} \times K \mid \lambda \geq J(v)\} \tag{3.21}$$

est un sous-ensemble convexe, fermé et non-vide de $\mathbb{R} \times V$. Soit $v_0 \in K$ et $\lambda_0 < J(v_0)$, si bien que $(\lambda_0, v_0) \notin K$. En appliquant le corollaire 3.4, on déduit l'existence d'une paire $(\beta, L) \in \mathbb{R} \times V'$ et d'un réel α tels que

$$\beta\lambda_0 + L(v_0) \leq \alpha < \beta\lambda + L(v), \quad \forall (\lambda, v) \in \mathcal{E}.$$

Comme λ peut être arbitrairement grand au membre de droite, on doit avoir $\beta \geq 0$; de plus, en prenant $v = v_0$, on voit que $\beta \neq 0$. D'où $\beta > 0$. En prenant $\lambda = J(v)$, on en déduit que

$$J(v) = \lambda > \frac{\alpha}{\beta} - \frac{1}{\beta} L(v).$$

(2) Preuve de (3.20). Soit $v \in K$ arbitraire et soit $u \in K$ fixé. En appliquant (3.16) puis (3.19), on voit que

$$\begin{aligned} \frac{J(v) + J(u)}{2} &\geq J\left(\frac{v+u}{2}\right) + \frac{\alpha}{8}\|v-u\|_V^2 \\ &\geq \frac{L(v) + L(u)}{2} + \delta + \frac{\alpha}{8}\|v-u\|_V^2. \end{aligned}$$

On en déduit que

$$J(v) \geq \frac{\alpha}{4}\|v\|_V^2 + c_1\|v\|_V + c_2,$$

avec $c_1 = -\frac{\alpha}{2}\|u\|_V - \|L\|_{V'}$ et $c_2 = -J(u) + L(u) + 2\delta + \frac{\alpha}{4}\|u\|_V^2$. Comme $c_1\|v\|_V \geq -\frac{\alpha}{8}\|v\|_V^2 - \frac{2}{\alpha}c_1^2$, on obtient la minoration (3.20) avec $\gamma = \frac{\alpha}{8}$ et $\delta' = c_2 - \frac{2}{\alpha}c_1^2$. □

Théorème 3.7 (Minimisation de fonctionnelles convexes)**.** *Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V et soit $J : K \to \mathbb{R}$ une fonctionnelle convexe et continue sur K. On suppose que la fonctionnelle J est infinie à l'infini dans K, ce qui signifie que pour toute suite $(v_n)_{n\in\mathbb{N}}$ de K telle que $\|v_n\|_V \to +\infty$, on a $J(v_n) \to +\infty$. Alors, il* ***existe au moins un*** *minimiseur de J sur K.*

Démonstration. Comme J est infinie à l'infini, toute suite minimisante $(v_n)_{n\in\mathbb{N}}$ est bornée. Du théorème 2.23, on déduit qu'à une sous-suite près, cette suite converge faiblement vers une limite $v \in V$, et grâce au théorème 2.24, on montre que $v \in K$. Comme l'épigraphe de J (cf. (3.21)) est un ensemble convexe fermé, il est fermé pour la topologie faible et on conclut que $J(v) \leq \inf_{w\in K} J(w)$, ce qui prouve l'existence du minimiseur de J sur K. □

Théorème 3.8 (Minimisation de fonctionnelles fortement convexes)**.** *Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V et soit $J : K \to \mathbb{R}$ une fonctionnelle fortement convexe et continue sur K. Alors, il* ***existe un unique*** *minimiseur de J sur K.*

Démonstration. (1) Pour l'existence du minimiseur, on peut invoquer le théorème 3.7 et la minoration (3.20) qui montre que la fonctionnelle J est infinie à l'infini. On peut également donner une preuve directe en considérant une suite minimisante $(v_n)_{n\in\mathbb{N}}$ dans K. En utilisant (3.16), il vient

$$\begin{aligned}\frac{\alpha}{8}\|v_n - v_p\|_V^2 &\le \frac{J(v_n)+J(v_p)}{2} - J\left(\frac{v_n+v_p}{2}\right)\\ &\le \frac{1}{2}\left(J(v_n) - \inf_{w\in K} J(w)\right) + \frac{1}{2}\left(J(v_p) - \inf_{w\in K} J(w)\right),\end{aligned}$$

où nous avons utilisé la convexité de K pour obtenir $-J(\frac{v_n+v_p}{2}) \le -\inf_{w\in K} J(w)$. La majoration ci-dessus montre que la suite $(v_n)_{n\in\mathbb{N}}$ est de Cauchy dans V donc converge vers une limite $v \in V$ qui est dans K puisque K est fermé.
(2) L'unicité du minimiseur résulte à nouveau de (3.16) puisque si v_1 et v_2 sont deux minimiseurs de J sur K, en raisonnant comme ci-dessus, on obtient

$$\frac{\alpha}{8}\|v_1 - v_2\|_V^2 \le \frac{1}{2}\left(J(v_1) - \inf_{w\in K} J(w)\right) + \frac{1}{2}\left(J(v_2) - \inf_{w\in K} J(w)\right) = 0,$$

ce qui montre que $v_1 = v_2$. □

3.2.3 Conditions de minimalité

Nous allons maintenant nous intéresser à des conditions nécessaires de minimalité en supposant que la fonctionnelle J est différentiable sur K (au sens de Fréchet) et nous allons également voir dans quelles situations ces conditions nécessaires de minimalité sont également suffisantes.

On renvoie le lecteur à la sous-section 2.4.2 pour la notion de différentielle d'une fonctionnelle et quelques exemples importants (qu'il est essentiel de bien maîtriser !). Comme nous nous plaçons ici dans le cadre des espaces de Hilbert, nous allons privilégier la notion de gradient plutôt que celle de forme linéaire continue. Soit V un espace de Hilbert ; on désigne par $(\cdot,\cdot)_V$ le produit scalaire dans V et $\|\cdot\|_V$ la norme associée. Soit $J : K \to \mathbb{R}$ une fonctionnelle différentiable sur K, i.e., pour tout $v \in K$, il existe un élément de V noté $\nabla J(v)$ tel que

$$J(v+\delta v) = J(v) + (\nabla J(v), \delta v)_V + o(\delta v), \qquad \forall \delta v \in V,\ v+\delta v \in K, \tag{3.22}$$

où la notation $o(\delta v)$ signifie que $\lim_{\delta v\to 0} \frac{o(\delta v)}{\|\delta v\|_V} = 0$. Notons que si la fonctionnelle J est différentiable sur K, elle est *a fortiori* continue sur K.

Avant d'entrer dans le vif du sujet sur les conditions de minimalité, voyons comment la notion de différentielle nous permet d'étudier la convexité d'une fonctionnelle.

Proposition 3.9 (Caractérisation de la convexité)**.** *Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V et soit $J : K \to \mathbb{R}$ une fonctionnelle différentiable sur K. Soit $\alpha > 0$ un réel strictement positif. Les assertions suivantes sont équivalentes :*

(i) *J est fortement convexe de paramètre α ;*

(ii) $J(v) \geq J(u) + (\nabla J(u), v-u)_V + \frac{1}{2}\alpha\|u-v\|_V^2$, $\forall u, v \in K$;

(iii) $(\nabla J(u) - \nabla J(v), u-v)_V \geq \alpha\|u-v\|_V^2$, $\forall u, v \in K$.

En outre, la convexité de la fonctionnelle J est équivalente aux assertions (ii) *et* (iii) *avec* $\alpha = 0$.

Démonstration. (i) $\Rightarrow$ (ii). Pour tout entier $k \in \mathbb{N}$, en posant $\theta_k = 2^{-k}$, on montre par récurrence sur k à partir de (3.16) que pour tout $u, v \in K$, on a

$$J((1-\theta_k)u + \theta_k v) \leq (1-\theta_k)J(u) + \theta_k J(v) - \frac{\alpha}{2}\theta_k(1-\theta_k)\|u-v\|_V^2,$$

et en ré-arrangeant les différents termes, il vient

$$\frac{1}{\theta_k}\big(J(u+\theta_k(v-u)) - J(u)\big) \leq J(v) - J(u) - \frac{\alpha}{2}(1-\theta_k)\|u-v\|_V^2.$$

En faisant tendre $k \to +\infty$ et en utilisant la différentiabilité de J, on obtient la minoration de $J(v)$ dans (ii).
(ii) $\Rightarrow$ (iii). Il suffit d'écrire la propriété (ii) avec u et v, puis d'échanger les rôles de u et de v et de sommer.
(iii) $\Rightarrow$ (i). Soit $u, v \in K$. On définit la fonction $\psi : \mathbb{R} \to \mathbb{R}$ telle que $\psi(t) = J(u + t(v-u))$ pour tout $t \in \mathbb{R}$. On vérifie facilement que la fonction ψ est dérivable (et donc continue) sur $\mathbb{R}$ et on a $\psi'(t) = (\nabla J(u + t(v-u)), v-u)_V$. En utilisant la minoration (iii), il vient

$$(t-s)(\psi'(t)-\psi'(s)) = (\nabla J(u+t(v-u)) - \nabla J(u+s(v-u)), (t-s)(v-u))_V \geq \alpha(t-s)^2\|u-v\|_V^2.$$

Pour $s \leq t$, il vient $\psi'(t) - \psi'(s) \geq \alpha(t-s)\|u-v\|_V^2$. Soit $\theta \in [0,1]$. En intégrant cette inégalité pour $(t,s) \in [\theta, 1] \times [0, \theta]$, on obtient

$$\theta\psi(1) + (1-\theta)\psi(0) - \psi(\theta) \geq \frac{\alpha}{2}\theta(1-\theta)\|u-v\|_V^2,$$

qui n'est rien d'autre que (3.16) pour $\theta = \frac{1}{2}$. □

Proposition 3.10 (Équation d'Euler, minimisation sans contraintes)**.** *Soit V un espace de Hilbert et soit $J : V \to \mathbb{R}$ une fonctionnelle différentiable sur V. On considère le problème de minimisation sans contraintes* (3.10).
(i) *Une* ***condition nécessaire*** *de minimalité pour* (3.10) *est*

$$\nabla J(\overline{v}) = 0 \ (\in V). \tag{3.23}$$

Cette condition, appelée équation d'Euler, signifie que si $\overline{v} \in V$ est solution de (3.10)*, alors $\overline{v}$ vérifie* (3.23).
(ii) *Si la fonctionnelle J est* ***convexe****, la condition* (3.23) *est également* ***suffisante*** *; ceci signifie que si $\overline{v} \in V$ vérifie* (3.23)*, alors $\overline{v}$ est solution de* (3.10).

Démonstration. (1) Soit $\overline{v} \in V$ une solution de (3.10). Pour tout $\delta v \in V$, on a

$$0 \leq J(\overline{v} + \delta v) - J(\overline{v}) = (\nabla J(\overline{v}), \delta v)_V + o(\delta v).$$

On divise par $\|\delta v\|_V$ puis on fait tendre δv vers 0. On en déduit que $(\nabla J(\overline{v}), \delta v)_V \geq 0$; comme δv est arbitraire dans V, on peut changer δv en $-\delta v$. On conclut ainsi que $\nabla J(\overline{v}) = 0$ dans V, i.e., $\overline{v}$ vérifie (3.23).
(2) Supposons maintenant que la fonctionnelle J est convexe et que $\overline{v} \in V$ vérifie (3.23). En utilisant la proposition 3.9 dans le cas convexe ($\alpha = 0$), on en déduit que

$$J(v) \geq J(\overline{v}) + (\underbrace{\nabla J(\overline{v})}_{=0}, v - \overline{v})_V = J(\overline{v}), \quad \forall v \in V.$$

On conclut ainsi que $\overline{v}$ est solution de (3.10). □

Proposition 3.11 (Inéquation d'Euler, minimisation avec contraintes)**.** *Soit K un sous-ensemble convexe fermé non-vide d'un espace de Hilbert V. Soit $J : K \to \mathbb{R}$ une fonctionnelle différentiable sur K. On considère le problème de minimisation avec contraintes* (3.9).
(i) *Une* ***condition nécessaire*** *de minimalité pour* (3.9) *est*

$$(\nabla J(\overline{v}), v - \overline{v})_V \geq 0, \quad \forall v \in K. \tag{3.24}$$

Cette condition, appelée inéquation d'Euler, signifie que si $\overline{v} \in K$ est solution de (3.9)*, alors $\overline{v}$ vérifie* (3.24).
(ii) *Si la fonctionnelle J est* ***convexe****, la condition* (3.24) *est également* ***suffisante*** *; ceci signifie que si $\overline{v} \in K$ vérifie* (3.24)*, alors $\overline{v}$ est solution de* (3.9).

Démonstration. (1) Soit $\overline{v} \in V$ une solution de (3.9). Pour tout $v \in K$, $v \neq \overline{v}$, et tout $\theta \in]0, 1]$, on a $\overline{v} + \theta(v - \overline{v}) \in K$ car le sous-ensemble K est convexe. Par suite, il vient

$$0 \leq J(\overline{v} + \theta(v - \overline{v})) - J(\overline{v}) = \theta(\nabla J(\overline{v}), v - \overline{v})_V + o(\theta(v - \overline{v})).$$

On divise par $\theta\|v - \overline{v}\|_V$ puis on fait tendre θ vers 0 (par valeurs positives). On en déduit que $\overline{v}$ vérifie bien (3.24).
(2) La preuve du caractère suffisant de l'inéquation d'Euler dans le cas où J est une fonctionnelle convexe est identique au cas sans contraintes. □

Remarque 3.12. [Cas d'un point intérieur] Si $\bar{v}$ est situé à l'intérieur de l'ensemble K (on dit que $\bar{v}$ ne sature pas la contrainte), l'inéquation d'Euler (3.24) devient $\nabla J(\bar{v}) = 0$ ($\in V$). Cela résulte du fait qu'on peut prendre $v = \bar{v} + \rho z$ avec ρ suffisamment petit et z arbitraire dans V. □

Exemple 3.13. [Projection sur un convexe] Pour la projection d'un élément $v \in V$ sur un ensemble convexe K (cf. la proposition 3.3), la fonctionnelle à minimiser est $J(w) = \|v - w\|_V^2$. La fonctionnelle J est fortement convexe (de paramètre $\alpha = 2$) grâce à la formule de la médiane. Un calcul élémentaire montre que $\nabla J(w) = 2(w - v)$. On voit donc que la caractérisation (3.13) de la projection convexe n'est rien d'autre que l'inéquation d'Euler pour $\Pi_K(v)$ obtenue à la proposition 3.11. □

3.2.4 Application au contrôle optimal sous critère quadratique

Nous reprenons brièvement l'exemple du problème de contrôle optimal sous critère quadratique introduit à la section 3.1. Ce problème rentre dans le cadre abstrait de la section 3.2 en posant

$$V = L^2([0,T];\mathbb{R}^k), \tag{3.25a}$$

$$J : V \to \mathbb{R}, \quad J(u) = \int_0^T |u(t)|^2_{\mathbb{R}^k} \,\mathrm{d}t = \|u\|^2_V, \tag{3.25b}$$

$$K = \left\{ u \in V \mid \int_0^T e^{(T-s)A} B u(s) \,\mathrm{d}s = x_1 - e^{TA} x_0 \right\}. \tag{3.25c}$$

L'espace V est bien un espace de Hilbert. La fonctionnelle J est fortement convexe sur V de paramètre $\alpha = 2$ (utiliser la formule de la médiane). Enfin, K est un sous-ensemble convexe fermé non-vide de V. En effet,

— K est non-vide car nous avons supposé que le système de contrôle linéaire est contrôlable ;

— K est convexe car pour deux contrôles $u_1, u_2 \in K$ et pour tout $\theta \in [0,1]$, on a

$$\int_0^T e^{(T-s)A} B(\theta u_1(s) + (1-\theta) u_2(s)) \,\mathrm{d}s = \theta\Big(x_1 - e^{TA}x_0\Big) + (1-\theta)\Big(x_1 - e^{TA}x_0\Big) = x_1 - e^{TA}x_0;$$

— enfin, si $(u_n)_{n\in\mathbb{N}}$ est une suite de K qui converge vers u dans V, on a

$$x_1 - e^{TA}x_0 = \int_0^T e^{(T-s)A} B u_n(s) \,\mathrm{d}s \to \int_0^T e^{(T-s)A} B u(s) \,\mathrm{d}s,$$

ce qui montre que la limite u amène x_0 en x_1 en temps T, i.e., $u \in K$; le sous-ensemble K est donc bien fermé.

En appliquant le théorème 3.8, on en déduit qu'il existe un unique contrôle optimal $\overline{u} \in K$ minimisant J sur K. En outre, de par la proposition 3.11, une condition nécessaire et suffisante d'optimalité est

$$(\overline{u}, u - \overline{u})_V \geq 0, \quad \forall u \in K. \tag{3.26}$$

Or, lorsque u décrit K, le vecteur $h = u - \overline{u}$ décrit le sous-espace vectoriel

$$H = \left\{ h \in V \mid \int_0^T e^{(T-s)A} B h(s) \,\mathrm{d}s = 0 \right\}. \tag{3.27}$$

On a donc $(\overline{u}, h)_V \geq 0$ pour tout $h \in H$ et comme H est un sous-espace vectoriel, on peut considérer les vecteurs h et $-h$ dans l'inégalité ci-dessus, ce qui montre que $\overline{u} \in H^\perp$. Soit $(e_i)_{1\leq i\leq d}$ une base cartésienne de $\mathbb{R}^d$. Posons $y_i(t) = B^\dagger e^{(T-t)A^\dagger} e_i$; on a $y_i \in V$ pour tout $i \in \{1,\dots,d\}$ et par définition $H = (\mathrm{vect}(y_i)_{1\leq i\leq d})^\perp$. Par suite, $\overline{u} \in \mathrm{vect}(y_i)_{1\leq i\leq d}$ (qui est de dimension finie donc fermé) ; en d'autres termes, il existe un vecteur $y \in \mathbb{R}^d$ tel que

$$\overline{u}(t) = B^\dagger e^{(T-t)A^\dagger} y, \quad \forall t \in [0,T]. \tag{3.28}$$

Finalement, le vecteur $y \in \mathbb{R}^d$ s'obtient en imposant que $x_{\overline{u}}(T) = x_1$, et par la formule de Duhamel, on retrouve bien l'expression donnée dans l'équation (3.7).

3.3 Exemple : temps-optimalité (cas linéaire)

Dans cette section, on considère le système de contrôle linéaire autonome

$$\dot{x}_u(t) = Ax_u(t) + Bu(t), \quad \forall t \geq 0 \qquad x_u(0) = x_0, \tag{3.29}$$

avec des matrices $A \in \mathbb{R}^{d\times d}$ et $B \in \mathbb{R}^{d\times k}$. On suppose que le contrôle est à valeurs dans un sous-ensemble **compact non-vide** $U \subset \mathbb{R}^k$. On pose, pour tout $t > 0$,

$$\mathcal{U}_t = L^\infty([0,t]; U). \tag{3.30}$$

Le problème de **temps-optimalité** est le suivant : on se donne une **cible** $x_1 \in \mathbb{R}^d$, on suppose qu'il existe au moins une trajectoire reliant x_0 à x_1 en temps fini, et parmi toutes ces trajectoires, on cherche celle(s) qui le font en **temps minimal**. On rappelle que l'ensemble atteignable en temps t à partir de x_0 est défini comme suit (cf. la section 1.3) :

$$\mathcal{A}(t, x_0) = \{y \in \mathbb{R}^d \mid \exists u \in \mathcal{U}_t \text{ tel que } x_u(t) = y\}. \tag{3.31}$$

Nous avons vu (cf. la proposition 1.19) que pour tout $t \geq 0$, l'ensemble atteignable $\mathcal{A}(t, x_0)$ est compact, convexe, et varie continûment en temps. Le problème de temps-optimalité se formalise alors comme suit : étant donnée une cible $x_1 \in \mathbb{R}^d$ atteignable en temps fini, i.e., telle que l'ensemble $\{t \geq 0 \mid x_1 \in \mathcal{A}(t, x_0)\}$ est non-vide, on cherche

$$t_* = \inf\{t \geq 0 \mid x_1 \in \mathcal{A}(t, x_0)\}. \tag{3.32}$$

Comme l'ensemble $\mathcal{A}(t, x_0)$ varie continûment en t, l'ensemble $\{t \geq 0 \mid x_1 \in \mathcal{A}(t, x_0)\}$ est fermé dans $\mathbb{R}$, si bien que la borne inférieure t_* est atteinte (considérer une suite minimisante $t_n \downarrow t_*$ et constater que $0 = d(x_1, \mathcal{A}(t_n, x_0)) \to d(x_1, \mathcal{A}(t_*, x_0))$ de par la dépendance continue en temps).

Lemme 3.14 (Atteinte à la frontière)**.** *On a $x_1 \in \partial\mathcal{A}(t_*, x_0)$.*

Démonstration. On raisonne par l'absurde en supposant que $x_1 \in \mathring{\mathcal{A}}(t_*, x_0)$. Il existe donc un réel $\rho > 0$ tel que $B(x_1, \rho) \subset \mathcal{A}(t_*, x_0)$ où $B(x_1, \rho)$ désigne la boule ouverte de centre x_1 et de rayon ρ. Comme l'ensemble atteignable $\mathcal{A}(t, x_0)$ varie continûment en t, il existe $\delta > 0$ tel que $d(\mathcal{A}(t_*, x_0), \mathcal{A}(t_* - \delta, x_0)) \leq \frac{\rho}{2}$, i.e., pour tout $z \in \mathcal{A}(t_*, x_0)$, on peut trouver $y \in \mathcal{A}(t_* - \delta, x_0)$ tel que $|z - y|_{\mathbb{R}^d} \leq \frac{1}{2}\rho$. De plus, comme x_1 a été atteint en temps optimal et que $\delta > 0$, $x_1 \notin \mathcal{A}(t_* - \delta, x_0)$. L'ensemble $\mathcal{A}(t_* - \delta, x_0)$ étant convexe, le corollaire 3.4 implique qu'il existe un hyperplan affine H séparant x_1 de $\mathcal{A}(t_* - \delta, x_0)$. En notant $\Pi_H(x_1)$ la projection orthogonale de x_1 sur H, on a donc $\|x_1 - \Pi_H(x_1)\|_V > 0$. On considère le point

$$z = x_1 + \frac{\rho}{2}\frac{x_1 - \Pi_H(x_1)}{\|x_1 - \Pi_H(x_1)\|_V}.$$

Par construction, $\|z - x_1\|_V = \frac{\rho}{2}$, i.e., $z \in B(x_1, \rho)$, et donc $z \in \mathcal{A}(t_*, x_0)$. On peut donc trouver $y \in \mathcal{A}(t_* - \delta, x_0)$ tel que $|z - y|_{\mathbb{R}^d} \leq \frac{1}{2}\rho$. Or, par construction, on a

$$d(z, \mathcal{A}(t_* - \delta, x_0)) > \frac{1}{2}\rho,$$

d'où la contradiction. En conclusion, on a bien $x_1 \in \partial\mathcal{A}(t_*, x_0)$. □

Remarque 3.15. [Convexité] L'argument de convexité est essentiel dans la preuve ci-dessus. On pourra s'en persuader en considérant les ensembles $\mathcal{B}(t) = \{v \in V \mid 1 - t \leq \|v\|_V \leq 2\}$ pour tout $t \in [0,1]$. L'ensemble $\mathcal{B}(t)$ est fermé non-vide mais il n'est pas convexe pour $t < 1$; de plus, $\mathcal{B}(t)$ varie continûment en temps. Or, le point $v = 0$ appartient à $\mathcal{B}(1)$ mais pas à $\mathcal{B}(s)$ pour tout $s < 1$; or, $v = 0$ est situé à l'intérieur de $\mathcal{B}(1)$. □

Un raisonnement identique à celui de la preuve du lemme 3.14 s'applique pour tout $t \in [0,t_*]$: si $x_*(t)$ est la trajectoire associée à un contrôle temps-minimal, on a

$$x_*(t) \in \partial\mathcal{A}(t, x_0), \quad \forall t \in [0,t_*]. \tag{3.33}$$

Une illustration est présentée à la figure 3.1.

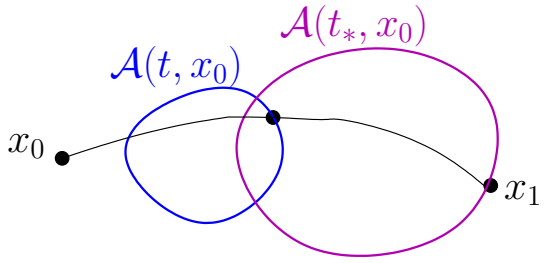

FIGURE 3.1 – Illustration d'une trajectoire associée à un contrôle temps-optimal, et plus généralement d'une trajectoire associée à un contrôle extrémal.

Définition 3.16 (Contrôle extrémal)**.** *On dit qu'un contrôle $u \in \mathcal{U}_t$ est **extrémal** si la trajectoire associée vérifie $x_u(s) \in \partial\mathcal{A}(s, x_0)$ pour tout $s \in [0,t]$.*

On a donc montré qu'un contrôle temps-optimal est nécessairement extrémal. La réciproque est bien sûr fausse car la notion d'extrémalité ne distingue pas entre minimalité et maximalité. Nous allons maintenant établir une condition nécessaire et suffisante d'extrémalité.

Théorème 3.17 (Condition nécessaire et suffisante d'extrémalité)**.** *On fixe un horizon temporel $T > 0$. On suppose que l'ensemble U est non-vide, compact et convexe. Alors, le contrôle $\overline{u} \in \mathcal{U}_T$ est extrémal sur $[0,T]$ **si et seulement si** il existe une solution non-triviale (i.e., qui n'est pas identiquement nulle) $p : [0,T] \to \mathbb{R}^d$ de l'équation*

$$\dot{p}(t) = -A^\dagger p(t), \quad t \in [0,T], \tag{3.34}$$

telle que

$$p(t)^\dagger B\overline{u}(t) = \min_{v \in U} p(t)^\dagger Bv, \quad p.p.\ t \in [0,T]. \tag{3.35}$$

*La fonction $p : [0,T] \to \mathbb{R}^d$ est appelée **état adjoint**.*

Remarque 3.18. [État adjoint] On notera que la condition initiale sur l'état adjoint p n'est pas spécifiée dans (3.34). On notera également que l'état adjoint est ici une fonction régulière du temps. En outre, seule la direction de l'état adjoint $p(t)$ compte dans (3.35), mais pas son amplitude. □

Remarque 3.19. [Contrôle extrémal] La condition nécessaire et suffisante d'extrémalité du théorème 3.17 montre que si le contrôle $\overline{u}$ est extrémal sur $[0,T]$, il l'est sur $[0,t]$ pour tout $t \in [0,T]$. □

Démonstration. (1) Condition nécessaire. Soit $u \in \mathcal{U}_T$ un contrôle extrémal et soit $x_u : [0,T] \to \mathbb{R}^d$ la trajectoire associée. Comme $x_u(T) \in \partial\mathcal{A}(T,x_0)$ et que l'ensemble atteignable $\mathcal{A}(T,x_0)$ est convexe (cf. la proposition 1.19), il existe un hyperplan séparant au sens large $x_u(T)$ et $\mathcal{A}(T,x_0)$, i.e.,

$$\exists p_T \in \mathbb{R}^d \setminus \{0\}, \quad p_T^\dagger(y - x_u(T)) \geq 0, \ \forall y \in \mathcal{A}(T,x_0). \tag{3.36}$$

Une illustration est présentée à la figure 3.2. En notant $\hat{u} \in \mathcal{U}_T$ un contrôle quelconque associé

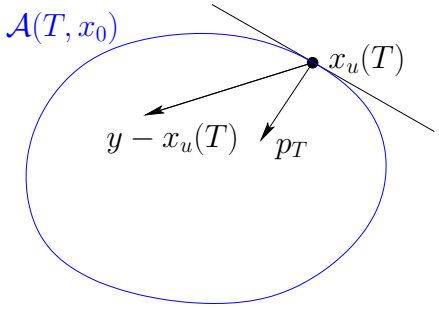

FIGURE 3.2 – Illustration de la séparation au sens large du point $x_u(T)$ et de l'ensemble convexe $\mathcal{A}(T,x_0)$.

à une trajectoire amenant au point $y \in \mathcal{A}(T,x_0)$, l'inégalité $p_T^\dagger(y - x_u(T)) \geq 0$ s'écrit

$$\int_0^T p_T^\dagger e^{(T-t)A} B\hat{u}(t)\,\mathrm{d}t \geq \int_0^T p_T^\dagger e^{(T-t)A} Bu(t)\,\mathrm{d}t.$$

En introduisant l'état adjoint tel que $\dot{p}(t) = -A^\dagger p(t)$, pour tout $t \in [0,T]$ et $p(T) = p_T$, l'inégalité ci-dessus se récrit

$$\int_0^T p(t)^\dagger B\hat{u}(t)\,\mathrm{d}t \geq \int_0^T p(t)^\dagger Bu(t)\,\mathrm{d}t.$$

On peut alors raisonner par l'absurde. Supposons que $p(t)^\dagger Bu(t) > \min_{v\in U} p(t)^\dagger Bv$ sur un sous-ensemble de $[0,T]$ de mesure strictement positive. Ceci implique que

$$\int_0^T p(t)^\dagger Bu(t)\,\mathrm{d}t > \int_0^T \min_{v\in U} p(t)^\dagger Bv\,\mathrm{d}t.$$

On considère alors un contrôle $\hat{u}$ sur $[0,T]$ à valeurs dans U tel que

$$p(t)^\dagger B\hat{u}(t) = \min_{v\in U} p(t)^\dagger Bv.$$

En invoquant un résultat de sélection mesurable (cf. la sous-section 2.4.3), on montre que $\hat{u}$ peut être choisi mesurable sur [0,T], i.e., on a bien $\hat{u} \in \mathcal{U}_T$. On a ainsi obtenu

$$\begin{aligned}\int_0^T \min_{v\in U} p(t)^\dagger Bv \,\mathrm{d}t &= \int_0^T p(t)^\dagger B\hat{u}(t)\,\mathrm{d}t \\ &\geq \int_0^T p(t)^\dagger Bu(t)\,\mathrm{d}t \\ &> \int_0^T \min_{v\in U} p(t)^\dagger Bv\,\mathrm{d}t,\end{aligned}$$

ce qui fournit la contradiction cherchée. En conclusion, on a bien (3.35).
(2) Condition suffisante. Supposons qu'il existe un état adjoint non trivial tel que le contrôle vérifie

$$p(t)^\dagger Bu(t) = \min_{v\in U} p(t)^\dagger Bv, \quad \text{p.p. } t \in [0,T].$$

En remontant les calculs précédents, on en déduit que

$$p(t)^\dagger(y - x_u(t)) \geq 0,\ \forall y \in \mathcal{A}(t,x_0),\ \forall t \in [0,T].$$

Si $x_u(t) \in \mathring{\mathcal{A}}(t,x_0)$, il existerait $\epsilon > 0$ tel que $x_u(t) - \epsilon p(t) \in \mathcal{A}(t,x_0)$; d'où

$$p(t)^\dagger(y - x(t)) = -\epsilon |p(t)|^2_{\mathbb{R}^d} \geq 0,$$

ce qui fournit la contradiction cherchée. En conclusion, on a bien $x_u(t) \in \partial\mathcal{A}(t,x_0)$. □

Remarque 3.20. [Fonction de commutation] Dans le cas mono-entrée (i.e., dans le cas d'un contrôle scalaire où $k = 1$), l'ensemble U où le contrôle peut prendre ses valeurs est un intervalle. Considérons pour simplifier le cas où $U = [-a,a]$ avec $a > 0$. La condition de minimisation (3.35) implique alors que

$$u(t) = -a\,\mathrm{signe}(p(t)^\dagger B).$$

La fonction

$$t \mapsto p(t)^\dagger B \in \mathbb{R}$$

est la **fonction de commutation**, et les temps t_c où le contrôle u change de valeur sont appelés les **temps de commutation**. Ces temps de commutation correspondent aux zéros de la fonction de commutation. On montre (voir par exemple la proposition 3.4 dans [11]) que dans le cas d'un système de contrôle linéaire autonome dont les matrices A et B vérifient la condition de Kalman et si toutes les valeurs propres de A sont réelles, alors tout contrôle extrémal a au plus $(d-1)$ commutations. □

Exemple 3.21. [Contrôle d'un tram] Reprenons l'exemple 1.11 du tram. On rappelle que le tram est repéré par sa position $x(t)$ le long d'un axe unidirectionnel et qu'on contrôle son accélération par le biais du contrôle $u(t)$. En considérant une masse unité, l'équation du mouvement est donc

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

En posant $X(t) = (x(t), v(t))^\dagger$ où $v(t) = \dot{x}(t)$, on obtient

$$\dot{X}(t) = AX(t) + Bu(t), \quad A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

On part d'une condition initiale $(x_0, v_0)^\dagger \in \mathbb{R}^2$ et on souhaite atteindre la cible $(0,0)^\dagger$ en temps minimal. En appliquant le théorème 3.17, on introduit l'état adjoint $p(t) = (p_x(t), p_v(t))^\dagger$ tel que $\dot{p}(t) = -A^\dagger p(t)$. Il vient $\dot{p}_x(t) = 0$, $\dot{p}_v(t) = -p_x(t)$, i.e.,

$$p_x(t) = p_{x0}, \quad p_v(t) = p_{v0} - p_{x0}t.$$

La condition de minimalité (3.35) s'écrit

$$u(t) = -\operatorname{signe}(p(t)^\dagger B) = -\operatorname{signe}(p_v(t)).$$

Comme la fonction p_v est affine en t, cela nous permet déjà de montrer qu'il y a au plus une commutation et que le contrôle temps-optimal est nécessairement bang-bang. Pour aller plus loin, on calcule les trajectoires dans l'espace des phases (i.e., dans le plan (x, v)).

— Si le contrôle est constant et égal à 1, on a $x(t) - \frac{1}{2}v(t)^2 = cste$ (car $\frac{d}{dt}(x(t) - \frac{1}{2}v(t)^2) = \dot{x}(t) - v(t)\dot{v}(t) = v(t) - v(t)u(t) = 0$) ; les trajectoires sont donc des paraboles d'axe Ox, parcourues dans le sens des v croissants.

— Si le contrôle est constant et égal à -1, on a $x(t) + \frac{1}{2}v(t)^2 = cste$; les trajectoires sont donc des paraboles d'axe $-Ox$, parcourues dans le sens des v décroissants.

Ces paraboles sont illustrées à la figure 3.3. Les deux demi-paraboles en rouge sur la figure 3.3

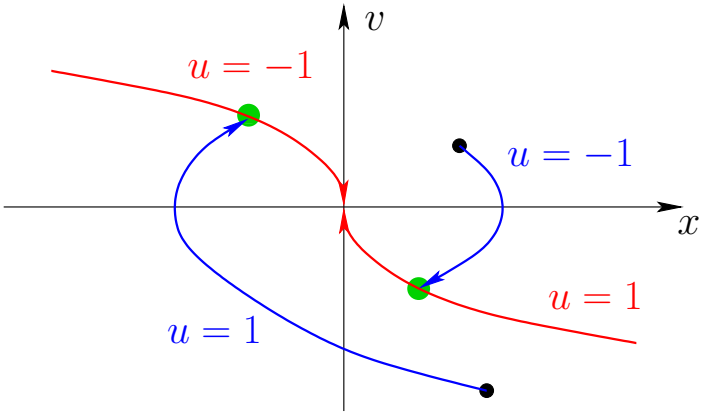

FIGURE 3.3 – Temps-optimalité pour l'arrêt d'un tram : trajectoires et courbe de commutation.

forment la courbe de commutation, et le point vert indique la commutation sur chaque trajectoire. On notera que le contrôle optimal s'écrit comme un **feedback** en fonction de l'état puisque l'on a

— $u = 1$ si X est au-dessous de la courbe de commutation (dans ce cas, il faut accélérer)

— $u = -1$ si X est au-dessus de cette courbe (dans ce cas, il faut décélérer).

L'obtention d'un contrôle optimal sous forme de feedback est très intéressant en pratique. □

Chapitre 4

Le système linéaire-quadratique (LQ)

Ce chapitre est consacré à l'étude du système linéaire-quadratique (LQ). Il s'agit d'un problème de contrôle optimal régi par une dynamique linéaire et où le critère à minimiser est quadratique en le contrôle et en la trajectoire associée. Ce problème étant relativment simple, il nous sera possible d'en mener une analyse mathématique complète. D'une part, nous montrerons l'existence et l'unicité du contrôle optimal. D'autre part, cette analyse nous permettra de dégager plusieurs notions importantes pour la suite : l'**état adjoint** pour le calcul de la différentielle du critère, le **Hamiltonien** pour la formulation du contrôle optimal à tout temps comme un minimiseur fonction des valeurs instantanées de l'état adjoint et enfin, celle de **feedback** grâce à l'équation de Riccati afin de formuler le contrôle optimal en **boucle fermée**, c'est-à-dire comme une fonction instantanée de l'état du système.

4.1 Présentation du système LQ

On se donne un intervalle de temps $[0,T]$, avec $T > 0$, une matrice $A \in \mathbb{R}^{d\times d}$ et une matrice $B \in \mathbb{R}^{d\times k}$. On se donne également une condition initiale $x_0 \in \mathbb{R}^d$ et (pour un peu plus de généralité) un terme de dérive $f \in L^1([0,T];\mathbb{R}^d)$. Le système de contrôle linéaire s'écrit sous la forme

$$\dot{x}_u(t) = Ax_u(t) + Bu(t) + f(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0. \tag{4.1}$$

L'ensemble des contrôles admissibles est ici l'espace

$$V = L^2([0,T];\mathbb{R}^k). \tag{4.2}$$

Pour chaque contrôle $u \in L^2([0,T];\mathbb{R}^k)$, il existe une unique trajectoire $x_u \in AC([0,T];\mathbb{R}^d)$ associée à ce contrôle.

L'objectif de ce chapitre est de chercher un **contrôle optimal** (en fait **le** contrôle optimal, car nous verrons qu'il est unique) qui minimise dans $L^2([0,T];\mathbb{R}^k)$ le critère

$$J(u) = \frac{1}{2}\int_0^T u(t)^\dagger Ru(t)\,\mathrm{d}t + \frac{1}{2}\int_0^T e_{x_u}(t)^\dagger Qe_{x_u}(t)\,\mathrm{d}t + \frac{1}{2}e_{x_u}(T)^\dagger De_{x_u}(T), \tag{4.3}$$

où $e_{x_u} = x_u - \xi$ et où $\xi \in C^0([0,T];\mathbb{R}^d)$ est une **trajectoire cible** donnée. On s'intéresse donc au problème suivant :

$$\text{Chercher } \overline{u} \in V \text{ tel que } J(\overline{u}) = \inf_{u \in V} J(u). \tag{4.4}$$

Dans la définition du critère J, les matrices $Q, D \in \mathbb{R}^{d\times d}$ sont symétriques **semi-définies positives**, tandis que la matrice $R \in \mathbb{R}^{k\times k}$ est symétrique **définie positive**. La définie positivité de la matrice R jouera un rôle clé pour assurer l'existence et l'unicité du contrôle optimal minimisant J sur $L^2([0,T];\mathbb{R}^k)$. On notera que le critère J résulte d'une pondération au sens des moindres carrés entre l'atteinte de la trajectoire cible décrite par la fonction ξ et le fait que le contrôle ne soit pas "trop grand" dans $L^2([0,T];\mathbb{R}^k)$. En revanche, on ne s'impose pas ici d'atteindre exactement la cible au temps final T (ni à aucun temps intermédiaire). Une illustration générale du problème de contrôle optimal LQ est présentée à la figure 4.1.

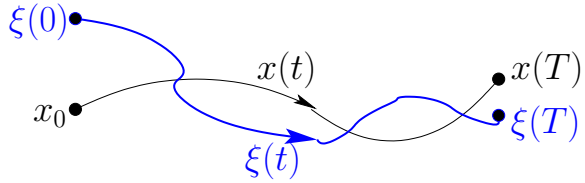

FIGURE 4.1 – Illustration du problème de contrôle optimal LQ : trajectoire cible et trajectoire optimale.

Remarque 4.1. [$Q = D = 0$] On peut prendre $Q = D = 0$ dans le critère (4.3). La solution du problème (4.4) est alors triviale : $u \equiv 0$ sur $[0,T]$. □

Afin d'étudier les propriétés de la fonctionnelle J, il sera utile de poser

$$J(u) = J_R(u) + J_{QD}(u), \quad \forall u \in V, \tag{4.5}$$

avec

$$J_R(u) = \frac{1}{2}\int_0^T u(t)^\dagger R u(t)\,\mathrm{d}t, \tag{4.6a}$$

$$J_{QD}(u) = \frac{1}{2}\int_0^T e_{x_u}(t)^\dagger Q e_{x_u}(t)\,\mathrm{d}t + \frac{1}{2} e_{x_u}(T)^\dagger D e_{x_u}(T). \tag{4.6b}$$

Lemme 4.2 (Forte convexité et continuité de J)**.** *La fonctionnelle J définie en* (4.3) *est fortement convexe et continue sur l'espace de Hilbert $V = L^2([0,T];\mathbb{R}^k)$.*

Démonstration. Comme la matrice R est symétrique définie positive, la fonctionnelle J_R est fortement convexe sur V de paramètre $\alpha = \lambda_{\min}(R)$ (la plus petite valeur propre de la matrice R). En effet, pour deux vecteurs $v_1, v_2 \in \mathbb{R}^k$, on a

$$\begin{aligned}\left(\frac{v_1+v_2}{2}\right)^\dagger R\left(\frac{v_1+v_2}{2}\right) &= \frac{v_1^\dagger R v_1 + v_2^\dagger R v_2}{2} - \frac{1}{4}(v_1-v_2)^\dagger R (v_1-v_2)\\ &\le \frac{v_1^\dagger R v_1 + v_2^\dagger R v_2}{2} - \frac{1}{4}\lambda_{\min}(R)|v_1-v_2|^2_{\mathbb{R}^k}.\end{aligned}$$

On en déduit que pour deux contrôles $u_1, u_2 \in V$, on a

$$J_R\left(\frac{u_1+u_2}{2}\right) \leq \frac{J_R(u_1)+J_R(u_2)}{2} - \frac{1}{8}\lambda_{\min}(R)\|u_1-u_2\|_V^2,$$

ce qui prouve la forte convexité de la fonctionnelle J_R sur V avec paramètre $\alpha = \lambda_{\min}(R)$. De plus, la fonctionnelle J_R est clairement continue en u. Par ailleurs, la fonctionnelle J_{QD} est convexe sur V comme composée d'une application convexe par une application affine. En effet,

- comme $x_u(t) = e^{tA}x_0 + \int_0^t e^{(t-s)A}(Bu(s)+f(s))\,\mathrm{d}s$, l'application qui à $u \in L^2([0,T];\mathbb{R}^k)$ associe $e_{x_u} = x_u - \xi \in C^0([0,T];\mathbb{R}^d)$ est affine ;
- comme les matrices Q et D sont symétriques semi-définies positives, on montre facilement que l'application qui à $y \in C^0([0,T];\mathbb{R}^d)$ associe $\frac{1}{2}\int_0^T y(t)^\dagger Q y(t)\,\mathrm{d}t + \frac{1}{2}y(T)^\dagger D y(T) \in \mathbb{R}$ est convexe (même raisonnement que ci-dessus).

La fonctionnelle J_{QD} est en outre continue comme composée de deux applications continues. En conclusion, la fonctionnelle J est fortement convexe sur V comme somme d'une application fortement convexe (J_R) et d'une application convexe (J_{QD}), et J est également continue comme somme de deux applications continues. □

Corollaire 4.3 (Existence et unicité)**.** *Il existe un unique contrôle optimal $\overline{u} \in V$ solution de* (4.4)*.*

Démonstration. Il suffit de combiner le théorème 3.8 (avec $K = V$) avec le lemme 4.2. □

4.2 Différentielle du critère : état adjoint

L'objectif de cette section est d'utiliser les résultats de la section 3.2.3 afin d'établir une condition nécessaire et suffisante d'optimalité formulée à l'aide de la différentielle de la fonctionnelle J.

Lemme 4.4 (Différentiabilité de J)**.** *La fonctionnelle J est différentiable sur V et on a, pour tout $u \in V$,*

$$\nabla J(u) = Ru + B^\dagger p\ (\in V), \tag{4.7}$$

*où l'**état adjoint** $p \in C^1([0,T];\mathbb{R}^d)$ est l'unique solution de l'équation différentielle rétrograde en temps*

$$\dot{p}(t) = -A^\dagger p(t) - Q e_{x_u}(t), \quad \forall t \in [0,T], \qquad p(T) = D e_{x_u}(T). \tag{4.8}$$

Démonstration. Comme $J = J_R + J_{QD}$, nous allons considérer séparément la différentiabilité des fonctionnelles J_R et J_{QD}.
(1) La différentiabilité de J_R est immédiate puisque, en utilisant la symétrie de la matrice R,

il vient, pour toute perturbation du contrôle $\delta u \in V$,

$$\begin{aligned} J_R(u+\delta u) &= \frac{1}{2}\int_0^T (u(t)+\delta u(t))^\dagger R(u(t)+\delta u(t))\,\mathrm{d}t \\ &= J_R(u) + \int_0^T \delta u(t)^\dagger R u(t)\,\mathrm{d}t + J_R(\delta u) \\ &= J_R(u) + (Ru, \delta u)_V + J_R(\delta u). \end{aligned}$$

Comme $\frac{J_R(\delta u)}{\|\delta u\|_V} \leq \frac{1}{2}\lambda_{\max}(R)\|\delta u\|_V$, on conclut que $\nabla J_R(u) = Ru \in V$, ce qui signifie que p.p. sur $[0,T]$, on a $(\nabla J_R(u))(t) = Ru(t)$.
(2) Pour différentier J_{QD}, on considère la trajectoire perturbée $x_{u+\delta u}$, associée au contrôle perturbé $u+\delta u$. Par linéarité, on a $x_{u+\delta u} = x_u + \delta x$ avec

$$\frac{d}{dt}\delta x(t) = A\delta x(t) + B\delta u(t), \quad \forall t \in [0,T], \qquad \delta x(0) = 0.$$

La perturbation de la trajectoire δx est donc linéaire en δu et on a $\|\delta x\|_{C^0([0,T];\mathbb{R}^d)} \leq C\|\delta u\|_V$ car $\delta x(t) = \int_0^t e^{(t-s)A}B\delta u(s)\,\mathrm{d}s$, où C est une constante dépendant de A, B et T mais qui est uniforme en δu. Comme les matrices Q et D sont symétriques, et en raisonnant comme ci-dessus, on obtient

$$\begin{aligned} J_{QD}(u+\delta u) = J_{QD}(u) &+ \int_0^T \delta x(t)^\dagger Q e_{x_u}(t)\,\mathrm{d}t + \delta x(T)^\dagger D e_{x_u}(T) \\ &+ \frac{1}{2}\int_0^T \delta x(t)^\dagger Q\delta x(t)\,\mathrm{d}t + \frac{1}{2}\delta x(T)^\dagger D\delta x(T), \end{aligned}$$

ce qui montre que

$$(\nabla J_{QD}(u), \delta u)_V = \int_0^T \delta x(t)^\dagger Q e_{x_u}(t)\,\mathrm{d}t + \delta x(T)^\dagger D e_{x_u}(T).$$

Au membre de droite, la perturbation du contrôle δu n'apparaît pas explicitement, mais uniquement de manière implicite par le fait que la perturbation de la trajectoire δx dépend (linéairement) de la perturbation du contrôle δu. Afin de faire apparaître explicitement δu au membre de droite, on utilise l'état adjoint $p \in C^1([0,T];\mathbb{R}^d)$ solution de (4.8). En effet, en intégrant par parties en temps, on constate que

$$\begin{aligned} (\nabla J_{QD}(u), \delta u)_V &= \int_0^T \delta x(t)^\dagger Q e_x(t)\,\mathrm{d}t + \delta x(T)^\dagger D e_x(T) \\ &= -\int_0^T \delta x(t)^\dagger(\dot{p}(t) + A^\dagger p(t))\,\mathrm{d}t + \delta x(T)^\dagger p(T) \\ &= \int_0^T \big(\dot{\delta x}(t)^\dagger p(t) - \delta x(t)^\dagger A^\dagger p(t)\big)\,\mathrm{d}t \\ &= \int_0^T (B\delta u(t))^\dagger p(t)\,\mathrm{d}t = \int_0^T \delta u(t)^\dagger B^\dagger p(t)\,\mathrm{d}t = (B^\dagger p, \delta u)_V. \end{aligned}$$

En conclusion, on a montré que $\nabla J_{QD}(u) = B^\dagger p$, ce qui conclut la preuve. □

Théorème 4.5 (CNS d'optimalité)**.** *Le contrôle $\overline{u} \in V$ est optimal pour le problème LQ* ***si et seulement si*** *on a*

$$\overline{u}(t) = -R^{-1}B^\dagger \overline{p}(t) \quad \forall t \in [0,T], \tag{4.9}$$

où l'état adjoint $\overline{p} : [0,T] \to \mathbb{R}^d$ est tel que

$$\frac{d\overline{p}}{dt}(t) = -A^\dagger \overline{p}(t) - Qe_{\overline{x}}(t), \quad \forall t \in [0,T], \qquad \overline{p}(T) = De_{\overline{x}}(T), \tag{4.10}$$

où $e_{\overline{x}} = \overline{x} - \xi$ et où $\overline{x} = x_{\overline{u}}$ est la trajectoire associée au contrôle optimal $\overline{u}$, i.e.,

$$\frac{d\overline{x}}{dt}(t) = A\overline{x}(t) + B\overline{u}(t) + f(t), \quad \forall t \in [0,T], \qquad \overline{x}(0) = x_0. \tag{4.11}$$

Le triplet $(\overline{x}, \overline{p}, \overline{u})$ satisfaisant les conditions ci-dessus est appelé une ***extrémale****.*

Démonstration. Il suffit de combiner la proposition 3.10 avec le lemme 4.4, le caractère suffisant de la condition (4.9) résultant de la convexité de la fonctionnelle J. □

Remarque 4.6. [État adjoint] Attention, il n'y a pas de condition initiale sur $\overline{p}$, mais une condition finale en T. Par ailleurs, dans la littérature, la convention est parfois de définir l'état adjoint comme un vecteur ligne $\widehat{p} := \overline{p}^\dagger$. Dans ce cas, le système différentiel rétrograde s'écrit $\frac{d}{dt}\widehat{p}(t) = -\widehat{p}(t)A - e_{\overline{x}}(t)^\dagger Q$, pour tout $t \in [0,T]$, et $\widehat{p}(T) = e_{\overline{x}}(T)^\dagger D$. Enfin, le contrôle optimal est $u(t) = -R^{-1}B^\dagger \widehat{p}(t)^\dagger$. □

Remarque 4.7. [Régularité] On notera que si $(\overline{x}, \overline{p}, \overline{u})$ est une extrémale, on a $\overline{p} \in C^1([0,T]; \mathbb{R}^d)$ et par conséquent $\overline{u} \in C^1([0,T]; \mathbb{R}^k)$. Il n'y a pas ici de phénomène de commutation pour le contrôle optimal. □

Remarque 4.8. [Unicité de l'extrémale] Même si on sait déjà qu'on a unicité du contrôle optimal $\overline{u}$, donc de la trajectoire optimale $\overline{x}$ et de la trajectoire adjointe $\overline{p}$, il est instructif de montrer directement l'unicité de l'extrémale. Par linéarité (considérer la différence entre deux extrémales), il suffit de montrer que dans le cas sans dérive et avec cible nulle, une extrémale est nécessairement nulle. Considérons donc une extrémale $(\overline{x}, \overline{p}, \overline{u})$ telle que

$$\begin{aligned}
&\frac{d\overline{x}}{dt}(t) = A\overline{x}(t) + B\overline{u}(t), \qquad \overline{x}(0) = 0,\\
&\frac{d\overline{p}}{dt}(t) = -A^\dagger \overline{p}(t) - Q\overline{x}(t), \quad \overline{p}(T) = D\overline{x}(T),\\
&\overline{u}(t) = -R^{-1}B^\dagger \overline{p}(t).
\end{aligned}$$

L'observation cruciale est que

$$\begin{aligned}
\frac{d}{dt}\big(\overline{p}(t)^\dagger \overline{x}(t)\big) &= \left(\frac{d\overline{p}}{dt}(t)\right)^\dagger \overline{x}(t) + \overline{p}(t)^\dagger \frac{d\overline{x}}{dt}(t)\\
&= -\overline{x}(t)^\dagger Q\overline{x}(t) - (B^\dagger \overline{p}(t))^\dagger R^{-1}B^\dagger \overline{p}(t) \le 0.
\end{aligned}$$

Comme $\overline{x}(0) = 0$, en intégrant de 0 à T, il vient

$$\begin{aligned} 0 &= \overline{p}(T)^\dagger \overline{x}(T) - \int_0^T \frac{d}{dt}\big(\overline{p}(t)^\dagger \overline{x}(t)\big)\,\mathrm{d}t \\ &= \overline{x}(T)^\dagger D\overline{x}(T) + \int_0^T \Big\{ \overline{x}(t)^\dagger Q \overline{x}(t) + (B^\dagger \overline{p}(t))^\dagger R^{-1} B^\dagger \overline{p}(t) \Big\}\,\mathrm{d}t. \end{aligned}$$

Comme les matrices D et Q sont positives et que la matrice R est définie positive, on en déduit que $B^\dagger \overline{p}(t) = 0$ sur $[0,T]$. Donc, $\overline{u}(t) = 0$, ce qui implique que $\overline{x}(t) = 0$, et ce qui implique enfin que $\overline{p}(t) = 0$. □

Exemple 4.9. [Mouvement d'un point matériel] On considère un point matériel qui peut se déplacer sur une droite et dont on contrôle la vitesse (cf. l'exemple 1.22). Le système de contrôle linéaire s'écrit, avec $d = k = 1$,

$$\dot{x}_u(t) = u(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0.$$

Le critère à minimiser dans $V = L^2([0,T];\mathbb{R})$ est

$$J(u) = \frac{1}{2}\int_0^T x_u(t)^2\,\mathrm{d}t + \frac{1}{2}\int_0^T u(t)^2\,\mathrm{d}t,$$

qui réalise une pondération au sens des moindres carrés entre l'atteinte de la cible nulle sur $[0,T]$ et le fait que le contrôle ne soit pas trop grand dans $L^2([0,T];\mathbb{R})$. Ce problème rentre dans le cadre du système LQ introduit à la section 4.1 en posant

$$A = 0, \quad B = 1, \quad R = 1, \quad Q = 1, \quad D = 0, \quad \xi \equiv 0.$$

En appliquant le théorème 4.5, on déduit que le contrôle optimal est

$$\overline{u}(t) = -\overline{p}(t),$$

où l'état adjoint est solution de

$$\frac{d\overline{p}}{dt}(t) = -\overline{x}(t), \quad \forall t \in [0,T], \qquad \overline{p}(T) = 0.$$

On a donc

$$\frac{d}{dt}\begin{pmatrix}\overline{x}(t)\\ \overline{p}(t)\end{pmatrix} = \underbrace{\begin{pmatrix}0 & -1\\ -1 & 0\end{pmatrix}}_{=Z}\begin{pmatrix}\overline{x}(t)\\ \overline{p}(t)\end{pmatrix}, \qquad e^{tZ} = \begin{pmatrix}\cosh(t) & -\sinh(t)\\ -\sinh(t) & \cosh(t)\end{pmatrix},$$

si bien que

$$\begin{aligned} \overline{x}(t) &= x_0\cosh(t) - \overline{p}(0)\sinh(t), \\ \overline{p}(t) &= -x_0\sinh(t) + \overline{p}(0)\cosh(t). \end{aligned}$$

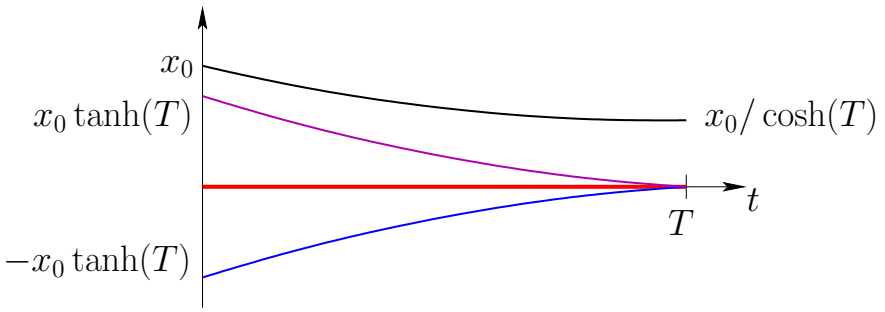

FIGURE 4.2 – Illustration de l'extrémale obtenue à l'exemple 4.9 (mouvement d'un point matériel) : trajectoire $\overline{x}(t)$, état adjoint $\overline{p}(t)$, contrôle optimal $\overline{u}(t)$; la cible $\xi(t)$ est identiquement nulle.

On notera que l'état adjoint initial est, à ce stade, encore inconnu. Afin de le déterminer, on utilise la condition en $t = T$ sur l'état adjoint, à savoir $\overline{p}(T) = 0$. On obtient facilement que $\overline{p}(0) = x_0 \tanh(T)$. En conclusion, l'extrémale s'écrit

$$\begin{aligned}
\overline{x}(t) &= x_0 \tfrac{1}{\cosh(T)} \cosh(T-t), \\
\overline{p}(t) &= x_0 \tfrac{1}{\cosh(T)} \sinh(T-t), \\
\overline{u}(t) &= -\overline{p}(t) = -x_0 \tfrac{1}{\cosh(T)} \sinh(T-t).
\end{aligned}$$

Cette extrémale est illustrée à la figure 4.2. □

4.3 Principe du minimum : Hamiltonien

L'objectif de cette section est de reformuler le théorème 4.5 à l'aide de la notion de Hamiltonien. Ce point de vue nous sera très utile au chapitre suivant lorsque nous aborderons les systèmes de contrôle non-linéaires et formulerons le principe du minimum de Pontryaguine.

Définition 4.10 (Hamiltonien)**.** *Le* ***Hamiltonien*** *associé au système de contrôle linéaire* (4.1) *et à la fonctionnelle J définie en* (4.3) *est l'application $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ telle que*

$$H(t,x,p,u) = p^\dagger(Ax + Bu + f(t)) + \frac{1}{2}u^\dagger Ru + \frac{1}{2}(x - \xi(t))^\dagger Q(x - \xi(t)). \tag{4.12}$$

On notera bien que dans cette écriture, (x,p,u) désigne un vecteur générique de $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^k$.

Un calcul élémentaire sur les dérivées partielles du Hamiltonien (qui sont ici identifiées à des vecteurs colonne) montre que

$$\nabla_x H(t,x,p,u) = A^\dagger p + Q(x - \xi(t)), \tag{4.13a}$$
$$\nabla_p H(t,x,p,u) = Ax + Bu + f(t), \tag{4.13b}$$
$$\nabla_u H(t,x,p,u) = B^\dagger p + Ru. \tag{4.13c}$$

On considère maintenant l'extrémale $(\overline{x}, \overline{p}, \overline{u})$ obtenue au théorème 4.5. Pour tout $t \in [0,T]$, on évalue H et ses dérivées partielles en $(t, \overline{x}(t), \overline{p}(t), \overline{u}(t))$. On constate d'une part que

$$\frac{d\overline{x}}{dt}(t) = A\overline{x}(t) + B\overline{u}(t) + f(t) = \nabla_p H(t, \overline{x}(t), \overline{p}(t), \overline{u}(t)), \tag{4.14a}$$

$$\frac{d\overline{p}}{dt}(t) = -A^\dagger \overline{p}(t) - Q(\overline{x}(t) - \xi(t)) = -\nabla_x H(t, \overline{x}(t), \overline{p}(t), \overline{u}(t)), \tag{4.14b}$$

et d'autre part que

$$\nabla_u H(t, \overline{x}(t), \overline{p}(t), \overline{u}(t)) = 0. \tag{4.15}$$

Comme la fonction $v \mapsto H(t,x,p,v)$ est fortement convexe en $v \in \mathbb{R}^k$ pour tout triplet (t,x,p) fixé dans $[0,T] \times \mathbb{R}^d \times \mathbb{R}^d$, l'équation (4.15) ne signifie rien d'autre que

$$\overline{u}(t) = \arg\min_{v \in \mathbb{R}^k} H(t, \overline{x}(t), \overline{p}(t), v), \quad \forall t \in [0,T]. \tag{4.16}$$

Il s'agit du **principe du minimum de Pontryaguine (PMP)** dans le cas particulier du système LQ. Résumons ce résultat sous la forme d'une proposition.

Proposition 4.11 (PMP pour le système LQ)**.** *Le contrôle $\overline{u} \in V$ est optimal pour le problème LQ* ***si et seulement si*** *on a*

$$\overline{u}(t) = \arg\min_{v \in \mathbb{R}^k} H(t, \overline{x}(t), \overline{p}(t), v), \quad \forall t \in [0,T], \tag{4.17}$$

avec

$$\frac{d\overline{x}}{dt}(t) = \nabla_p H(t, \overline{x}(t), \overline{p}(t), \overline{u}(t)) = A\overline{x}(t) + B\overline{u}(t), \qquad \overline{x}(0) = x_0, \tag{4.18a}$$

$$\frac{d\overline{p}}{dt}(t) = -\nabla_x H(t, \overline{x}(t), \overline{p}(t), \overline{u}(t)) = -A^\dagger \overline{p}(t) - Qe_{\overline{x}}(t), \quad \overline{p}(T) = De_{\overline{x}}(T), \tag{4.18b}$$

où $e_{\overline{x}}(t) = \overline{x}(t) - \xi(t)$.

Remarque 4.12. [Convention de signe] On aurait pu définir $\widehat{H} := -H$ et aboutir à un principe du maximum pour $\widehat{H}$. □

Dans le cas particulier avec dérive et cible nulles, i.e., lorsque $f \equiv 0$ et $\xi \equiv 0$ sur $[0,T]$, le Hamiltonien H ne dépend pas du temps, i.e., on a

$$\frac{\partial H}{\partial t}(t,x,p,u) = 0. \tag{4.19}$$

On dit que le Hamiltonien est **autonome**.

Proposition 4.13 (Conservation du Hamiltonien le long de l'extrémale)**.** *On suppose que dérive et cible sont nulles, i.e., que le Hamiltonien est autonome. Alors, la valeur du Hamiltonien se conserve le long de l'extrémale $(\overline{x}, \overline{p}, \overline{u})$.*

Démonstration. On considère l'application $\mathcal{H} : [0,T] \to \mathbb{R}$ telle que

$$\mathcal{H}(t) = H(\overline{x}(t), \overline{p}(t), \overline{u}(t)), \quad \forall t \in [0,T].$$

En dérivant cette fonction par rapport au temps, il vient

$$\begin{aligned}\frac{d\mathcal{H}}{dt}(t) &= \big(\nabla_x H\big)^\dagger \frac{d\overline{x}}{dt}(t) + \big(\nabla_p H\big)^\dagger \frac{d\overline{p}}{dt}(t) + \big(\nabla_u H\big)^\dagger \frac{d\overline{u}}{dt}(t) \\ &= -\frac{d\overline{p}}{dt}(t)^\dagger \frac{d\overline{x}}{dt}(t) + \frac{d\overline{x}}{dt}(t)^\dagger \frac{d\overline{p}}{dt}(t) + 0 = 0,\end{aligned}$$

ce qui conclut la preuve. □

Exemple 4.14. [Mouvement d'un point matériel] On reprend l'exemple 4.9 du mouvement d'un point matériel le long d'une droite et dont on contrôle la vitesse, i.e., $\dot{x}_u(t) = u(t)$, pour tout $t \in [0,T]$, et $x_u(0) = x_0$. Le critère à minimiser est à nouveau $J(u) = \frac{1}{2}\int_0^T x_u(t)^2\,\mathrm{d}t + \frac{1}{2}\int_0^T u(t)^2\,\mathrm{d}t$. Ce problème rentre dans le cadre du système LQ avec $d = k = 1$ et

$$A = 0, \quad B = 1, \quad R = 1, \quad Q = 1, \quad D = 0, \quad \xi \equiv 0.$$

Le Hamiltonien est l'application de $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ dans $\mathbb{R}$ telle que

$$H(x,p,u) = pu + \frac{1}{2}u^2 + \frac{1}{2}x^2.$$

À (x,p) étant fixés, l'application $u \mapsto H(x,p,u)$ est quadratique. Le principe du minimum de Pontryaguine (cf. la proposition 4.11) implique que le contrôle optimal $\overline{u}(t)$ est, pour tout $t \in [0,T]$, le minimiseur de $u \mapsto H(\overline{x}(t), \overline{p}(t), u)$ sur $\mathbb{R}$. En utilisant l'expression de H, on obtient facilement

$$\overline{u}(t) = -\overline{p}(t).$$

On retrouve ainsi le même résultat que celui obtenu en considérant la différentielle de J. De plus, si on évalue le Hamiltonien le long de l'extrémale, il vient

$$\mathcal{H}(t) = H(\overline{x}(t), \overline{p}(t), \overline{u}(t)) = \frac{1}{2}(\overline{x}(t)^2 - \overline{p}(t)^2) = \frac{1}{2}\left(\frac{x_0}{\cosh(T)}\right)^2,$$

car

$$\overline{x}(t) = \frac{x_0}{\cosh(T)}\cosh(T-t), \qquad \overline{p}(t) = \frac{x_0}{\cosh(T)}\sinh(T-t).$$

Ce calcul confirme que le Hamiltonien est bien constant le long de l'extrémale, comme annoncé à la proposition 4.13. □

4.4 Équation de Riccati : feedback

L'objectif de cette section est de montrer qu'il est possible, en résolvant l'équation de Riccati, de formuler à tout temps $t \in [0,T]$ le contrôle optimal $\overline{u}(t)$ comme un feedback sur l'état $\overline{x}(t)$. Pour simplifier, on suppose que dérive et cible sont nulles.

Théorème 4.15 (Équation de Riccati)**.** *On suppose que dérive et cible sont nulles. Il existe une unique matrice $P \in C^1([0,T];\mathbb{R}^{d\times d})$ solution de l'équation de Riccati*

$$\dot{P}(t) = -A^\dagger P(t) - P(t)A + P(t)BR^{-1}B^\dagger P(t) - Q,\ \forall t \in [0,T], \quad P(T) = D, \tag{4.20}$$

et on a

$$\overline{p}(t) = P(t)\overline{x}(t), \quad \forall t \in [0,T], \tag{4.21}$$

si bien que le contrôle optimal s'écrit sous forme de ***boucle fermée*** *:*

$$\overline{u}(t) = K(t)\overline{x}(t), \quad K(t) = -R^{-1}B^\dagger P(t), \quad \forall t \in [0,T]. \tag{4.22}$$

De plus, la matrice $P(t)$ est symétrique semi-définie positive, et définie positive si la matrice D est définie positive. Enfin, la valeur optimale du critère est $J(\overline{u}) = \frac{1}{2}x_0^\dagger P(0)x_0$.

Démonstration. (1) Dépendance linéaire. Le problème LQ étant bien posé, on sait qu'il existe un unique couple $(\overline{x},\overline{p}) \in C^1([0,T];\mathbb{R}^d \times \mathbb{R}^d)$ tel que

$$\begin{aligned}
\frac{d\overline{x}}{dt}(t) &= A\overline{x}(t) - BR^{-1}B\overline{p}(t), \quad \overline{x}(0) = x_0,\\
\frac{d\overline{p}}{dt}(t) &= -A^\dagger\overline{p}(t) - Q\overline{x}(t), \qquad \overline{p}(T) = D\overline{x}(T).
\end{aligned}$$

Par linéarité, le couple $(\overline{x},\overline{p})$ dépend linéairement de la condition initiale $x_0 \in \mathbb{R}^d$. Il existe donc des matrices $\mathcal{X},\mathcal{P}$ dans $C^1([0,T];\mathbb{R}^{d\times d})$ telles que

$$\overline{x}(t) = \mathcal{X}(t)x_0, \quad \overline{p}(t) = \mathcal{P}(t)x_0, \quad \forall t \in [0,T],$$

et on a $\mathcal{X}(0) = I_d$.
(2) Inversibilité de $\mathcal{X}(t)$. Nous allons montrer que la matrice $\mathcal{X}(t)$ est inversible pour tout $t \in [0,T]$. Pour ce faire, on raisonne par l'absurde. Soit $s \in [0,T]$ et $0 \neq x_0 \in \mathbb{R}^d$ tels que $\overline{x}(s) = \mathcal{X}(s)x_0 = 0$. On a nécessairement $s > 0$ car $\mathcal{X}(0) = I_d$. De plus, on a vu que

$$\frac{d}{dt}\big(\overline{p}(t)^\dagger\overline{x}(t)\big) = -\overline{x}(t)^\dagger Q\overline{x}(t) - (B^\dagger\overline{p}(t))^\dagger R^{-1}B^\dagger\overline{p}(t).$$

En intégrant de s à T, et comme $\overline{x}(s) = 0$, il vient

$$0 = (D\overline{x}(T))^\dagger\overline{x}(T) + \int_s^T \Big(\overline{x}(t)^\dagger Q\overline{x}(t) + (B^\dagger\overline{p}(t))^\dagger R^{-1}B^\dagger\overline{p}(t)\Big)\,\mathrm{d}t \geq 0.$$

Les matrices D,Q,R étant symétriques (semi-)définies positives, on en déduit que

$$\overline{u}(t) = -R^{-1}B^\dagger\overline{p}(t) = 0, \quad \forall t \in [s,T].$$

On a donc $\frac{d\overline{x}}{dt}(t) = A\overline{x}(t)$ et $\overline{x}(s) = 0$; d'où $\overline{x}(t) = 0$ sur $[s,T]$. De même, comme on a $\frac{d\overline{p}}{dt}(t) = -A^\dagger\overline{p}(t)$ et $\overline{p}(T) = D\overline{x}(T) = 0$, il vient $\overline{p}(t) = 0$ sur $[s,T]$. On en déduit que $(\overline{x},\overline{p})$ vérifie un système différentiel linéaire avec conditions finales $\overline{x}(T) = \overline{p}(T) = 0$. Ceci implique

que $\overline{x}(t) = \overline{p}(t) = 0$ sur $[0,T]$; en particulier, on obtient $x_0 = 0$, d'où la contradiction.
(3) Équation de Riccati. On pose

$$P(t) = \mathcal{P}(t)\mathcal{X}(t)^{-1}, \quad \forall t \in [0,T].$$

Par construction, on a $P \in C^1([0,T];\mathbb{R}^{d\times d})$. De plus, on constate que

$$\begin{aligned}\frac{d\overline{p}}{dt}(t) &= \frac{dP}{dt}(t)\overline{x}(t) + P(t)\frac{d\overline{x}}{dt}(t)\\ &= \left(\frac{dP}{dt}(t) + P(t)A - P(t)BR^{-1}B^\dagger P(t)\right)\overline{x}(t),\end{aligned}$$

et par ailleurs, on a également $\frac{d\overline{p}}{dt}(t) = -A^\dagger\overline{p}(t) - Q\overline{x}(t)$. On en déduit que

$$\left(\frac{dP}{dt}(t) + P(t)A + A^\dagger P(t) - P(t)BR^{-1}B^\dagger P(t) + Q\right)\overline{x}(t) = 0,$$

pour tout $t \in [0,T]$ et pour tout $x_0 \in \mathbb{R}^d$. Pour tout $t \in [0,T]$ fixé, le vecteur $\overline{x}(t)$ décrit $\mathbb{R}^d$ lorsque x_0 décrit $\mathbb{R}^d$ (car $\mathcal{X}(t)$ est inversible). Par conséquent, la fonction $t \mapsto P(t)$ est bien solution de l'équation de Riccati pour tout $t \in [0,T]$. En raisonnant de manière analogue, on constate que $\overline{p}(T) = D\overline{x}(T) = P(T)\overline{x}(T)$. Comme $\overline{x}(T)$ décrit $\mathbb{R}^d$ lorsque x_0 décrit $\mathbb{R}^d$, on conclut que $P(T) = D$.
(4) Propriétés de $P(t)$. La fonction $t \mapsto P(t)$ est solution d'un système différentiel quadratique. La non-linéarité satisfait donc une condition de Lipschitz locale, ce qui assure l'unicité de la solution. L'unicité prouve que $P(t)$ est symétrique pour tout $t \in [0,T]$ car la fonction $t \mapsto P(t)^\dagger$ satisfait la même équation. Afin d'établir la positivité de $P(t)$ pour tout $t \in [0,T]$, on raisonne comme suit. Soit $x \in \mathbb{R}^d$. Posons $x_0 = \mathcal{X}(t)^{-1}x$ de sorte que $x = \overline{x}(t)$ où $\overline{x}$ est la trajectoire optimale issue de x_0. Comme la fonction $t \mapsto \overline{p}(t)^\dagger\overline{x}(t)$ est décroissante, il vient

$$x^\dagger P(t)x = \overline{x}(t)^\dagger P(t)\overline{x}(t) \geq \overline{x}(T)^\dagger D\overline{x}(T) \geq 0,$$

ce qui montre que $P(t)$ est semi-définie positive. Enfin, si la matrice D est définie positive, cela entraîne $\overline{x}(T) = 0$, d'où $x = \mathcal{X}(t)\mathcal{X}(T)^{-1}\overline{x}(T) = 0$, i.e., la matrice $P(t)$ est alors définie positive.
(5) Valeur optimale du critère. Il vient

$$\begin{aligned}J(\overline{u}) &= \frac{1}{2}\int_0^T \left(\overline{x}(t)^\dagger Q\overline{x}(t) + \overline{u}(t)^\dagger R\overline{u}(t)\right)\mathrm{d}t + \frac{1}{2}\overline{x}(T)^\dagger D\overline{x}(T)\\ &= \frac{1}{2}\int_0^T \left(\overline{x}(t)^\dagger Q\overline{x}(t) - \overline{p}(t)^\dagger B\overline{u}(t)\right)\mathrm{d}t + \frac{1}{2}\overline{x}(T)^\dagger D\overline{x}(T)\\ &= \frac{1}{2}\int_0^T \left(\overline{x}(t)^\dagger Q\overline{x}(t) - \overline{p}(t)^\dagger B\overline{u}(t)\right)\mathrm{d}t + \frac{1}{2}\overline{p}(T)^\dagger\overline{x}(T)\\ &= \frac{1}{2}\int_0^T -\frac{d}{dt}\big(\overline{p}(t)^\dagger\overline{x}(t)\big)\,\mathrm{d}t + \frac{1}{2}\overline{p}(T)^\dagger\overline{x}(T)\\ &= \frac{1}{2}\overline{p}(0)^\dagger\overline{x}(0) = \frac{1}{2}\overline{x}(0)^\dagger P(0)\overline{x}(0) = \frac{1}{2}x_0^\dagger P(0)x_0,\end{aligned}$$

ce qui conclut la preuve. □

Remarque 4.16. [Représentation linéaire de l'équation de Riccati] Au lieu de résoudre un système différentiel quadratique de taille $\frac{d(d+1)}{2}$ (P est symétrique), on peut considérer le système différentiel **linéaire** suivant qui est de taille $2d$:

$$\frac{d}{dt}\begin{pmatrix} x(t) \\ p(t) \end{pmatrix} = \underbrace{\begin{pmatrix} A & -BR^{-1}B^\dagger \\ -Q & -A^\dagger \end{pmatrix}}_{=\mathbb{A}\in\mathbb{R}^{(2d)\times(2d)}} \begin{pmatrix} x(t) \\ p(t) \end{pmatrix}$$

On note $R(t) = e^{(T-t)\mathbb{A}}$ la résolvante associée à ce système différentiel (telle que $R(T) = I_{2d}$). On pose

$$R(t) = \begin{pmatrix} R_1(t) & R_2(t) \\ R_3(t) & R_4(t) \end{pmatrix} \in \mathbb{R}^{(2d)\times(2d)},$$

où les quatre blocs sont à valeurs dans $\mathbb{R}^{d\times d}$. On a $x(t) = R_1(t)x(T) + R_2(t)p(T)$ et $p(t) = R_3(t)x(T) + R_4(t)p(T)$. Or $p(T) = Dx(T)$, si bien qu'en posant $\mathcal{X}_T(t) = R_1(t) + R_2(t)D$ et $\mathcal{P}_T(t) = R_3(t) + R_4(t)D$, il vient $x(t) = \mathcal{X}_T(t)x(T)$ et $p(t) = \mathcal{P}_T(t)x(T)$. En conclusion, la matrice $P(t)$ solution de l'équation de Riccati s'obtient également à partir de la résolvante du système linéaire de taille $2d$ ci-dessus en posant

$$P(t) = (R_3(t) + R_4(t)D)(R_1(t) + R_2(t)D)^{-1} \in \mathbb{R}^{d\times d}.$$

Cette expression est intéressante en pratique car elle évite de devoir résoudre un système différentiel non-linéaire. □

Exemple 4.17. [Mouvement d'un point matériel] On reprend l'exemple 4.9 du mouvement d'un point matériel le long d'une droite et dont on contrôle la vitesse, i.e., $\dot{x}_u(t) = u(t)$, pour tout $t \in [0,T]$, et $x_u(0) = x_0$. Le critère à minimiser est à nouveau $J(u) = \frac{1}{2}\int_0^T x_u(t)^2\,\mathrm{d}t + \frac{1}{2}\int_0^T u(t)^2\,\mathrm{d}t$. Ce problème rentre dans le cadre du système LQ avec $d = k = 1$ et

$$A = 0, \quad B = 1, \quad R = 1, \quad Q = 1, \quad D = 0, \quad \xi \equiv 0.$$

L'équation de Riccati pour la fonction $P(t)$, ici à valeurs scalaires, s'écrit

$$\dot{P}(t) = P(t)^2 - 1, \quad \forall t \in [0,T], \qquad P(T) = 0.$$

On obtient $P(t) = \tanh(T-t)$. Le contrôle optimal se met alors sous forme de boucle fermée

$$\overline{u}(t) = K(t)\overline{x}(t), \quad K(t) = -P(t) = -\tanh(T-t).$$

Pour mémoire, on avait trouvé que

$$\begin{aligned} \overline{x}(t) &= \tfrac{x_0}{\cosh(T)}\cosh(T-t), \\ \overline{u}(t) &= -\overline{p}(t) = -\tfrac{x_0}{\cosh(T)}\sinh(T-t), \end{aligned}$$

ce qui permet de retrouver l'expression ci-dessus liant $\overline{u}(t)$ à $\overline{x}(t)$. Enfin, la valeur optimale du critère est $J(\overline{u}) = \frac{1}{2}x_0^2 P(0) = \frac{1}{2}x_0^2\tanh(T)$. □

Chapitre 5

Principe du minimum de Pontryaguine (PMP)

Ce chapitre est consacré au problème de contrôle optimal pour des systèmes non-linéaires. Le résultat phare est le **principe du minimum de Pontryaguine** (PMP) dont nous nous contenterons de donner l'énoncé dans ce chapitre, une esquisse de preuve étant présentée au chapitre suivant. Nous verrons que le PMP ne fournit que des **conditions nécessaires d'optimalité** dont la formulation fait intervenir, comme pour le système LQ du chapitre précédent, les notions d'**état adjoint** et de **Hamiltonien**. En revanche, le PMP ne dit rien sur l'existence d'un contrôle optimal ni sur le caractère suffisant de ces conditions. L'intérêt pratique du PMP est de nous permettre de faire un premier tri des contrôles candidats à l'optimalité ; en espérant que les contrôles vérifiant les conditions nécessaires d'optimalité du PMP ne sont pas trop nombreux, on pourra ensuite les examiner individuellement pour en déterminer le caractère optimal ou non. Afin de nous familiariser avec l'emploi du PMP, nous présentons dans ce chapitre deux exemples d'application : le système LQ avec des contraintes sur le contrôle d'une part et un modèle non-linéaire de dynamique de populations d'autre part.

5.1 Systèmes de contrôle non-linéaires

On se donne un intervalle de temps $[0,T]$ avec $T > 0$, on considère un état à valeurs dans $\mathbb{R}^d$, $d \geq 1$, et un contrôle à valeurs dans un sous-ensemble fermé non-vide $U \subset \mathbb{R}^k$. On s'intéresse au système de contrôle non-linéaire

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in [0,T], \qquad x_u(0) = x_0, \tag{5.1}$$

avec une dynamique décrite par la fonction $f : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$. L'ensemble des contrôles admissibles est ici le sous-ensemble

$$\mathcal{U} = L^1([0,T]; U) \subset L^1([0,T]; \mathbb{R}^k). \tag{5.2}$$

L'objectif est de trouver un contrôle optimal $\overline{u} \in \mathcal{U}$ qui minimise le critère

$$J(u) = \int_0^T g(t, x_u(t), u(t))\, \mathrm{d}t + h(x_u(T)), \tag{5.3}$$

où les fonctions $g : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}$ et $h : \mathbb{R}^d \to \mathbb{R}$ sont données. Le problème de contrôle optimal est donc le suivant :

$$\text{Chercher } \overline{u} \in \mathcal{U} \text{ tel que } J(\overline{u}) = \inf_{u \in \mathcal{U}} J(u). \tag{5.4}$$

Nous allons formuler quelques hypothèses (en général, raisonnables) sur les différents ingrédients intervenant dans la formulation du problème de contrôle optimal (5.4), à savoir la fonction f pour la dynamique et les fonctions g et h pour le critère. Commençons par les hypothèses sur la dynamique. On suppose que

(a) $f \in C^0([0,T] \times \mathbb{R}^d \times U; \mathbb{R}^d)$ et f est de classe C^1 par rapport à x ;

(b) $\exists C$, $|f(t,y,v)|_{\mathbb{R}^d} \leq C(1 + |y|_{\mathbb{R}^d} + |v|_{\mathbb{R}^k})$, $\forall t \in [0,T]$, $\forall y \in \mathbb{R}^d$, $\forall v \in U$;

(c) Pour tout $R > 0$, $\exists C_R$, $|\frac{\partial f}{\partial x}(t,y,v)|_{\mathbb{R}^{d\times d}} \leq C_R(1 + |v|_{\mathbb{R}^d})$, $\forall t \in [0,T]$, $\forall y \in \overline{B}(0,R)$, $\forall v \in U$.

Dans ces hypothèses, C et C_R désignent des constantes génériques indépendantes de (t,y,v), C_R dépendant du rayon R de la boule fermée $\overline{B}(0,R)$; par la suite, nous utiliserons les symboles C et C_R avec la convention que les valeurs de C et de C_R peuvent changer à chaque utilisation tant qu'elles restent indépendantes du temps, de l'état du système et de la valeur du contrôle. L'objectif des trois hypothèses ci-dessus est d'assurer, pour tout contrôle $u \in \mathcal{U}$, l'existence et l'unicité de la trajectoire associée $x_u \in AC([0,T]; \mathbb{R}^d)$.

Lemme 5.1 (Existence et unicité des trajectoires)**.** *Dans le cadre des hypothèses* (a), (b), (c) *ci-dessus, pour tout contrôle $u \in \mathcal{U}$, il existe une unique trajectoire associée $x_u \in AC([0,T]; \mathbb{R}^d)$ solution de* (5.1)*.*

Démonstration. Il s'agit d'une conséquence de la version locale du théorème de Cauchy–Lipschitz avec une dynamique mesurable en temps uniquement (cf. le théorème 2.6). On considère le système dynamique $\dot{x}(t) = F(t,x(t))$ avec la fonction $F : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ telle que $F(t,x) = f(t,x,u(t))$. La fonction F est mesurable en t, et elle est continue en x. De plus, F est localement lipschitzienne par rapport à x puisque l'on a, pour tout $t \in [0,T]$ et tout $x_1, x_2 \in \overline{B}(0,R)$,

$$|F(t,x_1) - F(t,x_2)|_{\mathbb{R}^d} \leq C_0(t)|x_1 - x_2|_{\mathbb{R}^d}, \quad C_0(t) = \sup_{y \in \overline{B}(0,R)} \left|\tfrac{\partial f}{\partial x}(t,y,u(t))\right|_{\mathbb{R}^{d\times d}}.$$

Comme $C_0(t) \leq C_R(1 + |u(t)|_{\mathbb{R}^k})$ grâce à l'hypothèse (c), on a bien $C_0 \in L^1([0,T]; \mathbb{R}_+)$. En outre, la fonction F est localement intégrable grâce à l'hypothèse (b) puisque l'on a, pour tout $x \in \mathbb{R}^d$ et tout $t \in [0,T]$,

$$|F(t,x)|_{\mathbb{R}^d} \leq C(1 + |x|_{\mathbb{R}^d} + |u(t)|_{\mathbb{R}^k}) \in L^1([0,T]; \mathbb{R}_+).$$

Il reste enfin à s'assurer que la trajectoire maximale est bien définie sur tout l'intervalle $[0,T]$ (i.e., qu'il n'y a pas eu d'explosion en un temps $t_* < T$). Pour cela, on utilise le lemme de Gronwall rappelé ci-dessous. Comme on a $x(t) = x_0 + \int_0^t f(s,x(s),u(s))\,\mathrm{d}s$, on peut appliquer ce lemme avec $z(t) = |x(t)|_{\mathbb{R}^d}$ et $\psi(t) \equiv C$. L'estimation (5.5) est satisfaite avec $\alpha = |x_0|_{\mathbb{R}^d} + C(T + \|u\|_{L^1([0,T];\mathbb{R}^k)})$ grâce à l'hypothèse (b). On en déduit que la trajectoire reste bien bornée sur $[0,T]$, i.e., il n'y a pas d'explosion. □

Lemme 5.2 (Gronwall). *Soit $\psi, z : [0,T] \to \mathbb{R}_+$ deux fonctions continues telles que*

$$\exists \alpha \geq 0, \quad \forall t \in [0,T], \quad z(t) \leq \alpha + \int_0^t \psi(s) z(s) \, \mathrm{d}s. \tag{5.5}$$

Alors, on a $z(t) \leq \alpha e^{\int_0^t \psi(s)\,\mathrm{d}s}$ pour tout $t \in [0,T]$.

Démonstration. Posons $\Psi(t) = \int_0^t \psi(s)\,\mathrm{d}s$ et considérons la fonction $v(t) = e^{-\Psi(t)} \int_0^t \psi(s) z(s)\,\mathrm{d}s$. En utilisant (5.5), on constate que

$$\begin{aligned} \frac{dv}{dt}(t) &= -\psi(t) e^{-\Psi(t)} \int_0^t \psi(s) z(s)\,\mathrm{d}s + e^{-\Psi(t)} \psi(t) z(t) \\ &= \psi(t) e^{-\Psi(t)} \left(z(t) - \int_0^t \psi(s) z(s)\,\mathrm{d}s \right) \leq \alpha \psi(t) e^{-\Psi(t)}. \end{aligned}$$

Comme $v(0) = 0$ et $\Psi(0) = 0$, en intégrant cette majoration de 0 à t, il vient

$$e^{-\Psi(t)} \int_0^t \psi(s) z(s)\,\mathrm{d}s = v(t) \leq \alpha \int_0^t \psi(s) e^{-\Psi(s)}\,\mathrm{d}s = \alpha(1 - e^{-\Psi(t)}),$$

et en ré-arrangeant les termes, on obtient

$$\alpha + \int_0^t \psi(s) z(s)\,\mathrm{d}s \leq \alpha e^{\Psi(t)}.$$

On conclut en utilisant à nouveau la borne (5.5) sur $z(t)$. □

Venons en maintenant aux hypothèses sur le critère. On suppose que

(d) $g \in C^0([0,T] \times \mathbb{R}^d \times U; \mathbb{R})$ et g est de classe C^1 par rapport à x ; de plus, $h \in C^1(\mathbb{R}^d; \mathbb{R})$;

(e) Pour tout $R > 0$, $\exists C_R$, $|g(t,y,v)| \leq C_R(1 + |v|_{\mathbb{R}^k})$, $\forall t \in [0,T]$, $\forall y \in \overline{B}(0,R)$, $\forall v \in U$;

(f) Pour tout $R > 0$, $\exists C_R$, $|\frac{\partial g}{\partial x}(t,y,v)|_{\mathbb{R}^d} \leq C_R(1 + |v|_{\mathbb{R}^k})$, $\forall t \in [0,T]$, $\forall y \in \overline{B}(0,R)$, $\forall v \in U$;

(g) Les fonctions g et h sont minorées respectivement sur $[0,T] \times \mathbb{R}^d \times U$ et sur $\mathbb{R}^d$.

Ces hypothèses nous permettent d'affirmer que, pour tout $u \in \mathcal{U}$, le critère $J(u)$ est bien défini car la trajectoire associée x_u est bien définie et $x_u(t) \in \overline{B}(0, R(u))$, pour tout $t \in [0,T]$, si bien que grâce à l'hypothèse (e), la fonction $t \mapsto g(t, x(t), u(t))$ est bien intégrable. En outre, l'infimum de J sur $\mathcal{U}$ est bien fini grâce à l'hypothèse (g). Il est donc raisonnable de considérer le problème de minimisation (5.4). Les hypothèses (d) et (f) nous seront utiles à la section suivante pour définir l'état adjoint.

5.2 PMP : énoncé et commentaires

L'objectif de cette section est d'énoncer le principe du minimum de Pontryaguine (PMP) pour le système de contrôle non-linéaire (5.1) et la fonctionnelle J définie en (5.3). Dans ce chapitre, nous nous contenterons d'énoncer le PMP et d'en voir quelques premiers exemples d'application. La preuve du PMP sera esquissée au chapitre suivant. Comme dans le cas plus simple du système linéaire-quadratique (cf. la section 4.3), le PMP repose sur la notion de Hamiltonien.

Définition 5.3 (Hamiltonien)**.** *Le* ***Hamiltonien*** *associé au système de contrôle non-linéaire* (5.1) *et à la fonctionnelle J définie en* (5.3) *est l'application $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times U \to \mathbb{R}$ telle que*

$$H(t,x,p,u) = p^\dagger f(t,x,u) + g(t,x,u). \tag{5.6}$$

On notera bien que dans cette écriture, (x,p,u) désigne un vecteur générique de $\mathbb{R}^d \times \mathbb{R}^d \times U$. Lorsque l'application H ne dépend pas explicitement du temps, on dit que le Hamiltonien est ***autonome****.*

Théorème 5.4 (PMP)**.** *Si $\overline{u} \in \mathcal{U}$ est un contrôle optimal, i.e., si $\overline{u}$ est une solution de* (5.4)*, alors en notant $\overline{x} = x_{\overline{u}} \in AC([0,T];\mathbb{R}^d)$ la trajectoire associée au contrôle $\overline{u}$ et en définissant l'****état adjoint*** *$\overline{p} \in AC([0,T];\mathbb{R}^d)$ solution de*

$$\frac{d\overline{p}}{dt}(t) = -\overline{A}(t)^\dagger \overline{p}(t) - \overline{b}(t),\ \forall t \in [0,T], \quad \overline{p}(T) = \frac{\partial h}{\partial x}(\overline{x}(T)) \in \mathbb{R}^d, \tag{5.7}$$

où pour tout $t \in [0,T]$,

$$\overline{A}(t) = \frac{\partial f}{\partial x}(t,\overline{x}(t),\overline{u}(t)) \in \mathbb{R}^{d\times d}, \qquad \overline{b}(t) = \frac{\partial g}{\partial x}(t,\overline{x}(t),\overline{u}(t)) \in \mathbb{R}^d, \tag{5.8}$$

on a, p.p. $t \in [0,T]$,

$$\overline{u}(t) \in \arg\min_{v\in U} H(t,\overline{x}(t),\overline{p}(t),v), \tag{5.9}$$

où le Hamiltonien $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times U \to \mathbb{R}$ est défini en (5.6)*. Un triplet $(\overline{x},\overline{p},\overline{u})$ satisfaisant les conditions ci-dessus est appelé une* ***extrémale****. On notera que avec les conventions adoptées, $\frac{\partial g}{\partial x}$ et $\frac{\partial h}{\partial x}$ sont des vecteurs colonne.*

Remarque 5.5. [État adjoint] L'état adjoint $\overline{p}$ est solution d'un système linéaire (à $(\overline{x},\overline{u})$ fixés) instationnaire et rétrograde en temps. Ce système, ainsi que la condition finale sur $\overline{p}(T)$, sont bien définis grâce aux hypothèses (a) et (d) ci-dessus. De plus, ce système admet une unique solution car la fonction $\overline{b}$ est bien intégrable en temps grâce à l'hypothèse (f) et la fonction $\overline{A}$ est dans $L^1([0,T];\mathbb{R}^{d\times d})$ grâce à l'hypothèse (c). □

Remarque 5.6. [Condition nécessaire] Dans le cas du système de contrôle non-linéaire (5.1) avec la fonctionnelle J définie en (5.3), le PMP ne fournit qu'une **condition nécessaire d'optimalité**. En revanche, le PMP ne dit rien sur l'existence d'un contrôle optimal, et il ne fournit pas *en général* de condition suffisante (cf. toutefois la proposition 5.12 ci-dessous). L'intérêt pratique du PMP est de restreindre le champ des possibles en vue de l'obtention d'un contrôle optimal : on commence par considérer les extrémales et, en espérant qu'elles ne sont pas trop nombreuses, on en fait ensuite le tri. □

Remarque 5.7. [Amplitude de $\overline{p}$] Si on multiplie les fonctions g et h par un facteur $\lambda \in \mathbb{R}_+$, le nouveau critère à minimiser est $J_\lambda = \lambda J$, le nouvel état adjoint est $\overline{p}_\lambda = \lambda\overline{p}$, et le nouvel Hamiltonien est $H_\lambda = \lambda H$. Comme H_λ et H ont les mêmes minimiseurs, cela montre que l'amplitude de $\overline{p}$ n'apporte pas d'information en vue de la résolution du problème de contrôle optimal. □

Remarque 5.8. [Hamiltonien autonome] Lorsque le Hamiltonien est autonome, i.e., que l'application H ne dépend pas explicitement du temps, la condition (5.9) devient

$$\overline{u}(t) \in \arg\min_{v\in U} H(\overline{x}(t), \overline{p}(t), v).$$

On observera que le contrôle optimal $\overline{u}$ dépend (en général) du temps car $\overline{x}(t)$ et $\overline{p}(t)$ dépendent (en général) du temps. □

Exemple 5.9. [Système LQ] Appliquons le théorème 5.4 au système LQ étudié au chapitre précédent. Pour simplifier, on omet le terme de dérive. On a

$$f(t,x,u) = Ax + Bu, \quad g(t,x,u) = \tfrac{1}{2}u^\dagger Ru + \tfrac{1}{2}e_x(t)^\dagger Q e_x(t), \quad h(x) = \tfrac{1}{2}e_x(T)^\dagger D e_x(T),$$

où $e_x(t) = x - \xi(t)$; on rappelle que les matrices $Q, D \in \mathbb{R}^{d\times d}$ sont symétriques semi-définies positives, que la matrice $R \in \mathbb{R}^{k\times k}$ est symétrique définie positive et que $\xi \in C^0([0,T];\mathbb{R}^d)$ est la trajectoire cible. Pour le système LQ, il n'y a pas de contraintes sur le contrôle, on a donc $U = \mathbb{R}^k$. Le Hamiltonien s'écrit

$$H(t,x,p,u) = p^\dagger(Ax + Bu) + \tfrac{1}{2}u^\dagger Ru + \tfrac{1}{2}e_x(t)^\dagger Q e_x(t).$$

On a donc (noter l'unicité du minimiseur)

$$\overline{u}(t) = \arg\min_{v\in\mathbb{R}^k} \left(\overline{p}^\dagger Bv + \frac{1}{2}v^\dagger Rv\right),$$

ce qui équivaut à

$$\overline{u}(t) = -R^{-1}B^\dagger \overline{p}(t).$$

Comme $\frac{\partial f}{\partial x} = A$, $\frac{\partial g}{\partial x} = Qe_x$, $\frac{\partial h}{\partial x} = De_x$, l'équation (5.7) sur l'état adjoint devient

$$\frac{d\overline{p}}{dt}(t) = -A^\dagger \overline{p}(t) - Qe_{\overline{x}}(t), \quad \forall t \in [0,T], \qquad \overline{p}(T) = De_{\overline{x}}(T),$$

qui est bien l'équation différentielle rétrograde et la condition finale qui avaient été obtenues au chapitre précédent pour l'état adjoint (cf. le théorème 4.5). □

Exemple 5.10. [Non-existence de contrôle optimal] Donnons un exemple relativement simple de non-existence de contrôle optimal. On considère le système de contrôle linéaire $\dot{x}_u(t) = u(t)$ avec $x_u(0) = x_0 = 0$ et $T = 1$. Le critère à minimiser est

$$J(u) = \int_0^1 x_u(t)^2\,\mathrm{d}t + \int_0^1 (u(t)^2 - 1)^2\,\mathrm{d}t, \quad U = [-1,1].$$

Alors, on a $\inf_{u\in\mathcal{U}} J(u) = 0$ et il n'existe pas de contrôle optimal. Pour le montrer, on considère pour tout $n \in \mathbb{N}_*$ la suite minimisante de contrôles

$$u_n(t) = (-1)^k, \quad t \in [\tfrac{k}{2n}, \tfrac{k+1}{2n}[, \quad k \in \{0,\dots,2n-1\},$$

dont la trajectoire associée, x_n, est en dents de scie et vérifie $\|x_n\|_{L^\infty(0,1)} \le \frac{1}{2n}$ (cf. la figure 5.1). On en déduit que $J(u_n) \le \frac{1}{4n^2}$. S'il existait $\overline{u} \in \mathcal{U}$ tel que $J(\overline{u}) = 0$, alors on aurait $\overline{x}(t) \equiv 0$ et $\overline{u}(t) \in \{-1,1\}$, mais $\overline{u}(t) = \frac{d\overline{x}}{dt}(t) = 0$. La difficulté rencontrée dans cet exemple provient de la non-convexité du critère. □

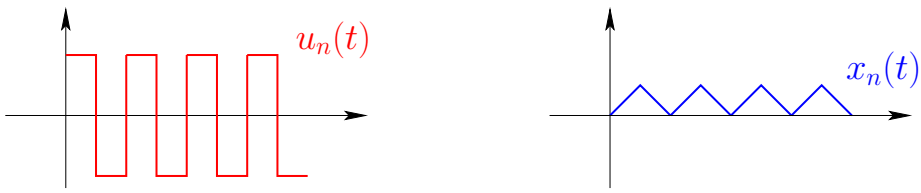

FIGURE 5.1 – Illustration du (contre-)exemple 5.10 : contrôle issu d'une suite minimisante et trajectoire associée.

Exemple 5.11. [Absence de condition suffisante] Donnons maintenant un exemple où le PMP ne fournit pas de condition suffisante d'optimalité. On considère à nouveau le système de contrôle linéaire $\dot{x}(t) = u(t)$ avec $x_0 = 0$ et $T = 1$. Le critère à minimiser est cette fois

$$J(u) = \int_0^1 (x_u(t)^2 - 1)^2 \, \mathrm{d}t, \quad U = [-1,1].$$

On cherche donc à minimiser la distance de $x(t)$ à l'ensemble $\{-1, 1\}$; les contraintes sur u font que $x(t) \in [-1,1]$, $\forall t \in [0,T]$. Il y a donc deux contrôles optimaux, qui sont $\overline{u}_\pm(t) \equiv \pm 1$, pour tout $t \in [0,T]$, et on a $\inf_{u\in\mathcal{U}} J(u) = \int_0^1 (t^2-1)^2 \, \mathrm{d}t = \frac{8}{15}$. Or, si on considère le contrôle $\overline{u}(t) \equiv 0$, celui-ci vérifie les conditions du PMP mais ce n'est pas un contrôle optimal car $J(0) = 1 > \frac{8}{15}$. En effet, on a $f(t,x,u) = u$, $g(t,x,u) = (x^2-1)^2$, $h = 0$, la trajectoire associée est $\overline{x}(t) \equiv 0$ et l'état adjoint est $\overline{p}(t) \equiv 0$. Le Hamiltonien à minimiser est $H(t,\overline{x}(t),\overline{p}(t),v) = (\overline{x}(t)^2-1)^2$ dont un minimiseur est bien $v = 0$. La difficulté rencontrée dans cet exemple provient à nouveau de la non-convexité du critère. □

Concluons cette section par un résultat positif quant au caractère suffisant de la condition d'optimalité du PMP.

Proposition 5.12 (Condition suffisante)**.** *Le PMP fournit une* ***condition suffisante*** *d'optimalité sous les hypothèses suivantes :*

- *$f(t,x,u) = A(t)x + B(t)u$ avec $A \in C^0([0,T];\mathbb{R}^{d\times d})$ et $B \in C^0([0,T];\mathbb{R}^{d\times k})$;*
- *$\mathcal{U} = L^2([0,T];U)$ où U est un ensemble* ***convexe*** *fermé non-vide ;*
- *la fonction g est* ***convexe*** *et différentiable en $(x,u) \in \mathbb{R}^d \times U$;*
- *la fonction h est* ***convexe*** *et différentiable en $x \in \mathbb{R}^d$.*

Démonstration. Nous nous contenterons d'esquisser la preuve. La fonctionnelle J est convexe en u sur l'ensemble convexe $K = L^2([0,T];U)$ (on travaille dans L^2 afin de se placer dans le cadre des espaces de Hilbert). De par la proposition 3.11, $\overline{u}$ est un contrôle optimal dans K si et seulement si

$$(\nabla J(\overline{u}), v - \overline{u})_{L^2([0,T];\mathbb{R}^k)} \geq 0, \quad \forall v \in K.$$

Grâce à l'introduction de l'état adjoint $\overline{p}$ solution de (5.7), ceci se récrit

$$\int_0^T \left(\overline{p}(t)^\dagger B(v(t) - \overline{u}(t)) + \frac{\partial g}{\partial u}(t,\overline{x}(t),\overline{u}(t))^\dagger (v(t) - \overline{u}(t)) \right) \mathrm{d}t \geq 0, \quad \forall v \in K.$$

Cette inégalité, toujours grâce à la proposition 3.11, équivaut au fait que $\overline{u}$ soit minimiseur sur K de la fonctionnelle

$$\tilde{J}(u) = \int_0^T \Big(\overline{p}(t)^\dagger B u(t) + g(t, \overline{x}(t), u(t)) \Big) \, \mathrm{d}t.$$

En raisonnant comme dans la preuve du théorème 3.17, on montre que cela équivaut au fait que $\overline{u}(t)$ soit minimiseur instantané de $v \mapsto \overline{p}(t)^\dagger B v + g(t, \overline{x}(t), v)$, ce qui n'est rien d'autre que minimiser le Hamiltonien par rapport à v. □

5.3 Application au système LQ avec contraintes

L'objectif de cette section est d'illustrer le PMP dans le cas du système LQ (dynamique linéaire et critère quadratique), mais contrairement au chapitre 4, nous supposons ici qu'il y a des contraintes sur le contrôle. Malgré la présence de ces contraintes, ce nouveau problème de contrôle optimal reste relativement simple, et il nous sera en fait possible de prouver le PMP (et d'en établir le caractère suffisant) en nous appuyant sur l'inéquation d'Euler caractérisant le minimiseur d'une fonctionnelle convexe sur un sous-ensemble convexe, fermé, non-vide d'un espace de Hilbert (cf. la proposition 3.11).

Soit $T > 0$, une matrice $A \in \mathbb{R}^{d \times d}$, une matrice $B \in \mathbb{R}^{d \times k}$ et une condition initiale $x_0 \in \mathbb{R}^d$. Le système de contrôle linéaire s'écrit sous la forme

$$\dot{x}_u(t) = A x_u(t) + B u(t), \quad \forall t \in [0,T], \qquad x_u(0) = x_0. \tag{5.10}$$

Soit U un sous-ensemble **convexe, fermé, non-vide** de $\mathbb{R}^k$. L'ensemble des contrôles admissibles est ici le sous-ensemble

$$K = L^2([0,T]; U). \tag{5.11}$$

On s'intéresse au problème de minimisation sous contraintes

$$\text{Chercher } \overline{u} \in K \text{ tel que } J(\overline{u}) = \inf_{u \in K} J(u), \tag{5.12}$$

avec le critère quadratique

$$J(u) = \frac{1}{2} \int_0^T u(t)^\dagger R u(t) \, \mathrm{d}t + \frac{1}{2} \int_0^T e_{x_u}(t)^\dagger Q e_{x_u}(t) \, \mathrm{d}t + \frac{1}{2} e_{x_u}(T)^\dagger D e_{x_u}(T), \tag{5.13}$$

où $e_{x_u} = x_u - \xi$ et $\xi \in C^0([0,T]; \mathbb{R}^d)$ est la trajectoire cible. Comme dans le chapitre 4, les matrices $Q, D \in \mathbb{R}^{d \times d}$ sont symétriques semi-définies positives, tandis que la matrice $R \in \mathbb{R}^{k \times k}$ est symétrique définie positive.

Lemme 5.13 (Existence et unicité)**.** *Il existe une unique solution au problème* (5.12)*, i.e., la fonctionnelle J définie par* (5.13) *admet un unique minimiseur sur le sous-ensemble K défini par* (5.11)*.*

Démonstration. Nous allons appliquer le théorème 3.8. D'une part, K est un sous-ensemble convexe, fermé, non-vide de l'espace de Hilbert $V = L^2([0,T];\mathbb{R}^k)$. En effet,

— K est non-vide car le sous-ensemble U est non-vide (considérer un contrôle constant en temps égal à un élément de U) ;

— K est convexe car le sous-ensemble U est convexe (pour tout $u_1, u_2 \in K$ et $\theta \in [0,1]$, on a $\theta u_1(t)+(1-\theta)u_2(t) \in U$ p.p. $t \in [0,T]$ car U est convexe, si bien que $\theta u_1+(1-\theta)u_2 \in K$) ;

— enfin, K est fermé dans V car si $(u_n)_{n\in\mathbb{N}}$ est une suite de K convergeant vers u dans V, comme la convergence dans $L^2([0,T];\mathbb{R}^k)$ implique la convergence p.p. (à une sous-suite près) et que le sous-ensemble U est fermé, on en déduit que $u(t) \in U$ p.p. $t \in [0,T]$, i.e., $u \in K$.

D'autre part, la fonctionnelle J est fortement convexe et continue (elle est même différentiable) sur V (cf. les lemmes 4.2 et 4.4). □

Dans la suite de cette section, on notera $\overline{u} \in K = L^2([0,T];U)$ l'unique contrôle optimal solution de (5.12) et $\overline{x} = x_{\overline{u}}$ la trajectoire associée. Le système LQ avec contraintes rentre dans le champ d'application du PMP. En procédant comme à l'exemple 5.9 (qui traitait le cas sans contraintes), on introduit l'état adjoint $\overline{p} \in C^1([0,T];\mathbb{R}^d)$ tel que

$$\frac{d\overline{p}}{dt}(t) = -A^\dagger \overline{p}(t) - Qe_{\overline{x}}(t), \quad \forall t \in [0,T], \qquad \overline{p}(T) = De_{\overline{x}}(T), \tag{5.14}$$

où $e_{\overline{x}}(t) = \overline{x}(t) - \xi(t)$ p.p. $t \in [0,T]$, et le Hamiltonien $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ tel que

$$H(t,x,p,u) = p^\dagger(Ax+Bu) + \frac{1}{2}u^\dagger Ru + \frac{1}{2}(x-\xi(t))^\dagger Q(x-\xi(t)). \tag{5.15}$$

En appliquant le PMP (cf. le théorème 5.4), on en déduit qu'une condition nécessaire d'optimalité est que, p.p. $t \in [0,T]$, $\overline{u}(t)$ est un minimiseur de $H(t,\overline{x}(t),\overline{p}(t),v)$ sur U, i.e.,

$$\overline{u}(t) \in \arg\min_{v\in U} H(t,\overline{x}(t),\overline{p}(t),v). \tag{5.16}$$

En inspectant l'expression de H, on voit que de manière équivalente, on a

$$\overline{u}(t) \in \arg\min_{v\in U} \left(v^\dagger B^\dagger \overline{p}(t) + \frac{1}{2}v^\dagger Rv \right). \tag{5.17}$$

Or, la fonctionnelle en v au membre de droite est quadratique et fortement convexe. On en déduit qu'elle admet un unique minimiseur sur le sous-ensemble convexe, fermé, non-vide U de $\mathbb{R}^k$. De manière plus précise, on a donc

$$\overline{u}(t) = \arg\min_{v\in U} \left(v^\dagger B^\dagger \overline{p}(t) + \frac{1}{2}v^\dagger Rv \right). \tag{5.18}$$

Lorsque $U = \mathbb{R}^k$, on retrouve bien le résultat du chapitre 4, à savoir $\overline{u}(t) = -R^{-1}B^\dagger\overline{p}(t)$. Dans le cas général pour le sous-ensemble U, on n'a pas forcément d'expression explicite de $\overline{u}(t)$ en fonction de $\overline{p}(t)$ car celle-ci dépend de la forme du sous-ensemble U.

Proposition 5.14 (Condition nécessaire et suffisante)**.** *La condition* (5.18) *est une* ***condition nécessaire et suffisante*** *d'optimalité pour le problème* (5.12)*. En outre, cette condition définit un unique contrôle optimal $\overline{u} \in K$ et celui-ci est une fonction lipschitzienne du temps.*

Remarque 5.15. [Fonction lipschitzienne] Le fait que le contrôle optimal $\overline{u} \in K$ soit une fonction lipschitzienne du temps montre que pour le système LQ avec contraintes, il n'y a pas de phénomènes de type bang-bang pour le contrôle optimal. □

Démonstration. (1) La fonctionnelle J étant convexe et différentiable sur V, une condition nécessaire et suffisante d'optimalité pour le problème (5.12) est l'inéquation d'Euler (cf. la proposition 3.11)

$$(\nabla J(\overline{u}), v - \overline{u})_V \geq 0, \quad \forall v \in K.$$

En utilisant l'expression de la différentielle de J obtenue au lemme 4.4, on en déduit que

$$(R\overline{u} + B^\dagger \overline{p}, v - \overline{u})_V \geq 0, \quad \forall v \in K,$$

ou encore, en explicitant le produit scalaire dans $V = L^2([0,T];\mathbb{R}^k)$,

$$\int_0^T (v(t) - \overline{u}(t))^\dagger (R\overline{u}(t) + B^\dagger \overline{p}(t))\,\mathrm{d}t \geq 0, \quad \forall v \in K = L^2([0,T];U).$$

En utilisant à nouveau l'inéquation d'Euler, ceci ne signifie rien d'autre que

$$\overline{u} = \arg\min_{v \in K} \mathcal{J}_{\overline{p}}(v),$$

où la fonctionnelle

$$\mathcal{J}_{\overline{p}} : V \to \mathbb{R}, \qquad \mathcal{J}_{\overline{p}}(v) = \int_0^T \left(v(t)^\dagger B^\dagger \overline{p}(t) + \frac{1}{2} v(t)^\dagger R v(t) \right) \mathrm{d}t$$

est quadratique, différentiable et fortement convexe sur V. On pose pour tout $t \in [0,T]$,

$$u_\sharp(t) = \arg\min_{v \in U} \left(v^\dagger B^\dagger \overline{p}(t) + \frac{1}{2} v^\dagger R v \right).$$

De l'inéquation d'Euler dans $U \subset \mathbb{R}^k$, on déduit que pour tout $t \in [0,T]$,

$$(v - u_\sharp(t))^\dagger (R u_\sharp(t) + B^\dagger \overline{p}(t)) \geq 0, \quad \forall v \in U.$$

(2) Montrons que la fonction $u_\sharp(t)$ ainsi définie est lipschitzienne en t sur $[0,T]$. Soit $t_1, t_2 \in [0,T]$. On a

$$\begin{aligned}(u_\sharp(t_2) - u_\sharp(t_1))^\dagger (R u_\sharp(t_1) + B^\dagger \overline{p}(t_1)) \geq 0,\\ (u_\sharp(t_1) - u_\sharp(t_2))^\dagger (R u_\sharp(t_2) + B^\dagger \overline{p}(t_2)) \geq 0.\end{aligned}$$

En posant $\delta u_\sharp = u_\sharp(t_2) - u_\sharp(t_1)$, il vient

$$(\delta u_\sharp)^\dagger R \delta u_\sharp \leq (\delta u_\sharp)^\dagger B^\dagger(\overline{p}(t_1) - \overline{p}(t_2)).$$

Comme la matrice R est par hypothèse définie positive, on en déduit que

$$|u_\sharp(t_2) - u_\sharp(t_1)|_{\mathbb{R}^k} = |\delta u_\sharp|_{\mathbb{R}^k} \leq \lambda_{\min}(R)^{-1} \|B^\dagger\|_{\mathbb{R}^{k\times d}} |\overline{p}(t_2) - \overline{p}(t_1)|_{\mathbb{R}^d},$$

où $\lambda_{\min}(R) > 0$ désigne la plus petite valeur propre de la matrice R. Comme la fonction $t \mapsto \overline{p}(t)$ est de classe C^1 en t, cela montre que la fonction $t \mapsto u_\sharp(t)$ est lipschitzienne en t.
(3) En conclusion, la fonction $u_\sharp : [0,T] \to \mathbb{R}^k$ est mesurable (car lipschitzienne), de carré sommable et à valeurs dans U. On a donc $u_\sharp \in K$. De plus, comme $\overline{u}(t) \in U$ p.p. $t \in [0,T]$, l'inégalité suivante est satisfaite p.p. $t \in [0,T]$:

$$\overline{u}(t)^\dagger B^\dagger \overline{p}(t) + \frac{1}{2}\overline{u}(t)^\dagger R \overline{u}(t) \geq u_\sharp(t)^\dagger B^\dagger \overline{p}(t) + \frac{1}{2} u_\sharp(t)^\dagger R u_\sharp(t).$$

En intégrant cette inégalité de 0 à T, il vient

$$\mathcal{J}_{\overline{p}}(\overline{u}) \geq \mathcal{J}_{\overline{p}}(u_\sharp).$$

Par unicité du minimiseur de $\mathcal{J}_{\overline{p}}$ sur K, on conclut que $\overline{u} = u_\sharp$. □

5.4 Exemple non-linéaire : ruche d'abeilles

On considère un modèle relativement simple de dynamique de populations. Pour fixer les idées, nous allons le décliner dans le contexte de la modélisation d'une ruche d'abeilles. On suppose que dans la ruche, la population d'abeilles $a(t)$ et celle des reines $r(t)$ évolue selon la dynamique

$$\dot{x}(t) = \begin{pmatrix} \dot{a}(t) \\ \dot{r}(t) \end{pmatrix} = \begin{pmatrix} \varphi(u(t))a(t) \\ \gamma u(t) a(t) \end{pmatrix}, \quad \forall t \in [0,T], \tag{5.19}$$

où le contrôle $u \in L^\infty([0,T]; U)$ avec $U = [0,1]$ représente l'effort des abeilles pour fournir des reines et où nous avons introduit la fonction

$$\varphi : [0,1] \to \mathbb{R}, \quad \varphi(v) = \alpha(1-v) - \beta. \tag{5.20}$$

Les paramètres du modèle α, β, γ sont des réels strictement positifs et on suppose que $\alpha > \beta$. On suppose également que $a(0) > 0$; comme $\dot{a}(t) = \varphi(u(t))a(t)$, on a $a(t) > 0$ pour tout $t \in [0,T]$. On notera également que

- si u est constant égal à 1, on a $\dot{a}(t) = -\beta a(t) < 0$: la population d'abeilles décroît (exponentiellement) ;
- si u est constant égal à 0, on a $\dot{a}(t) = (\alpha - \beta)a(t) > 0$: la population d'abeilles croît (exponentiellement).

Notre objectif ici est de chercher un contrôle optimal afin de maximiser la population de reines au temps T. En introduisant la fonctionnelle $J : \mathcal{U} = L^1([0,T];U) \to \mathbb{R}$ telle que

$$J(u) = -r(T), \tag{5.21}$$

le problème de contrôle optimal est donc le suivant :

$$\text{Chercher } \overline{u} \in \mathcal{U} \text{ tel que } J(\overline{u}) = \inf_{u\in\mathcal{U}} J(u). \tag{5.22}$$

On commence par chercher une condition nécessaire d'optimalité en appliquant le PMP. L'état de la ruche est décrit par le vecteur $x = (a, r)^\dagger \in \mathbb{R}^2$. Le problème de contrôle optimal (5.22) rentre dans le cadre d'application du PMP en posant

$$f(x,u) = \begin{pmatrix} \varphi(u)a \\ \gamma u a \end{pmatrix}, \quad g(x,u) = 0, \quad h(x) = -r. \tag{5.23}$$

Soit $\overline{u} \in \mathcal{U}$ un contrôle optimal, de trajectoire associée $(\overline{a}, \overline{r})^\dagger$. Comme $\frac{\partial f}{\partial x}(x,u) = \left(\begin{smallmatrix} \varphi(u) & 0 \\ \gamma u & 0 \end{smallmatrix}\right)$ et $\frac{\partial g}{\partial x}(x,u) = 0$, l'état adjoint $\overline{p} = (\overline{p}_a, \overline{p}_r)^\dagger : [0,T] \to \mathbb{R}^2$ est tel que

$$\begin{cases} \dfrac{d\overline{p}_a}{dt}(t) = -\varphi(\overline{u}(t))\overline{p}_a(t) - \gamma\overline{u}(t)\overline{p}_r(t), \\ \dfrac{d\overline{p}_r}{dt}(t) = 0, \end{cases} \quad \forall t \in [0,T], \tag{5.24}$$

et la condition finale sur l'état adjoint est

$$\overline{p}(T) = (\overline{p}_a(T), \overline{p}_r(T)) = (0, -1)^\dagger. \tag{5.25}$$

On a donc

$$\frac{d\overline{p}_a}{dt}(t) = -\varphi(\overline{u}(t))\overline{p}_a(t) + \gamma\overline{u}(t), \quad \overline{p}_r(t) \equiv -1, \quad \forall t \in [0,T]. \tag{5.26}$$

Par ailleurs, le Hamiltonien est autonome (cf. la définition 5.3) et s'écrit sous la forme

$$H(x,p,u) = p_a\varphi(u)a + \gamma p_r u a. \tag{5.27}$$

La condition de minimisation (5.9) s'écrit, en utilisant le fait que $\overline{a}(t) \neq 0$ pour tout $t \in [0,T]$,

$$\overline{u}(t) \in \arg\min_{v\in[0,1]} \psi(t)v, \tag{5.28}$$

où la **fonction de commutation** est donnée par

$$\psi(t) = -\overline{p}_a(t)\alpha - \gamma. \tag{5.29}$$

La solution du problème de minimisation (5.28) est élémentaire ; on obtient, pour tout $t \in [0,T]$,

— si $\psi(t) > 0$, $\overline{u}(t) = 0$;
— si $\psi(t) = 0$, $\overline{u}(t) \in [0,1]$;

— si $\psi(t) < 0$, $\overline{u}(t) = 1$.

Le contrôle optimal est donc nécessairement bang-bang, sauf si $\overline{p}_a(t) = -\frac{\gamma}{\alpha}$ sur un sous-intervalle de temps de mesure strictement positive. Reprenons alors l'équation de l'état adjoint :

— si $\overline{p}_a(t) > -\frac{\gamma}{\alpha}$, $\overline{u}(t) = 1$, et on a $\frac{d}{dt}\overline{p}_a(t) = \beta\overline{p}_a(t) + \gamma \geq 0$, i.e., $t \mapsto \overline{p}_a(t)$ est croissante ;

— si $\overline{p}_a(t) < -\frac{\gamma}{\alpha}$, $\overline{u} = 0$, et on a $\frac{d}{dt}\overline{p}_a(t) = (\beta - \alpha)\overline{p}_a(t) \geq 0$, i.e., $t \mapsto \overline{p}_a(t)$ est encore croissante ;

— enfin, il ne peut exister d'intervalle de mesure strictement positive où $\overline{p}_a$ est constant et égal à $-\frac{\gamma}{\alpha}$; en effet, dans ces conditions, on aurait $\varphi(u(t))\frac{\gamma}{\alpha} + \gamma u(t) = 1 - \frac{\beta}{\alpha} \neq 0$, donc $\overline{p}_a$ ne pourrait pas être constant.

Nous pouvons maintenant terminer la résolution du problème. Au temps final, $\psi(T) = -\gamma < 0$, ce qui montre que $\overline{u}(T) = 1$, i.e., au temps final, le contrôle optimal consiste à fournir des reines (ce qui n'est pas très surprenant puisque l'objectif est d'en maximiser le nombre). Le point qui reste à préciser est s'il est optimal d'en fournir depuis l'instant initial ou s'il convient plutôt de laisser d'abord croître la population d'abeilles avant de commencer à en fournir. Comme la fonction de commutation est continue, il existe un temps $t_* < T$ tel que $\overline{u}(t) = 1$ sur $]t_*, T]$. Sur cet intervalle, on a $\frac{d\overline{p}_a}{dt}(t) = \beta\overline{p}_a(t) + \gamma$ et par ailleurs la condition finale sur $\overline{p}_a$ étant $\overline{p}_a(T) = 0$, on en déduit que

$$\overline{p}_a(t) = -\frac{\gamma}{\beta}\left(1 - e^{\beta(t-T)}\right), \quad \forall t \in [t_*, T]. \tag{5.30}$$

La fonction $\overline{p}_a$ est donnée par l'expression ci-dessus tant que le contrôle optimal $\overline{u}$ reste égal à 1. Pour que la valeur du contrôle change, la fonction de commutation (qui est continue) doit s'annuler, i.e., $\overline{p}_a(t_*) = -\frac{\gamma}{\alpha}$. En utilisant l'expression de $\overline{p}_a$, on obtient

$$t_* = \frac{1}{\beta}\ln(1 - \frac{\beta}{\alpha}) + T. \tag{5.31}$$

On notera que $t_* < T$. Deux cas peuvent alors se produire en fonction des paramètres du problème.

— **Cas 1.** $t_* < 0$ (ce qui correspond au cas d'un horizon temporel T petit) ; le contrôle optimal est alors $\overline{u} \equiv 1$ sur [0,T], ce qui signifie que l'on fournit des reines en continu depuis $t = 0$ jusqu'à $t = T$;

— **Cas 2.** $t_* > 0$ (ce qui correspond au cas d'un horizon temporel T relativement grand) ; le contrôle optimal est $\overline{u} \equiv 0$ sur $[0, t_*[$ et $\overline{u} \equiv 1$ sur $]t_*, T]$. En effet, le contrôle $\overline{u}$ vérifie bien le PMP car $\frac{d\overline{p}_a}{dt}(t) = (\beta - \alpha)\overline{p}_a(t)$, $\overline{p}_a(t_*) = -\frac{\gamma}{\alpha}$, d'où $\overline{p}_a(t) = -\frac{\gamma}{\alpha}e^{(\beta-\alpha)(t-t_*)} < -\frac{\gamma}{\alpha}$ sur $[0, t_*]$, si bien que la fonction de commutation est positive, ce qui correspond bien à $\overline{u}(t) = 0$. L'ensemble $\{t \in [0,T] \mid \psi(t) = 0\}$ est réduit au singleton $\{t_*\}$ et est donc de mesure nulle.

Une illustration de la trajectoire, de l'état adjoint et du contrôle optimal est présentée à la figure 5.2 dans le cas où il y a une commutation.

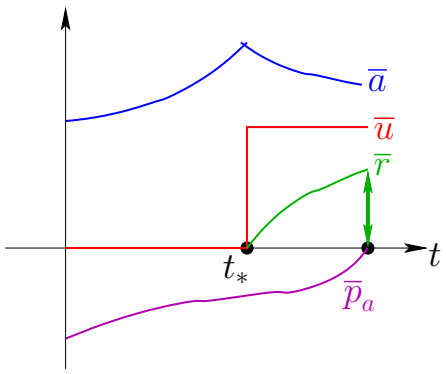

FIGURE 5.2 – Trajectoire, état adjoint et contrôle optimal pour le modèle de ruche.

Chapitre 6

PMP : preuve, extensions, application

Ce chapitre est consacré au principe du minimum de Pontryaguine (PMP) introduit au chapitre précédent. Dans ce chapitre, nous en esquissons la **preuve**, puis nous présentons une extension du PMP au cas où on rajoute une contrainte sur l'atteinte d'une **variété cible** au temps final. Nous présentons également un exemple d'application couvrant plusieurs cas de figure : le problème de Zermelo où on considère une barque traversant un canal sous un courant fort et où on cherche à atteindre la berge opposée en minimisant le déport latéral ou encore en minimisant le temps de traversée. Enfin, nous présentons une méthode de résolution numérique basée sur le PMP et utile dans les applications : la **méthode de tir**.

6.1 PMP : esquisse de preuve

On reprend le système de contrôle non-linéaire considéré à la section 5.1. On rappelle que la dynamique s'écrit sous la forme

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in [0,T], \qquad x_u(0) = x_0, \tag{6.1}$$

avec $T > 0$, $f : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$ et $x_0 \in \mathbb{R}^d$. L'ensemble des contrôles admissibles est

$$\mathcal{U} = L^1([0,T]; U), \tag{6.2}$$

où U est un sous-ensemble fermé non-vide de $\mathbb{R}^k$. L'objectif est de trouver un contrôle optimal $\overline{u} \in \mathcal{U}$ qui minimise le critère

$$J(u) = \int_0^T g(t, x_u(t), u(t))\,\mathrm{d}t + h(x_u(T)), \tag{6.3}$$

où les fonctions $g : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}$ et $h : \mathbb{R}^d \to \mathbb{R}$ sont données. Le problème de contrôle optimal est donc le suivant :

$$\text{Chercher } \overline{u} \in \mathcal{U} \text{ tel que } J(\overline{u}) = \inf_{u \in \mathcal{U}} J(u). \tag{6.4}$$

On rappelle les hypothèses qui avaient été introduites afin de garantir l'existence et l'unicité d'une trajectoire x_u pour un contrôle donné $u \in \mathcal{U}$ (cf. en particulier le lemme 5.1) et le fait que la fonctionnelle $J(u)$ est bien définie :

(a) $f \in C^0([0,T] \times \mathbb{R}^d \times U; \mathbb{R}^d)$ et f est de classe C^1 par rapport à x ;

(b) $\exists C$, $|f(t,y,v)|_{\mathbb{R}^d} \leq C(1 + |y|_{\mathbb{R}^d} + |v|_{\mathbb{R}^k})$, $\forall t \in [0,T]$, $\forall y \in \mathbb{R}^d$, $\forall v \in U$;

(c) Pour tout $R > 0$, $\exists C_R$, $|\frac{\partial f}{\partial x}(t,y,v)|_{\mathbb{R}^{d\times d}} \leq C_R(1 + |v|_{\mathbb{R}^d})$, $\forall t \in [0,T]$, $\forall y \in \overline{B}(0,R)$, $\forall v \in U$;

(d) $g \in C^0([0,T] \times \mathbb{R}^d \times U; \mathbb{R})$ et g est de classe C^1 par rapport à x ; de plus, $h \in C^1(\mathbb{R}^d; \mathbb{R})$;

(e) Pour tout $R > 0$, $\exists C_R$, $|g(t,y,v)| \leq C_R(1 + |v|_{\mathbb{R}^k})$, $\forall t \in [0,T]$, $\forall y \in \overline{B}(0,R)$, $\forall v \in U$;

(f) Pour tout $R > 0$, $\exists C_R$, $|\frac{\partial g}{\partial x}(t,y,v)|_{\mathbb{R}^d} \leq C_R(1 + |v|_{\mathbb{R}^k})$, $\forall t \in [0,T]$, $\forall y \in \overline{B}(0,R)$, $\forall v \in U$;

(g) Les fonctions g et h sont minorées respectivement sur $[0,T] \times \mathbb{R}^d \times U$ et sur $\mathbb{R}^d$.

Dans ces hypothèses, C et C_R désignent des constantes génériques indépendantes de (t,y,v), C_R dépendant du rayon R de la boule fermée $\overline{B}(0,R)$; comme précédemment, nous continuons à utiliser les symboles C et C_R avec la convention que les valeurs de C et de C_R peuvent changer à chaque utilisation tant qu'ils restent indépendants du temps, de l'état du système et de la valeur du contrôle.

Rappelons enfin l'énoncé du PMP (cf. le théorème 5.4).

Théorème 6.1 (PMP)**.** *Si $\overline{u} \in \mathcal{U}$ est un contrôle optimal, i.e., si $\overline{u}$ est une solution de* (6.4)*, alors, en notant $\overline{x} = x_{\overline{u}} \in AC([0,T]; \mathbb{R}^d)$ la trajectoire associée à $\overline{u}$, et en définissant l'**état adjoint** $\overline{p} \in AC([0,T]; \mathbb{R}^d)$ solution de*

$$\frac{d\overline{p}}{dt}(t) = -\overline{A}(t)^\dagger \overline{p}(t) - \overline{b}(t), \quad \forall t \in [0,T], \qquad \overline{p}(T) = \frac{\partial h}{\partial x}(\overline{x}(T)), \tag{6.5}$$

avec $\overline{A}(t) = \frac{\partial f}{\partial x}(t, \overline{x}(t), \overline{u}(t)) \in \mathbb{R}^{d\times d}$ et $\overline{b}(t) = \frac{\partial g}{\partial x}(t, \overline{x}(t), \overline{u}(t)) \in \mathbb{R}^d$ pour tout $t \in [0,T]$, on a, p.p. $t \in [0,T]$,

$$\overline{u}(t) \in \arg\min_{v\in U} H(t, \overline{x}(t), \overline{p}(t), v), \tag{6.6}$$

*où le **Hamiltonien** $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times U \to \mathbb{R}$ est défini par*

$$H(t,x,p,u) = p^\dagger f(t,x,u) + g(t,x,u). \tag{6.7}$$

On rappelle enfin qu'un triplet $(\overline{x}, \overline{p}, \overline{u})$ satisfaisant les conditions ci-dessus est appelé une **extrémale** et que le PMP ne fournit qu'une **condition nécessaire d'optimalité** ; en revanche, il ne dit rien sur l'existence d'un contrôle optimal et il ne fournit pas *a priori* de condition suffisante.

Démonstration. Nous allons nous contenter de donner une esquisse de la preuve, en insistant sur les idées principales sans nécessairement fournir tous les détails techniques pour certains résultats intermédiaires. Ce qui compte ici est donc davantage l'esprit de la démonstration que sa lettre.
(1) L'idée fondamentale est de tester l'optimalité de $J(\overline{u})$ en faisant des **variations aiguille** : il s'agit de perturbations de $\overline{u}$ d'ordre un(!) mais sur un intervalle de temps de longueur très petite $\delta \ll 1$. Soit $t \in [0,T[$ et $\delta \in]0, T-t[$, avec $\delta \ll 1$. La perturbation reste donc petite

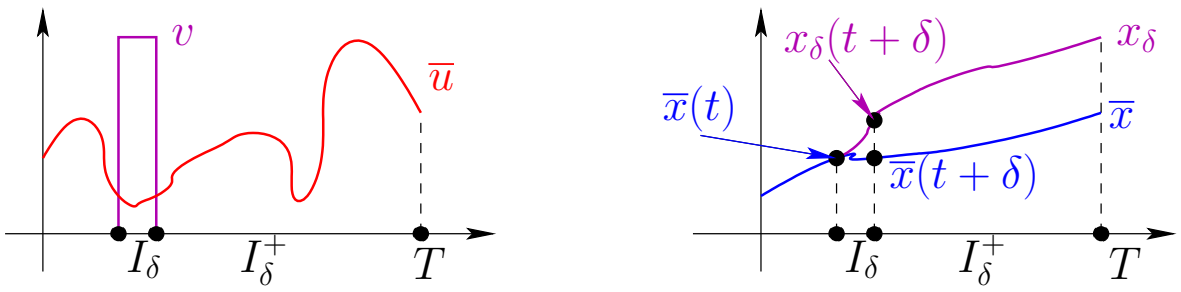

FIGURE 6.1 – Principe de la variation aiguille pour le contrôle optimal $\overline{u}$ (à gauche), trajectoire optimale et trajectoire perturbée (à droite).

dans $L^1([0,T];\mathbb{R}^k)$. Soit $v \in U$ arbitraire. On pose $I_\delta = [t,t+\delta]$ et on considère le contrôle perturbé

$$u_\delta(t) = \begin{cases} \overline{u}(t), & \forall t \in [0,T] \setminus I_\delta, \\ v, & \forall t \in I_\delta. \end{cases}$$

On note x_δ la trajectoire associée au contrôle perturbé. On admet par la suite que p.p. $t \in [0,T[$ (de tels points sont appelés points de Lebesgue), pour $\psi = f$ et $\psi = g$,

$$\lim_{\delta \to 0^+} \frac{1}{\delta} \int_{I_\delta} \psi(s, \overline{x}(s), \overline{u}(s))\, \mathrm{d}s = \psi(t, \overline{x}(t), \overline{u}(t)).$$

On suppose dans la suite de la preuve que t est un point de Lebesgue ; le résultat ci-dessus justifie donc que l'on considère bien tous les instants $t \in [0,T]$ à un sous-ensemble de mesure nulle près.
(2) Comparaison des trajectoires. Comme $x_\delta(t) = \overline{x}(t)$ et $x_\delta(s) = \overline{x}(s) + O(\delta)$ pour tout $s \in I_\delta$, on peut invoquer la continuité de f en (t,x) et la propriété des points de Lebesgue afin d'obtenir les estimations suivantes :

$$x_\delta(t+\delta) = \overline{x}(t) + \int_{I_\delta} f(s, x_\delta(s), v)\, \mathrm{d}s = \overline{x}(t) + \delta f(t, \overline{x}(t), v) + o(\delta),$$
$$\overline{x}(t+\delta) = \overline{x}(t) + \int_{I_\delta} f(s, \overline{x}(s), \overline{u}(s))\, \mathrm{d}s = \overline{x}(t) + \delta f(t, \overline{x}(t), \overline{u}(t)) + o(\delta),$$

si bien que

$$x_\delta(t+\delta) - \overline{x}(t+\delta) = \delta(f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t))) + o(\delta).$$

Une illustration est présentée à la figure 6.1. On va maintenant comparer $x_\delta(s)$ et $\overline{x}(s)$ pour tout $s \in I_\delta^+ = [t+\delta,T]$. Il est clair que $x_\delta(s) - \overline{x}(s) = O(\delta)$ pour tout $s \in I_\delta^+$, et on cherche à préciser la différence à l'ordre un en δ. On introduit la solution $y_\delta \in AC(I_\delta^+;\mathbb{R}^d)$ de l'équation différentielle

$$\dot{y}_\delta(s) = \overline{A}(s) y_\delta(s), \quad \forall s \in I_\delta^+, \qquad y_\delta(t+\delta) = f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t)),$$

où on rappelle que $\overline{A}(s) = \frac{\partial f}{\partial x}(s, \overline{x}(s), \overline{u}(s))$. On en déduit que

$$x_\delta(s) - \overline{x}(s) = \delta y_\delta(s) + \Phi_\delta(s), \quad \forall s \in I_\delta^+, \qquad \Phi_\delta = o(\delta) \text{ unif. sur } I_\delta^+.$$

En effet, on a vu que $\Phi_\delta(t+\delta) = o(\delta)$ et $\dot{\Phi}_\delta(s) = \Psi_\delta(s) + \overline{A}(s)\Phi_\delta(s)$, pour tout $s \in I_\delta^+$, où $\Psi_\delta(s) = o(s)$ uniformément sur I_δ^+, car

$$\Psi_\delta(s) = f(s, x_\delta(s), \overline{u}(s)) - f(s, \overline{x}(s), \overline{u}(s)) - \overline{A}(s)(x_\delta(s) - \overline{x}(s)).$$

En conclusion de cette première étape de la preuve, on a donc

$$x_\delta(s) - \overline{x}(s) = \delta y_\delta(s) + o(\delta) \text{ unif. sur } I_\delta^+.$$

(3) Comparaison des critères. Grâce à la comparaison des trajectoires, à la continuité de g en (t,x) et à la propriété des points de Lebesgue, il vient

$$\begin{aligned}
J(u_\delta) - J(\overline{u}) &= \int_t^T g(s, x_\delta(s), u_\delta(s)) - g(s, \overline{x}(s), \overline{u}(s))\,\mathrm{d}s + h(x_\delta(T)) - h(\overline{x}(T)) \\
&= \int_{I_\delta} g(s, x_\delta(s), v) - g(s, \overline{x}(s), \overline{u}(s))\,\mathrm{d}s + \int_{I_\delta^+} g(s, x_\delta(s), \overline{u}(s)) - g(s, \overline{x}(s), \overline{u}(s))\,\mathrm{d}s \\
&\quad + \delta \frac{\partial h}{\partial x}(\overline{x}(T))^\dagger y_\delta(T) + o(\delta) \\
&= \delta(g(t, \overline{x}(t), v) - g(t, \overline{x}(t), \overline{u}(t))) + \delta \int_{t+\delta}^T \overline{b}(s)^\dagger y_\delta(s)\,\mathrm{d}s \\
&\quad + \delta \frac{\partial h}{\partial x}(\overline{x}(T))^\dagger y_\delta(T) + o(\delta),
\end{aligned}$$

où on rappelle que $\overline{b}(s) = \frac{\partial g}{\partial x}(s, \overline{x}(s), \overline{u}(s))$. L'optimalité de $\overline{u}$ implique donc que

$$0 \le g(t, \overline{x}(t), v) - g(t, \overline{x}(t), \overline{u}(t)) + \int_{t+\delta}^T \overline{b}(s)^\dagger y_\delta(s)\,\mathrm{d}s + \frac{\partial h}{\partial x}(\overline{x}(T))^\dagger y_\delta(T) + o(1).$$

(4) Introduction de l'état adjoint et conclusion. L'état adjoint $\overline{p}$, qui est par définition tel que $\frac{d\overline{p}}{dt}(s) = -\overline{A}(s)^\dagger \overline{p}(s) - \overline{b}(s)$ sur $[0,T]$ et $\overline{p}(T) = \frac{\partial h}{\partial x}(\overline{x}(T))$, nous permet d'éliminer la fonction y_δ. En effet, il vient

$$\begin{aligned}
\int_{t+\delta}^T \overline{b}(s)^\dagger y_\delta(s)\,\mathrm{d}s + \frac{\partial h}{\partial x}(\overline{x}(T))^\dagger y_\delta(T) &= \int_{t+\delta}^T \Big(-\frac{d\overline{p}}{dt}(s) - \overline{A}(s)^\dagger \overline{p}(s) \Big)^\dagger y_\delta(s)\,\mathrm{d}s + \overline{p}(T)^\dagger y_\delta(T) \\
&= \int_{t+\delta}^T -\frac{d\overline{p}}{dt}(s)^\dagger y_\delta(s)\,\mathrm{d}s + \overline{p}(T)^\dagger y_\delta(T) \\
&= \overline{p}(t+\delta)^\dagger y_\delta(t+\delta) \\
&= \overline{p}(t+\delta)^\dagger (f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t))).
\end{aligned}$$

En faisant tendre $\delta \downarrow 0$, il vient par continuité de $\overline{p}$,

$$0 \le g(t, \overline{x}(t), v) - g(t, \overline{x}(t), \overline{u}(t)) + \overline{p}(t)^\dagger (f(t, \overline{x}(t), v) - f(t, \overline{x}(t), \overline{u}(t))),$$

et en utilisant la définition du Hamiltonien, on obtient

$$0 \leq H(t, \overline{x}(t), \overline{p}(t), v) - H(t, \overline{x}(t), \overline{p}(t), \overline{u}(t)),$$

ce qui conclut la preuve car v est arbitraire dans U. □

6.2 Extensions du PMP : atteinte de cible

On considère à nouveau le système de contrôle non-linéaire présenté à la section 6.1, mais on rajoute la contrainte d'**atteindre une variété cible** M à l'instant $t = T$, i.e.,

$$x_u(T) \in M, \tag{6.8}$$

où M est une variété différentielle de classe C^1 de dimension $0 \leq d' \leq d$. Le cas d'une cible ponctuelle correspond à $M = \{x_1\}$ avec $x_1 \in \mathbb{R}^d$ (et $d' = 0$), et l'absence de contrainte de cible (ou contrainte de cible triviale) correspond au cas où $M = \mathbb{R}^d$ (et $d' = d$). L'ensemble des contrôles admissibles devient

$$\mathcal{U}_M = \{u \in L^1([0,T]; U) \mid x_u(T) \in M\}, \tag{6.9}$$

et le problème de contrôle optimal devient

$$\text{Chercher } \overline{u} \in \mathcal{U}_M \text{ tel que } J(\overline{u}) = \inf_{u \in \mathcal{U}_M} J(u), \tag{6.10}$$

où le critère J est toujours défini par (6.3).

Définition 6.2 (Espace tangent)**.** *En un point $x_1 \in M$, l'**espace tangent** $T_{x_1}M$ est l'ensemble des vecteurs vitesse des courbes tracées sur M passant par x_1. Si $M = \{x_1\}$ (cible ponctuelle), on a $T_{x_1}M = \{0\}$, et si $M = \mathbb{R}^d$ (pas de contrainte de cible), on a $T_{x_1}M = \mathbb{R}^d$. Une illustration de l'espace tangent est présentée à la figure 6.2.*

Nous admettrons le résultat suivant (voir par exemple la référence [7], ainsi que le théorème 7.18 dans [11] pour un énoncé plus général où la donnée initiale est uniquement prescrite dans une variété).

Théorème 6.3 (PMP avec cible)**.** *Si $\overline{u} \in \mathcal{U}_M$ est un contrôle optimal, i.e., si $\overline{u}$ est une solution de* (6.10)*, alors, en notant $\overline{x} = x_{\overline{u}} \in AC([0,T]; \mathbb{R}^d)$ la trajectoire associée à $\overline{u}$, on a, p.p. $t \in [0,T]$,*

$$\overline{u}(t) = \arg\min_{v \in U} H(t, \overline{x}(t), \overline{p}(t), \lambda, v), \tag{6.11}$$

*où $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}_+ \times U \to \mathbb{R}$ est le **Hamiltonien** tel que*

$$H(t, x, p, \lambda, u) = p^\dagger f(t, x, u) + \lambda g(t, x, u), \tag{6.12}$$

et où le couple $(\overline{p}, \lambda) \in AC([0,T]; \mathbb{R}^d) \times \mathbb{R}_+$ est tel que $(p, \lambda) \neq (0, 0)$ et l'état adjoint $\overline{p}$ vérifie

$$\frac{d\overline{p}}{dt}(t) = -\overline{A}(t)^\dagger \overline{p}(t) - \lambda \overline{b}(t), \ \forall t \in [0,T], \tag{6.13}$$

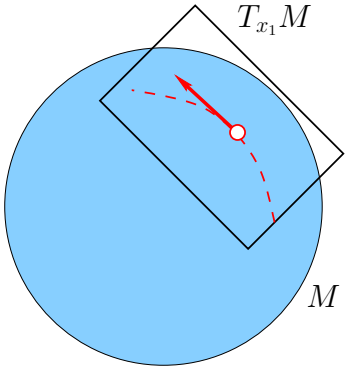

FIGURE 6.2 – Variété M (homéomorphe à une sphère), point $x_1 \in M$ et espace tangent $T_{x_1}M$.

où $\overline{A}(t) = \frac{\partial f}{\partial x}(t, \overline{x}(t), \overline{u}(t)) \in \mathbb{R}^{d\times d}$ *et* $\overline{b}(t) = \frac{\partial g}{\partial x}(t, \overline{x}(t), \overline{u}(t)) \in \mathbb{R}^d$. *Enfin, on a la* ***condition de transversalité*** *sur l'état adjoint au temps final*

$$\overline{p}(T) - \lambda \frac{\partial h}{\partial x}(\overline{x}(T)) \perp T_{\overline{x}(T)}M. \tag{6.14}$$

Un quadruplet $(\overline{x}, \overline{p}, \lambda, \overline{u})$ *satisfaisant les conditions ci-dessus est appelé une* ***extrémale****.*

Remarque 6.4. [Extrêmales normales et anormales] Comme $(p, \lambda) \neq (0, 0)$, deux cas peuvent se produire :

- $\lambda \neq 0$: le PMP étant invariant par un facteur d'échelle positif sur $(\overline{p}, \lambda)$, on peut supposer que $\lambda = 1$; on dit que l'extrémale est **normale** ;
- $\lambda = 0$: on a nécessairement $\overline{p} \not\equiv 0$; on dit que l'extrémale est **anormale**.

Lorsque $M = \mathbb{R}^d$ (pas de contrainte de cible), toute extrémale est normale. En effet, la condition de transversalité devient $\overline{p}(T) = \lambda \frac{\partial h}{\partial x}(\overline{x}(T))$ car $T_{\overline{x}(T)}M = \mathbb{R}^d$; si $\lambda = 0$, on a $\overline{p}(T) = 0$ et la dynamique de $\overline{p}$ implique que $\overline{p} \equiv 0$, ce qui est exclu. On a donc bien $\lambda \neq 0$ dans le cas où $M = \mathbb{R}^d$. En revanche, lorsque la variété M est de dimension $d' < d$, il peut y avoir des extrémales anormales. □

Remarque 6.5. [Méthode de pénalisation] On peut remplacer la contrainte de cible par une pénalisation dans le critère, ce qui conduit au problème (noter que ce problème est à nouveau posé sur $\mathcal{U}$ et non plus sur $\mathcal{U}_M$ comme (6.10))

$$\text{Chercher } \overline{u} \in \mathcal{U} \text{ tel que } J_\epsilon(\overline{u}) = \inf_{u\in\mathcal{U}} J_\epsilon(u),$$

avec la fonctionnelle pénalisée

$$J_\epsilon(u) = J(u) + \frac{1}{\epsilon} d(x_u(T), M)^2.$$

Si le coefficient de pénalisation ϵ est petit, on s'attend à ce que $d(x_u(T), M)$ soit également petit. On peut alors étudier les extrémales $(\overline{x}_\epsilon, \overline{p}_\epsilon, \overline{u}_\epsilon)$ du système pénalisé. Si $\overline{p}_\epsilon$ reste borné

quand $\epsilon \to 0^+$, l'extrémale du système pénalisé tend vers une extrémale normale du système contraint. En revanche, si $|\overline{p}_\epsilon|_{\mathbb{R}^d} \to +\infty$ quand $\epsilon \to 0^+$, on obtient une extrémale anormale. On pourra consulter la section 4.6 de [8] pour approfondir ces aspects. □

Remarque 6.6. [Cas où T n'est pas fixé] Lorsque l'horizon temporel T pour rejoindre la cible M n'est pas fixé *a priori*, on montre (voir à nouveau [7]) qu'on a également une **condition de transversalité** sur le Hamiltonien au temps final T, qui s'écrit

$$\min_{v \in U} H(T, \overline{x}(T), \overline{p}(T), \lambda, v) = 0. \tag{6.15}$$

Le **Hamiltonien minimisé** est la fonction $\mathcal{H} : [0,T] \to \mathbb{R}$ telle que

$$\mathcal{H}(t) = H(t, \overline{x}(t), \overline{p}(t), \lambda, \overline{u}(t)).$$

Comme le contrôle optimal $\overline{u}(t)$ minimise le Hamiltonien, en supposant que $U = \mathbb{R}^k$ (ou que $\overline{u}(t)$ appartient à l'intérieur de U), on en déduit que

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

De plus, en supposant suffisamment de régularité en temps pour que les manipulations ci-dessous soient licites, on observe que

$$\begin{aligned}
\frac{d\overline{x}}{dt}(t) &= \frac{\partial H}{\partial p}(t, \overline{x}(t), \overline{p}(t), \lambda, \overline{u}(t)), \\
\frac{d\overline{p}}{dt}(t) &= -\frac{\partial H}{\partial x}(t, \overline{x}(t), \overline{p}(t), \lambda, \overline{u}(t)).
\end{aligned}$$

Par conséquent, si le Hamiltonien est autonome, i.e., si $\frac{\partial H}{\partial t} = 0$, on a

$$\frac{d\mathcal{H}}{dt}(t) = \frac{d}{dt} H(\overline{x}(t), \overline{p}(t), \lambda, \overline{u}(t)) = 0.$$

Le Hamiltonien minimisé étant nul en T de par la condition de transversalité (6.15), on conclut que

$$\mathcal{H}(t) = H(\overline{x}(t), \overline{p}(t), \lambda, \overline{u}(t)) = 0, \quad \forall t \in [0,T],$$

i.e., le Hamiltonien minimisé pour un système autonome sans horizon temporel fixé est identiquement nul. □

Exemple 6.7. [Temps-optimalité (cas linéaire autonome)] Afin d'illustrer la remarque 6.6, considérons le problème de temps-optimalité pour un système linéaire autonome avec contrainte de cible ponctuelle $M = \{x_1\}$ (cf. la section 3.3). On a donc

$$f(t, x, u) = Ax + Bu, \quad g(t, x, u) = 1, \quad h(x) = 0.$$

L'état adjoint satisfait $\frac{d\overline{p}}{dt}(t) = -A^\dagger \overline{p}(t)$ pour tout $t \in [0,T]$, et la condition de transversalité sur l'état adjoint est triviale car elle s'écrit $\overline{p}(T) \perp 0$. Il n'y a donc pas de condition finale sur l'état adjoint. Le Hamiltonien vaut

$$H(x, p, \lambda, u) = p^\dagger(Ax + Bu) + \lambda,$$

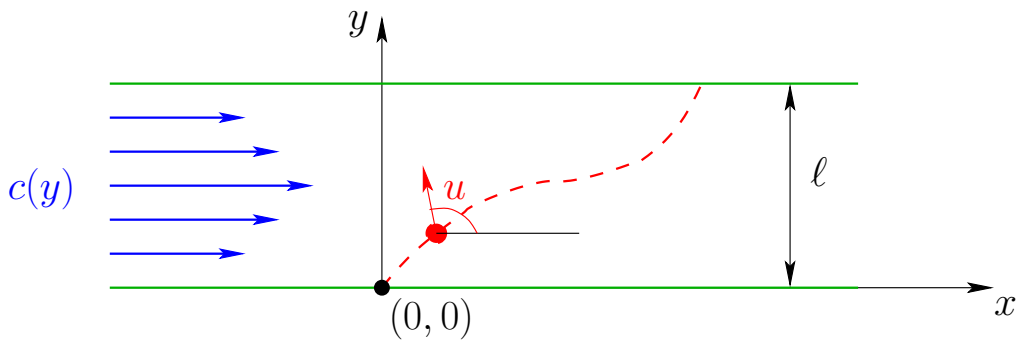

FIGURE 6.3 – Illustration du problème de Zermelo : barque traversant un canal.

et on constate qu'il est autonome ($\frac{\partial H}{\partial t} = 0$). La condition de minimisation donne

$$\overline{u}(t) = \arg\min_{v \in U} \overline{p}(t)^\dagger B v.$$

On retrouve bien le résultat du théorème 3.17. □

6.3 Application : problème de Zermelo

On considère une barque traversant un canal de largeur ℓ. On considère un repère cartésien où l'axe Ox coïncide avec la berge de départ et l'axe Oy est transverse au canal. La barque a une vitesse d'amplitude constante notée v, et le courant a une vitesse $c(y)$. On suppose que $c(y) > v$, pour tout $y \in [0, \ell]$; il s'agit d'une hypothèse dite de courant fort car l'amplitude du courant est toujours supérieure à la vitesse de la barque. La configuration est illustrée à la figure 6.3. Le contrôle est l'angle u de la vitesse de la barque par rapport à l'axe Ox, la vitesse étant considérée dans le repère du courant. L'état de la barque est décrit par le couple $X = (x, y)^\dagger \in \mathbb{R}^2$ donnant les coordonnées de la barque dans le repère Oxy. La trajectoire de la barque est régie par la dynamique suivante :

$$\dot{X}(t) = f(X(t), u(t)) = \begin{pmatrix} v\cos(u(t)) + c(y(t)) \\ v\sin(u(t)) \end{pmatrix}, \quad \forall t \in [0,T], \tag{6.16}$$

et la condition initiale est $X(0) = (0, 0)^\dagger$. Nous allons (brièvement) considérer trois problèmes de contrôle optimal pour atteindre la berge opposée :

1. minimiser le déport latéral ;
2. minimiser le temps de traversée ;
3. atteindre un point de la berge opposée en temps minimal.

Minimisation du déport latéral

Dans le problème de minimisation du déport latéral, le critère fait intervenir les fonctions

$$g(t, X, u) = 0, \quad h(X) = x, \tag{6.17}$$

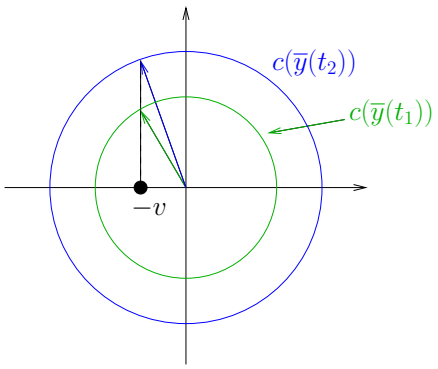

FIGURE 6.4 – Contrôle optimal pour la minimisation du déport latéral.

et on rajoute la contrainte de cible $y(T) = \ell$. Le temps final est libre. Commençons par considérer l'état adjoint. Comme $\overline{A}(t) = \left(\begin{smallmatrix} 0 & c'(\overline{y}(t)) \\ 0 & 0 \end{smallmatrix}\right)$ et $\overline{b}(t) = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, l'état adjoint $\overline{p} = (\overline{p}_x, \overline{p}_y)^\dagger$ satisfait

$$\overline{p}_x = cste, \quad \frac{d\overline{p}_y}{dt}(t) = -c'(\overline{y}(t))\overline{p}_x, \tag{6.18}$$

et la condition de transversalité sur l'état adjoint s'écrit

$$\begin{pmatrix} \overline{p}_x - \lambda \\ \overline{p}_y(T) \end{pmatrix} \perp \begin{pmatrix} 1 \\ 0 \end{pmatrix} \implies \overline{p}_x = \lambda. \tag{6.19}$$

Par ailleurs, le Hamiltonien vaut

$$H(X, p, \lambda, u) = (p_x \cos(u) + p_y \sin(u))v + p_x c(y). \tag{6.20}$$

Si $|\overline{p}|_{\mathbb{R}^2} \neq 0$, la condition de minimisation sur le Hamiltonien nous donne le contrôle optimal sous la forme

$$\cos(\overline{u}) = -\frac{\overline{p}_x}{|\overline{p}|}, \quad \sin(\overline{u}) = -\frac{\overline{p}_y}{|\overline{p}|}. \tag{6.21}$$

Le Hamiltonien minimisé vaut $\mathcal{H}(t) = H(\overline{X}(t), \overline{p}(t), \lambda, \overline{u}(t)) = -|\overline{p}(t)|v + \overline{p}_x c(\overline{y}(t))$ et la condition de transversalité sur le Hamiltonien au temps final donne

$$-|\overline{p}(T)|v + \overline{p}_x c(\overline{y}(T)) = 0. \tag{6.22}$$

Montrons qu'il n'y a pas d'extrémale anormale. On raisonne par l'absurde en supposant que $\lambda = 0$. Dans ce cas, on a $\overline{p}_x = \lambda = 0$, ce qui implique que $\overline{p}_y$ est constant. En utilisant la condition de transversalité sur H, on voit que $\overline{p}_y$ s'annule au temps final (sinon, le Hamiltonien minimisé au temps final vaudrait $-|\overline{p}_y(T)|v = 0$, ce qui serait une contradiction). Par conséquent, $\overline{p}_y$ serait également nul à tout temps, ce qui est exclu car $(\overline{p}, \lambda) \neq ((0,0)^\dagger, 0)$. L'extrémale étant normale, nous pouvons supposer que $\overline{p}_x = \lambda = 1$. Par conséquent, on a toujours $\overline{p} \neq (0,0)^\dagger$, si bien que le contrôle optimal est bien donné par l'équation (6.18).

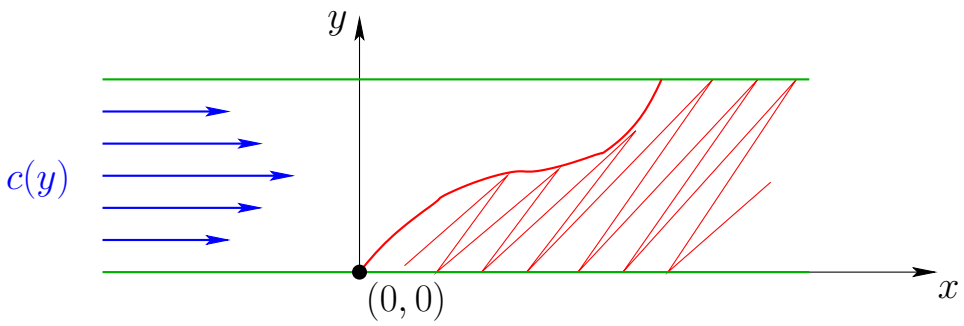

FIGURE 6.5 – Illustration de l'ensemble atteignable pour le problème de Zermelo sous hypothèse de courant fort.

Montrons maintenant que $|\overline{p}(t)|_{\mathbb{R}^2} = \frac{c(\overline{y}(t))}{v}$, pour tout $t \in [0,T]$. En effet, cette relation est satisfaite en T de par la condition de transversalité sur H. En dérivant par rapport au temps, et en utilisant le fait que $|\overline{p}|_{\mathbb{R}^2} = (1+\overline{p}_y^2)^{1/2}$, il vient

$$\frac{d|\overline{p}|}{dt} = \frac{\overline{p}_y}{|\overline{p}|}\frac{d\overline{p}_y}{dt} = \sin(\overline{u})c'(\overline{y}) = \frac{1}{v}\frac{d\overline{y}}{dt}c'(\overline{y}) = \frac{1}{v}\frac{d\overline{c}(\overline{y})}{dt}, \tag{6.23}$$

ce qui montre que

$$\frac{d}{dt}\left(|\overline{p}(t)|_{\mathbb{R}^2} - \frac{c(\overline{y}(t))}{v}\right) = 0. \tag{6.24}$$

On peut aussi montrer cette propriété en utilisant le fait que le Hamiltonien minimisé est constant en temps (cf. la remarque 6.6). En conclusion, le contrôle optimal s'écrit comme un feedback sous la forme

$$\cos(\overline{u}(t)) = -\frac{v}{c(\overline{y}(t))}, \quad \forall t \in [0,T]. \tag{6.25}$$

Une illustration montrant comment déterminer le contrôle optimal est présentée à la figure 6.4. Un calcul simple montre que $\frac{d}{dt}\overline{X}(t) = \frac{\sqrt{\overline{c}^2-v^2}}{\overline{c}}(\sqrt{\overline{c}^2-v^2}, v)^\dagger$, où $\overline{c} := c(\overline{y}(t))$, si bien que l'angle de la vitesse de la barque dans le repère cartésien avec l'axe Ox est $\overline{u}(t) - \frac{\pi}{2}$.

Remarque 6.8. [Ensemble atteignable] La résolution du problème de déport latéral permet de déterminer l'ensemble atteignable par tout contrôle. Celui-ci est illustré de manière schématique à la figure 6.5. □

Minimisation du temps de traversée

Le critère fait cette fois intervenir les fonctions $g(t,X,u) = 1$ et $h(X) = 0$. On a toujours la contrainte de cible $y(T) = \ell$ et le temps final reste libre. Les équations de l'état adjoint sont inchangées :

$$\overline{p}_x = cste, \quad \frac{d\overline{p}_y}{dt} = -c'(\overline{y})\overline{p}_x, \tag{6.26}$$

mais la condition de transversalité sur l'état adjoint est maintenant

$$\overline{p}_x(T) = 0 \implies \overline{p}_x \equiv 0, \tag{6.27}$$

ce qui implique que

$$\overline{p}_y = cste. \tag{6.28}$$

Le Hamiltonien à minimiser le long de l'extrémale vaut

$$\begin{aligned} H(\overline{X}, \overline{p}, \lambda, u) &= (\overline{p}_x \cos(u) + \overline{p}_y \sin(u))v + \overline{p}_x c(\overline{y}) + \lambda \\ &= \overline{p}_y \sin(u)v + \lambda, \end{aligned} \tag{6.29}$$

car $\overline{p}_x = 0$. On a nécessairement $\overline{p}_y \neq 0$, car sinon la condition de transversalité sur le Hamiltonien donnerait $\lambda = 0$, ce qui est exclu. Le contrôle optimal est donc $\sin(\overline{u}(t)) = 1$, i.e.,

$$\overline{u}(t) = \frac{\pi}{2}. \tag{6.30}$$

On retrouve un conseil (relativement) bien connu : "ne jamais naviguer contre le courant si on veut atteindre la rive opposée le plus vite possible."

Atteindre une cible en temps minimal

On suppose que la cible a pour coordonnées $(x_1, \ell)^\dagger$ où x_1 est situé en aval du point de déport minimal. On a $g = 1$ et $h = 0$ comme dans le cas précédent. Les équations d'évolution de l'état adjoint restent inchangées, mais la condition de transversalité sur l'état adjoint devient triviale. Le Hamiltonien est $H(X, p, \lambda, u) = (p_x \cos(u) + p_y \sin(u))v + p_x c(y) + \lambda$, et le Hamiltonien minimisé est

$$\mathcal{H}(t) = -|\overline{p}(t)|v + \overline{p}_x c(\overline{y}(t)) + \lambda \equiv 0, \quad \forall t \in [0,T]. \tag{6.31}$$

Deux situations peuvent se produire :

1. **Extrêmale anormale** ($\lambda = 0$) : on obtient à nouveau $\cos(\overline{u}(t)) = -\frac{v}{c(\overline{y}(t))}$; cette situation se produit lorsque x_1 est l'abscisse du point de déport minimal.
2. **Extrêmale normale** ($\lambda = 1$) : en utilisant le Hamiltonien minimisé et $\cos(\overline{u}(t)) = -\frac{\overline{p}_x}{|\overline{p}(t)|}$, il vient

$$\cos(\overline{u}(t)) = \frac{\overline{p}_x v}{1 - \overline{p}_x c(\overline{y}(t))}, \tag{6.32}$$

pourvu que $\overline{p}_x \in]-\infty, \frac{1}{v+c_\sharp}[$ où $c_\sharp = \max_{y\in[0,\ell]} c(y)$. On obtient une famille de courbes à un paramètre ; lorsque $\overline{p}_x \to -\infty$, on tend vers l'extrémale anormale ; par ailleurs, la valeur $\overline{p}_x = 0$ correspond à la traversée en temps minimum.

6.4 Résolution numérique : méthode de tir

Le PMP peut servir de base à une méthode numérique de résolution du problème de contrôle optimal : la **méthode de tir**. Cette méthode est intéressante lorsqu'il est facile de minimiser le Hamiltonien, i.e., lorsqu'on est capable d'évaluer une fonction

$$\zeta(t,x,p) \in \arg\min_{v\in U} H(t,x,p,v), \quad \forall (t,x,p) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d. \tag{6.33}$$

Dans ce cas, en posant $z(t) = (x(t),p(t))^\dagger \in \mathbb{R}^d\times\mathbb{R}^d$, on obtient le système différentiel

$$\dot z(t) = F(t,z(t)), \quad \forall t\in[0,T], \tag{6.34}$$

avec $F = (F_x,F_p) : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d\times\mathbb{R}^d$ telle que

$$F_x(t,(x,p)) = f(t,x,\zeta(t,x,p)), \tag{6.35a}$$

$$F_p(t,(x,p)) = -\frac{\partial f}{\partial x}(t,x,\zeta(t,x,p))^\dagger p - \frac{\partial g}{\partial x}(t,x,\zeta(t,x,p)). \tag{6.35b}$$

Le principe de la méthode de tir est le suivant :

1. On se donne une condition initiale $p_0\in\mathbb{R}^d$ sur l'état adjoint ; en intégrant le système différentiel (6.34), on obtient $p(T)\in\mathbb{R}^d$; ceci nous permet de définir l'application

$$\mathcal{F} : \mathbb{R}^d \to \mathbb{R}^d, \quad \mathcal{F}(p_0) := p(T) - \frac{\partial h}{\partial x}(x(T)). \tag{6.36}$$

2. On cherche $p_0\in\mathbb{R}^d$ tel que $\mathcal{F}(p_0)=0$; il s'agit d'un système de d équations non-linéaires couplées dans $\mathbb{R}^d$, que l'on peut (tenter de) résoudre par la méthode de Newton.

Les deux avantages de la méthode de tir sont une très grande précision numérique (si la convergence est atteinte ...) et une efficacité même en grande dimension sur l'état (i.e., même si $d\gg 1$). En revanche, les difficultés rencontrées avec la méthode de tir sont un (très) petit domaine de convergence, la nécessité de vérifier après coup l'optimalité de la solution trouvée, et le fait que la structure des commutations doit être connue à l'avance (la méthode de Newton exploitant la régularité de $\mathcal{F}$).

Remarque 6.9. [Méthodes directes] On peut aussi chercher un contrôle optimal en utilisant une méthode directe qui n'invoque pas le PMP. Dans ce cas, on considère directement le problème de contrôle optimal. On se ramène en dimension finie en discrétisant en temps la dynamique (en utilisant un schéma d'Euler par exemple) et on considère par exemple des contrôles constants par morceaux. On obtient ainsi un problème d'optimisation non-linéaire sous contraintes en dimension finie, que l'on peut résoudre de plusieurs façons (par exemple, en utilisant les techniques de Sequential Quadratic Programming). Les méthodes directes sont de mise en œuvre simple ; en revanche, elles sont souvent peu précises et deviennent très chères si d est grand. Enfin, on observera qu'on peut utiliser une méthode directe pour initialiser une méthode de tir. □

Chapitre 7

Programmation dynamique en temps discret

Ce chapitre introduit un nouveau point de vue pour la résolution des problèmes de contrôle optimal. Pour simplifier, nous considérons des problèmes de contrôle optimal en **temps discret** ; nous reviendrons aux problèmes en temps continu au chapitre suivant. L'idée clé de ce chapitre est de plonger le problème de contrôle optimal dans une **famille** de problèmes de contrôle optimal paramétrés par une paire (x, t_n) représentant un état générique du système $x \in \mathbb{R}^d$ à un instant discret $t_n \in \{t_0 = 0, \ldots, t_N = T\}$. On cherche alors pour chaque paire (x, t_n), la **fonction valeur** définie comme la valeur minimale du critère pour le problème de contrôle optimal défini à partir de t_n en prenant x comme état du système. C'est à première vue étonnant puisqu'on a maintenant à résoudre une famille de problèmes de contrôle optimal au lieu d'un seul (même si souvent on considère de toutes façons plusieurs conditions initiales...). Le point remarquable est que la famille de fonctions valeur est solution d'une équation de récurrence rétrograde, appelée **équation de programmation dynamique**. De plus, la résolution rétrograde de cette équation, du temps final au temps initial, nous fournit à chaque instant discret un contrôle optimal comme un feedback sur l'état du système. Nous pouvons alors revenir au problème de contrôle optimal de départ qui était posé avec une certaine condition initiale $x_0 \in \mathbb{R}^d$ au temps $t_0 = 0$, et en avançant du temps initial au temps final, nous pouvons déterminer les contrôles optimaux à tous les instants discrets ainsi que la trajectoire optimale associée. L'idée à la base de la programmation dynamique est plus générale que le cadre des problèmes de contrôle optimal. Nous en donnerons un exemple d'application en **optimisation combinatoire**.

7.1 Contrôle optimal en temps discret

On considère des instants discrets

$$0 = t_0 < t_1 < \cdots < t_N = T \quad (N > 0). \tag{7.1}$$

Plutôt que de décrire l'état du système et le contrôle par des fonctions $x : [0,T] \to \mathbb{R}^d$ et $u : [0,T] \to U \subset \mathbb{R}^k$ (où U est sous-ensemble fermé non-vide de $\mathbb{R}^k$), on les décrit ici par des

familles discrètes

$$(x_m)_{m\in\{0:N\}} \in \mathbb{R}^{d(N+1)}, \quad (u_m)_{m\in\{0:N-1\}} \in U^N. \tag{7.2}$$

La dynamique discrète d'évolution s'écrit sous la forme suivante :

$$x_{m+1} = F_m(x_m, u_m), \quad \forall m \in \{0{:}N{-}1\}, \qquad x_0 = x, \tag{7.3}$$

où $F_m : \mathbb{R}^d \times U \to \mathbb{R}^d$, pour tout $m \in \{0{:}N{-}1\}$, et $x \in \mathbb{R}^d$ est la condition initiale. Étant donnée la condition initiale x, la connaissance des contrôles $(u_m)_{m\in\{0:N-1\}}$ détermine par récurrence les états successifs $(x_m)_{m\in\{1:N\}}$ en utilisant (7.3). Au premier abord, il peut paraître surprenant de noter x la condition initiale ; c'est en fait tout à fait naturel dans le contexte de la programmation dynamique où une des idées clés est de considérer une famille de problèmes de contrôle optimal paramétrés par la condition initiale.

Afin de définir un critère d'optimalité en temps discret, on se donne une famille de fonctions $(G_m)_{m\in\{0:N-1\}}$ avec $G_m : \mathbb{R}^d \times U \to \mathbb{R}$, et une fonction $h : \mathbb{R}^d \to \mathbb{R}$. Le critère à minimiser $J_0 : \mathbb{R}^d \times U^N \to \mathbb{R}$ est tel que

$$J_0(x; u_0, \cdots, u_{N-1}) = \sum_{m\in\{0:N-1\}} G_m(x_m, u_m) + h(x_N). \tag{7.4}$$

On notera que l'on a explicité la dépendance du critère en la condition initiale $x \in \mathbb{R}^d$; de plus, l'indice 0 fait référence au fait que la condition initiale est prescrite en t_0. Le problème de contrôle optimal en temps discret est le suivant :

Chercher $(\overline{u}_0, \cdots, \overline{u}_{N-1}) \in U^N$ tel que

$$J_0(x; \overline{u}_0, \cdots, \overline{u}_{N-1}) = \min_{(u_0,\cdots,u_{N-1})\in U^N} J_0(x; u_0, \cdots, u_{N-1}). \tag{7.5}$$

Remarque 7.1. [Lien avec le contrôle optimal en temps continu] Pour le problème de contrôle optimal en temps continu, on rappelle que l'on a

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad J(u) = \int_0^T g(t, x_u(t), u(t))\,\mathrm{d}t + h(x_u(T)).$$

Le lien avec l'approche en temps discret se fait en considérant les approximations temporelles $x_m \approx x_u(t_m)$, $u_m \approx u(t_m)$, le schéma d'Euler explicite pour la dynamique (avec un pas de temps $\Delta t_m = t_{m+1} - t_m$) sous la forme

$$x_u(t_{m+1}) = x_u(t_m) + \int_{t_m}^{t_{m+1}} f(t, x_u(t), u(t))\,\mathrm{d}t = x_u(t_m) + \Delta t_m f(t_m, x_u(t_m), u(t_m)) + o(\Delta t),$$

ce qui conduit à la dynamique discrète (7.3) avec

$$F_m(y, v) = y + \Delta t_m f(t_m, y, v), \quad \forall (y, v) \in \mathbb{R}^d \times U,$$

et enfin une formule des rectangles pour évaluer le critère, i.e.,

$$J(u) = \sum_{m\in\{0:N-1\}} \Delta t_m g(t_m, x_u(t_m), u(t_m)) + o(\Delta t) + h(x_u(t_N)),$$

ce qui conduit à

$$G_m(y, v) = \Delta t_m g(t_m, y, v), \quad \forall (y, v) \in \mathbb{R}^d \times U,$$

tandis que la fonction h est la même que dans le cas continu. □

7.2 Fonction valeur et programmation dynamique

La fonction valeur pour le problème (7.5) est la fonction $V_0 : \mathbb{R}^d \to \mathbb{R}$ telle que

$$V_0(x) = \inf_{(u_0,\cdots,u_{N-1})\in U^N} J_0(x; u_0, \cdots, u_{N-1}). \tag{7.6}$$

L'application V_0 associe donc à la condition initiale $x \in \mathbb{R}^d$ la valeur optimale du critère.

Nous allons maintenant plonger le problème (7.5) dans une famille de problèmes de contrôle optimal en temps discret paramétrée par $n \in \{0{:}N{-}1\}$ et $x \in \mathbb{R}^d$. L'idée est que la dynamique démarre à l'instant discret t_n avec l'état x ; on a donc

$$x_{m+1} = F_m(x_m, u_m), \quad \forall m \in \{n{:}N{-}1\}, \qquad x_n = x, \tag{7.7}$$

et le critère est $J_n : \mathbb{R}^d \times U^{N-n} \to \mathbb{R}$ tel que

$$J_n(x; u_n, \cdots, u_{N-1}) = \sum_{m\in\{n:N-1\}} G_m(x_m, u_m) + h(x_N). \tag{7.8}$$

Définition 7.2 (Fonction valeur)**.** *La fonction valeur pour la dynamique discrète* (7.7) *et le critère* (7.8) *est l'application $V_n : \mathbb{R}^d \to \mathbb{R}$ telle que, pour tout $n \in \{0{:}N{-}1\}$,*

$$V_n(x) = \inf_{(u_n,\cdots,u_{N-1})\in U^{N-n}} J_n(x; u_n, \cdots, u_{N-1}). \tag{7.9}$$

Enfin, pour $n = N$, on définit $V_N : \mathbb{R}^d \to \mathbb{R}$ par $V_N(x) = h(x)$ pour tout $x \in \mathbb{R}^d$.

Le plongement opéré ci-dessus nous conduit à chercher la famille de fonctions valeur $(V_n)_{n\in\{0:N\}}$ avec $V_n : \mathbb{R}^d \to \mathbb{R}$, pour tout $n \in \{0{:}N\}$ et la condition finale (en $n = N$) $V_N(x) = h(x)$ pour tout $x \in \mathbb{R}^d$. L'idée clé est qu'il est possible d'obtenir toutes ces fonctions par la résolution d'une équation fonctionnelle rétrograde en temps discret.

Proposition 7.3 (Programmation dynamique en temps discret)**.** *Les fonctions valeur définies par* (7.9) *satisfont l'équation fonctionnelle*

$$V_n(x) = \inf_{v\in U} \big\{G_n(x, v) + V_{n+1}(F_n(x, v))\big\}, \quad \forall x \in \mathbb{R}^d,\ \forall n \in \{0{:}N{-}1\}, \tag{7.10}$$

qui se résout par récurrence rétrograde en n à partir de la condition finale $V_N(x) = h(x)$ pour tout $x \in \mathbb{R}^d$.

Remarque 7.4. [En pratique] Les problèmes de minimisation (7.10) sont bien posés si, par exemple, U est un ensemble compact, les fonctions F_n et G_n sont continues sur $\mathbb{R}^d \times U$ et la fonction h est continue sur $\mathbb{R}^d$. Les fonctions valeur V_n sont alors continues sur $\mathbb{R}^d$. La résolution des problèmes de minimisation (7.10) fournit de manière rétrograde les fonctions $\tilde{u}_{N-1}(x), \cdots, \tilde{u}_0(x)$ comme des feedbacks sur l'état x, i.e., comme des fonctions $\tilde{u}_n : \mathbb{R}^d \to \mathbb{R}^k$ telles que

$$\tilde{u}_n(x) \in \arg\min_{v \in U} \big\{ G_n(x,v) + V_{n+1}(F_n(x,v)) \big\}.$$

Une fois qu'on a déterminé tous ces feedbacks en remontant jusqu'à la fonction valeur V_0, on est en mesure de déterminer la trajectoire optimale et les contrôles optimaux en repartant de $n = 0$: il suffit en effet de poser $\overline{u}_0 = \tilde{u}_0(x_0)$, puis $\overline{x}_1 = F_0(x_0, \overline{u}_0)$ et $\overline{u}_1 = \tilde{u}_1(\overline{x}_1)$, puis $\overline{x}_2 = F_1(\overline{x}_1, \overline{u}_1)$ et $\overline{u}_2 = \tilde{u}_2(\overline{x}_2)$ et ainsi de suite. □

Démonstration. Pour $n = N-1$, comme $x_{N-1} = x$ et $x_N = F_{N-1}(x, u_{N-1})$, il vient

$$\begin{aligned} V_{N-1}(x) &= \inf_{u_{N-1} \in U} \Big\{ G_{N-1}(x, u_{N-1}) + h(x_N) \Big\} \\ &= \inf_{u_{N-1} \in U} \Big\{ G_{N-1}(x, u_{N-1}) + V_N(x_N) \Big\} \\ &= \inf_{u_{N-1} \in U} \Big\{ G_{N-1}(x, u_{N-1}) + V_N(F_{N-1}(x, u_{N-1}) \Big\} \\ &= \inf_{v \in U} \Big\{ G_{N-1}(x, v) + V_N(F_{N-1}(x, v)) \Big\}. \end{aligned}$$

Pour $n < N-1$, on a

$$\begin{aligned} V_n(x) &= \inf_{(u_n, \cdots, u_{N-1}) \in U^{N-n}} J_n(x; u_n, \cdots, u_{N-1}) \\ &= \inf_{u_n \in U} \inf_{(u_{n+1}, \cdots, u_{N-1}) \in U^{N-n-1}} \Big\{ \sum_{m \in \{n:N-1\}} G_m(x_m, u_m) + h(x_N) \Big\} \\ &= \inf_{u_n \in U} \Big\{ G_n(x, u_n) + \inf_{(u_{n+1}, \cdots, u_{N-1}) \in U^{N-n-1}} \Big\{ \sum_{m \in \{n+1:N-1\}} G_m(x_m, u_m) + h(x_N) \Big\} \Big\} \\ &= \inf_{u_n \in U} \Big\{ G_n(x, u_n) + V_{n+1}(x_{n+1}) \Big\} \\ &= \inf_{u_n \in U} \Big\{ G_n(x, u_n) + V_{n+1}(F_n(x, u_n)) \Big\} \\ &= \inf_{u \in U} \Big\{ G_n(x, u) + V_{n+1}(F_n(x, u)) \Big\}, \end{aligned}$$

ce qui complète la preuve. On notera que l'argument essentiel est que le critère est **additif** le long des trajectoires. □

7.3 Application : système LQ en temps discret

On considère deux matrices $\mathcal{A} \in \mathbb{R}^{d \times d}$ et $\mathcal{B} \in \mathbb{R}^{d \times k}$. La dynamique discrète s'écrit sous la forme

$$x_{m+1} = \mathcal{A} x_m + \mathcal{B} u_m, \quad \forall m \in \{0{:}N{-}1\}, \qquad x_0 = x \in \mathbb{R}^d, \tag{7.11}$$

qui est bien de la forme (7.3) avec $F(y,v) = \mathcal{A}y + \mathcal{B}v$. Le critère à minimiser s'écrit sous la forme

$$J_0(x;u_0,\cdots,u_{N-1}) = \sum_{m\in\{0:N-1\}} \Big\{\frac{1}{2}u_m^\dagger \mathcal{R} u_m + \frac{1}{2}x_m^\dagger \mathcal{Q} x_m\Big\} + \frac{1}{2}x_N^\dagger D x_N, \tag{7.12}$$

où la matrice $\mathcal{R} \in \mathbb{R}^{k\times k}$ est symétrique définie positive et les matrices $\mathcal{Q}, D \in \mathbb{R}^{d\times d}$ sont symétriques semi-définies positives. Le critère J_0 est bien de la forme (7.4) avec $G(y,v) = \frac{1}{2}v^\dagger\mathcal{R}v + \frac{1}{2}y^\dagger\mathcal{Q}y$ et $h(y) = \frac{1}{2}y^\dagger D y$. Pour simplifier, on considère un état cible nul et il n'y a pas de contraintes sur le contrôle si bien que $U = \mathbb{R}^k$.

En appliquant la proposition 7.3, l'équation de programmation dynamique s'écrit

$$V_n(x) = \inf_{u\in\mathbb{R}^k}\Big\{\frac{1}{2}u^\dagger\mathcal{R}u + \frac{1}{2}x^\dagger\mathcal{Q}x + V_{n+1}(\mathcal{A}x+\mathcal{B}u)\Big\}, \tag{7.13}$$

avec la condition finale $V_N(x) = \frac{1}{2}x^\dagger D x$ pour tout $x \in \mathbb{R}^d$.

Lemme 7.5 (Résolution de l'équation de programmation dynamique)**.** *Les fonctions valeur $(V_n)_{n\in\{0:N\}}$ de l'équation de programmation dynamique* (7.13) *sont telles que*

$$V_n(x) = \frac{1}{2}x^\dagger P_n x, \qquad \forall x \in \mathbb{R}^d, \tag{7.14}$$

où les matrices $(P_n)_{n\in\{0:N\}}$ sont symétriques semi-définies positives et données par la formule de récurrence rétrograde suivante :

$$P_N = D, \qquad P_n = \mathcal{A}^\dagger P_{n+1}\mathcal{A} - \mathcal{E}_{n+1}^\dagger(\mathcal{F}_{n+1})^{-1}\mathcal{E}_{n+1} + \mathcal{Q},\ \forall n \in \{0{:}N{-}1\}, \tag{7.15}$$

avec $\mathcal{E}_{n+1} = \mathcal{B}^\dagger P_{n+1}\mathcal{A}$ et $\mathcal{F}_{n+1} = \mathcal{R} + \mathcal{B}^\dagger P_{n+1}\mathcal{B}$.

Démonstration. La preuve se fait en raisonnant par récurrence rétrograde. Pour trouver le feedback $\tilde{u}_n(x)$ associé à la fonction valeur V_n, on considère l'application $u \mapsto u^\dagger\mathcal{R}u + (\mathcal{A}x + \mathcal{B}u)^\dagger\mathcal{P}_{n+1}(\mathcal{A}x+\mathcal{B}u)$. En utilisant l'hypothèse de récurrence sur la symétrie de la matrice P_{n+1} et en réarrangeant les termes, on se ramène à l'application $u \mapsto u^\dagger\mathcal{F}_{n+1}u + 2u^\dagger\mathcal{E}_{n+1}x$. Or, la matrice $\mathcal{F}_{n+1}$ est définie positive car $\mathcal{R}$ l'est et la matrice $\mathcal{B}^\dagger P_{n+1}\mathcal{B}$ est semi-définie positive. Par minimisation quadratique dans $\mathbb{R}^k$, on obtient qu'à l'étape n, le feedback est

$$\tilde{u}_n(x) = -(\mathcal{F}_{n+1})^{-1}\mathcal{E}_{n+1}x, \quad \forall x \in \mathbb{R}^d.$$

En reportant dans la définition de V_n, on obtient

$$\begin{aligned}
V_n(x) &= \frac{1}{2}x^\dagger(\mathcal{A}^\dagger P_{n+1}\mathcal{A} + \mathcal{Q})x + \inf_{u\in\mathbb{R}^k}\Big\{\frac{1}{2}u^\dagger\mathcal{F}_{n+1}u + u^\dagger\mathcal{E}_{n+1}x\Big\}\\
&= \frac{1}{2}x^\dagger(\mathcal{A}^\dagger P_{n+1}\mathcal{A} + \mathcal{Q})x + \frac{1}{2}\tilde{u}_n(x)^\dagger\mathcal{F}_{n+1}\tilde{u}_n(x) + \tilde{u}_n(x)^\dagger\mathcal{E}_{n+1}x\\
&= \frac{1}{2}x^\dagger(\mathcal{A}^\dagger P_{n+1}\mathcal{A} + \mathcal{Q})x - \frac{1}{2}\tilde{u}_n(x)^\dagger\mathcal{F}_{n+1}\tilde{u}_n(x)\\
&= \frac{1}{2}x^\dagger(\mathcal{A}^\dagger P_{n+1}\mathcal{A} + \mathcal{Q} - \mathcal{E}_{n+1}^\dagger(\mathcal{F}_{n+1})^{-1}\mathcal{E}_{n+1})x,
\end{aligned}$$

d'où l'expression de P_n en ré-arrangeant les termes. Cette expression montre que la matrice P_n est symétrique. De plus, la fonction valeur V_n vérifie $V_n(x) \geq 0$ pour tout $x \in \mathbb{R}^d$ car $J_n(x; u_n, \cdots, u_N) \geq 0$. Ceci montre que la matrice P_n est bien semi-définie positive. □

Il est intéressant de faire le lien avec le système LQ en temps continu en s'inspirant de la remarque 7.1. On rappelle que le système LQ en temps continu est régi par la dynamique

$$\dot{x}(t) = Ax(t) + Bu(t),\ \forall t \in [0,T], \quad x(0) = x, \tag{7.16}$$

et que le critère à minimiser est, en prenant un état cible nul,

$$J(u) = \int_0^T \left\{ \frac{1}{2} u(t)^\dagger R u(t) + \frac{1}{2} x(t)^\dagger Q x(t) \right\} \mathrm{d}t + \frac{1}{2} x(T)^\dagger D x(T). \tag{7.17}$$

On considère les instants discrets $0 = t_0 < t_1 < \cdots < t_N = T$ et on considère les approximations temporelles $x_m \approx x(t_m)$ et $u_m \approx u(t_m)$ pour tout $m \in \{0{:}N\}$. On suppose pour simplifier que le pas de temps Δt est constant. On considère un schéma d'Euler explicite pour la dynamique, ce qui donne

$$\mathcal{A} = I + \Delta t A, \quad \mathcal{B} = \Delta t B. \tag{7.18}$$

et la formule des rectangles pour évaluer le critère, ce qui donne

$$\mathcal{R} = \Delta t R, \quad \mathcal{Q} = \Delta t Q. \tag{7.19}$$

On rappelle que pour le système LQ en temps continu, l'état adjoint est donné par la formule $p(t) = P(t)x(t)$ où $P \in C^1([0,T]; \mathbb{R}^{d\times d})$ est solution de l'équation de Riccati (qui est rétrograde en temps)

$$\dot{P}(t) = -A^\dagger P(t) - P(t)A + P(t)BR^{-1}B^\dagger P(t) - Q,\ \forall t \in [0,T], \quad P(T) = D. \tag{7.20}$$

Soit maintenant une fonction $P_{\Delta t} \in C^1([0,T]; \mathbb{R}^{d\times d})$ telle que $P_{\Delta t}(t_n) = P_n$, pour tout $n \in \{0{:}N\}$. La forme précise de la dépendance temporelle de $P_{\Delta t}$ n'est pas importante tant que $P_{\Delta t}$ vérifie la propriété d'interpolation ci-dessus. En particulier, en $t_N = T$, on a $P_{\Delta t}(T) = P_N = D$. De plus, en effectuant des développements de Taylor en temps, il vient

$$\mathcal{A}^\dagger P_{n+1} \mathcal{A} = P_{\Delta t}(t_n) + \Delta t (A^\dagger P_{\Delta t}(t_n) + P_{\Delta t}(t_n) A + \dot{P}_{\Delta t}(t_n)) + o(\Delta t), \tag{7.21a}$$

$$\mathcal{E}_{n+1} = \mathcal{B}^\dagger P_{n+1} \mathcal{A} = \Delta t B^\dagger P_{\Delta t}(t_n) + o(\Delta t), \tag{7.21b}$$

$$\mathcal{F}_{n+1} = \mathcal{R} + \mathcal{B}^\dagger P_{n+1} \mathcal{B} = \Delta t R + o(\Delta t). \tag{7.21c}$$

Comme $P_n = \mathcal{A}^\dagger P_{n+1} \mathcal{A} - \mathcal{E}_{n+1}^\dagger (\mathcal{F}_{n+1})^{-1} \mathcal{E}_{n+1} + \mathcal{Q}$, on obtient, en simplifiant par $P_{\Delta t}(t_n)$ et en divisant par Δt, que

$$\dot{P}_{\Delta t}(t_n) = -A^\dagger P_{\Delta t}(t_n) - P_{\Delta t}(t_n) A + P_{\Delta t}(t_n) B R^{-1} B^\dagger P_{\Delta t}(t_n) - Q + o(1), \tag{7.22}$$

qui est, à $o(1)$ près, l'équation de Riccati pour le problème en temps continu à l'instant t_n.

Pour le système LQ en temps continu, le Hamiltonien est donné par

$$H(x,p,u) = p^\dagger(Ax+Bu) + \frac{1}{2}u^\dagger Ru + \frac{1}{2}x^\dagger Qx, \tag{7.23}$$

et le Hamiltonien minimisé par rapport à u est

$$H_\flat(x,p) = \min_{u\in\mathbb{R}^k} H(x,p,u). \tag{7.24}$$

Le Hamiltonien minimisé a bien un sens car on résout un problème de minimisation d'une fonctionnelle fortement convexe sur $\mathbb{R}^k$ (car la matrice R est symétrique définie positive). Dans le même esprit que ci-dessus, cherchons (formellement) une équation pour la fonction valeur du problème LQ en temps continu. Soit $V_{\Delta t} \in C^1([0,T]\times\mathbb{R}^d;\mathbb{R})$ une fonction de classe C^1 en (t,x) telle que

$$V_{\Delta t}(t_n,x) = V_n(x) = \frac{1}{2}x^\dagger P_n x, \quad \forall n\in\{0{:}N\},\ \forall x\in\mathbb{R}^d. \tag{7.25}$$

On rappelle que $V_n(x) = \inf_{u\in\mathbb{R}^k}\left\{\frac{1}{2}u^\dagger\mathcal{R}u + \frac{1}{2}x^\dagger\mathcal{Q}x + V_{n+1}(\mathcal{A}x+\mathcal{B}u)\right\}$ et que $\mathcal{A} = I+\Delta tA$, $\mathcal{B} = \Delta tB$, $\mathcal{R} = \Delta tR$, $\mathcal{Q} = \Delta tQ$. Un développement de Taylor nous montre alors que

$$\begin{aligned} V_{n+1}(\mathcal{A}x+\mathcal{B}u) - V_n(x) &= V_{\Delta t}(t_n+\Delta t, \mathcal{A}x+\mathcal{B}u) - V_{\Delta t}(t_n,x) \\ &= V_{\Delta t}(t_n+\Delta t, x+\Delta t(Ax+Bu)) - V_{\Delta t}(t_n,x) \\ &= \Delta t\frac{\partial V_{\Delta t}}{\partial t}(t_n,x) + \Delta t\frac{\partial V_{\Delta t}}{\partial x}(t_n,x)^\dagger(Ax+Bu) + o(\Delta t). \end{aligned} \tag{7.26}$$

En reportant dans l'équation de programmation dynamique donnant $V_n(x)$ et en réarrangeant les différents termes, puis en divisant par Δt, il vient

$$\inf_{u\in\mathbb{R}^k}\left\{\frac{1}{2}u^\dagger\mathcal{R}u + \frac{1}{2}x^\dagger\mathcal{Q}x + \frac{\partial V_{\Delta t}}{\partial t}(t_n,x) + \frac{\partial V_{\Delta t}}{\partial x}(t_n,x)^\dagger(Ax+Bu)\right\} = o(1), \tag{7.27}$$

ou encore

$$\frac{\partial V_{\Delta t}}{\partial t}(t_n,x) + \inf_{u\in\mathbb{R}^k}\left\{\frac{1}{2}u^\dagger\mathcal{R}u + \frac{1}{2}x^\dagger\mathcal{Q}x + \frac{\partial V_{\Delta t}}{\partial x}(t_n,x)^\dagger(Ax+Bu)\right\} = o(1), \tag{7.28}$$

En introduisant le Hamiltonien, il vient

$$\frac{\partial V_{\Delta t}}{\partial t}(t_n,x) + \inf_{u\in\mathbb{R}^k} H\left(x, \frac{\partial V_{\Delta t}}{\partial x}(t_n,x), u\right) = o(1), \tag{7.29}$$

ce qui se récrit de manière plus compacte avec le Hamiltonien minimisé sous la forme

$$\frac{\partial V_{\Delta t}}{\partial t}(t_n,x) + H_\flat\left(x, \frac{\partial V_{\Delta t}}{\partial x}(t_n,x)\right) = o(1). \tag{7.30}$$

Nous verrons au chapitre 8 que la fonction valeur $V : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ pour le problème LQ en temps continu satisfait l'équation de Hamilton–Jacobi–Bellman (HJB)

$$\frac{\partial V}{\partial t}(t,x) + H_\flat\left(x, \frac{\partial V}{\partial x}(t,x)\right) = 0, \quad \forall t\in[0,T],\ \forall x\in\mathbb{R}^d, \tag{7.31}$$

avec la condition en temps final $V(T,x) = \frac{1}{2}x^\dagger Dx$.

7.4 Optimisation combinatoire

La programmation dynamique a été introduite dans les années 1950 par R. Bellman. Son champ d'applications est bien plus vaste que les problèmes de contrôle optimal en temps discret. Parmi les exemples, nous pouvons mentionner la recherche opérationnelle (problème du plus court chemin, affectation de ressources, et plus généralement, la théorie des graphes) ou l'analyse statistique (détection de ruptures, i.e., estimer les instants où un signal présente des changements dans la distribution). Cette liste n'est de loin pas exhaustive.

Heuristiquement, le principe d'optimalité de Bellman est le suivant :

— toute solution optimale résulte de sous-problèmes résolus localement de façon optimale (i.e., lorsqu'on parcourt une trajectoire optimale, à tout instant, un contrôle optimal pour le problème restant est celui associé à la trajectoire optimale) ;
— on obtient ainsi une solution optimale en combinant des solutions optimales d'une série de sous-problèmes.

Nous allons nous contenter de donner ici un exemple simple de problème d'optimisation combinatoire : le problème du sac à dos. On considère un sac à dos de capacité Q et N objets, énumérés de 1 à N, de valeur individuelle v_n et d'encombrement e_n, pour tout $n \in \{1{:}N\}$. L'objectif est de remplir le sac à dos avec des objets en maximisant la valeur $\sum_{n\in\{1:N\}} u_n v_n$ du sac à dos sous la contrainte de capacité $\sum_{n\in\{1:N\}} u_n e_n \leq Q$. Ici, $u_n = 1$ signifie que l'objet d'indice n est sélectionné pour rentrer dans le sac à dos, sinon on a $u_n = 0$. Le problème de maximisation est donc le suivant :

$$\max_{\substack{(u_1,\cdots,u_N)\in\{0,1\}^N \\ \sum_{n\in\{1:N\}} u_n e_n \leq Q}} \Big\{ \sum_{n\in\{1:N\}} u_n v_n \Big\}. \tag{7.32}$$

Ce problème pourrait se résoudre en considérant les 2^N possibilités d'affection des objets, mais ce n'est pas réaliste si $N \gg 1$. La programmation dynamique permet de résoudre ce problème bien plus efficacement sous l'hypothèse (raisonnable) que Q et les encombrements e_n sont des entiers.

L'état du sac à dos est décrit par un entier $q \in \{0{:}Q\}$ quantifiant sa capacité. On considère la fonction valeur $V_n : \{0{:}Q\} \to \mathbb{R}$, où $V_n(q)$ est la valeur optimale d'un sac à dos de capacité q pour les objets énumérés de n à N. Pour $n = N$ (seul l'objet d'indice N est considéré), on a

$$V_N(q) = \begin{cases} v_N, & \text{si } e_N \leq q, \\ 0, & \text{sinon,} \end{cases} \qquad \tilde{u}_N(q) = \begin{cases} 1, & \text{si } e_N \leq q, \\ 0, & \text{sinon.} \end{cases}$$

Noter que le contrôle optimal est obtenu comme un feedback par rapport à l'état q. On stocke la fonction valeur et le contrôle optimal dans des tableaux à $(Q+1)$ lignes et N colonnes. Afin d'illustrer notre propos, considérons un exemple numérique avec $Q = 6$, $N = 3$, les valeurs $(3, 3, 5)$ pour les trois objets, et les encombrements $(3, 3, 4)$. On obtient alors les tableaux

suivants :

$V_n(q)$	1	2	3
6			5
5			5
4			5
3			0
2			0
1			0
0			0

$\tilde{u}_n(q)$	1	2	3
6			1
5			1
4			1
3			0
2			0
1			0
0			0

L'équation de programmation dynamique s'écrit

$$V_n(q) = \max_{\substack{u\in\{0,1\}\\ q\geq ue_n}} \Big\{ uv_n + V_{n+1}(q - ue_n) \Big\}. \tag{7.33}$$

Si $e_n > q$, la seule possibilité est $u = 0$ si bien que $V_n(q) = V_{n+1}(q)$, alors que si $e_n \leq q$, il vient

$$V_n(q) = \max \Big(V_{n+1}(q), v_n + V_{n+1}(q - e_n) \Big). \tag{7.34}$$

Le contrôle optimal comme feedback est donc

— $\tilde{u}_n(q) = 0$ si $e_n > q$ ou si $e_n \leq q$ et $V_{n+1}(q) > V_{n+1}(q - e_n) + v_n$;
— $\tilde{u}_n(q) = 1$ si $e_n \leq q$ et $V_{n+1}(q) < V_{n+1}(q - e_n) + v_n$;
— $\tilde{u}_n(q) \in \{0, 1\}$ si $e_n \leq q$ et $V_{n+1}(q) = V_{n+1}(q - e_n) + v_n$.

On continue à remplir les tableaux donnant $V_n(q)$ et $\tilde{u}_n(q)$ de la droite vers la gauche, ce qui pour le cas de notre exemple numérique conduit au résultat suivant :

$V_n(q)$	1	2	3
6	6	5	5
5	5	5	5
4	5	5	5
3	3	3	0
2	0	0	0
1	0	0	0
0	0	0	0

$\tilde{u}_n(q)$	1	2	3
6	1	0	1
5	0	0	1
4	0	0	1
3	{0,1}	1	0
2	0	0	0
1	0	0	0
0	0	0	0

Le problème de programmation dynamique est maintenant résolu : le contrôle optimal est $(\tilde{u}_1(6), \tilde{u}_2(3), \tilde{u}_3(0)) = (1, 1, 0)$, et la valeur optimale du sac à dos est de 6. Au total, on a effectué QN comparaisons (comparer à 2^N). On notera que l'algorithme glouton (qui consiste à choisir les objets par valeur décroissante) donne le contrôle $(0, 0, 1)$ et la valeur de 5 pour le sac à dos, ce qui est sous-optimal.

Chapitre 8

Équation de Hamilton–Jacobi–Bellman (HJB)

Ce chapitre est consacré à la programmation dynmaique en temps continu. En procédant de manière analogue au chapitre précédent, nous allons introduire une **fonction valeur** $V : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ et nous allons montrer, grâce au **principe d'optimalité de Bellman**, que cette fonction est solution d'une équation aux dérivées partielles en espace et en temps, appelée équation de **Hamilton–Jacobi–Bellman (HJB)**. Nous allons également montrer, sous certaines hypothèses, comment la fonction valeur nous permet de **synthétiser un contrôle optimal** sous forme de **feedback** et le lien qui peut être fait entre le gradient de la fonction valeur (par rapport à l'état) et l'état adjoint introduit dans le cadre du PMP. Nous conclurons en dressant un bref bilan comparatif des avantages et limites des deux approches considérées dans ce cours pour les problèmes de contrôle optimal : le PMP et l'équation HJB.

8.1 Fonction valeur

Soit $T > 0$, $x_0 \in \mathbb{R}^d$, et $f : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}^d$ où U est un sous-ensemble fermé non-vide de $\mathbb{R}^k$. On considère le système de contrôle non-linéaire

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in [0,T], \qquad x_u(0) = x_0, \tag{8.1}$$

où l'état du système est décrit par la fonction $x_u : [0,T] \to \mathbb{R}^d$ qui dépend du contrôle $u : [0,T] \to U$. L'ensemble des contrôles admissibles est le sous-ensemble

$$\mathcal{U} = L^1([0,T]; U) \subset L^1([0,T]; \mathbb{R}^k). \tag{8.2}$$

On considère une fonction $g : [0,T] \times \mathbb{R}^d \times U \to \mathbb{R}$ et une fonction $h : \mathbb{R}^d \to \mathbb{R}$, et on définit le critère suivant :

$$J(0, x_0; u) = \int_0^T g(t, x_u(t), u(t)) \, \mathrm{d}t + h(x_u(T)). \tag{8.3}$$

Comme dans le chapitre précédent, on explicite la dépendance du critère par rapport à la condition initiale $x_0 \in \mathbb{R}^d$, et le premier argument de J est là pour nous rappeler que cette

condition est imposée en $t = 0$. Le problème de contrôle optimal que l'on considère est le suivant :

$$\text{Chercher } \overline{u} \in \mathcal{U} \text{ tel que } J(0, x_0; \overline{u}) = \inf_{u \in \mathcal{U}} J(0, x_0; u). \tag{8.4}$$

Comme dans le chapitre précédent, on plonge le problème de minimisation (8.4) dans une famille de problèmes de contrôle optimal paramétrés par la paire (s, ξ) où $s \in [0,T]$ et $\xi \in \mathbb{R}^d$. Ces paramètres nous indiquent que le problème de contrôle optimal paramétré par (s, ξ) est posé sur l'intervalle $I_s = [s,T]$ avec la condition initiale $x(s) = \xi$. En résumé, on considère donc la famille de systèmes de contrôle non-linéaires et de critères

$$\dot{x}_u(t) = f(t, x_u(t), u(t)), \quad \forall t \in I_s, \qquad x_u(s) = \xi, \tag{8.5a}$$

$$J(s, \xi; u) = \int_{I_s} g(t, x_u(t), u(t))\, \mathrm{d}t + h(x_u(T)), \quad \forall u \in \mathcal{U}_s = L^1(I_s; U). \tag{8.5b}$$

Définition 8.1 (Fonction valeur)**.** *La fonction valeur $V : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ associée à la famille de systèmes de contrôle non-linéaires et de critères* (8.5) *est telle que*

$$V(s, \xi) = \inf_{u \in \mathcal{U}_s} J(s, \xi; u). \tag{8.6}$$

En $s = T$, on pose $V(T, \xi) = h(\xi)$ pour tout $\xi \in \mathbb{R}^d$.

Dans la suite de ce chapitre, nous faisons les hypothèses suivantes sur la dynamique :

a) la fonction f est continue sur $[0,T] \times \mathbb{R}^d \times U$, uniformément en $u \in U$;

b) la fonction f est dérivable par rapport à x et la fonction $\frac{\partial f}{\partial x}$ est continue bornée sur $[0,T] \times \mathbb{R}^d \times U$;

c) il existe une constante $C \geq 0$ telle que $|f(t, x, u)|_{\mathbb{R}^d} \leq C(1 + |x|_{\mathbb{R}^d} + |u|_{\mathbb{R}^k})$ sur $[0,T] \times \mathbb{R}^d \times U$.

et les hypothèses suivantes sur le critère :

d) la fonction g est continue sur $[0,T] \times \mathbb{R}^d \times U$, uniformément en $u \in U$;

e) il existe des constantes $\nu > 0$ et $C \geq 0$ telles que $g(t, x, u) \geq \nu |u|^2_{\mathbb{R}^k} - C$ sur $[0,T] \times \mathbb{R}^d \times U$;

f) la fonction h est continue et minorée sur $\mathbb{R}^d$.

On pourra vérifier que les hypothèses ci-dessus sont des conditions suffisantes d'une part pour qu'il existe une unique trajectoire $x_u \in AC([0,T]; \mathbb{R}^d)$ pour tout contrôle $u \in \mathcal{U} = L^1([0,T]; U)$, et d'autre part pour que le critère J ait bien un sens et $V(s, \xi) > -\infty$ pour tout $(s, \xi) \in [0,T] \times \mathbb{R}^d$.

Remarque 8.2. [Non-différentiabilité de la fonction valeur] On fera attention au fait que la fonction valeur n'est pas toujours différentiable en tout point. On considère par exemple le système de contrôle (linéaire) et le critère

$$\dot{x}_u(t) = u(t) \in U = [-1, 1],\ \forall t \in I_s, \quad x_u(s) = \xi, \quad J(s, \xi; u) = h(x_u(T)),$$

où la fonction $h : \mathbb{R} \to \mathbb{R}$ est supposée paire, régulière et telle que $h'(x) < 0$ si $x > 0$ (par exemple, on pourra considérer la fonction $h(x) = e^{-x^2}$ sur $\mathbb{R}$). Le problème de contrôle optimal ci-dessus se résout directement ; on constate en effet que

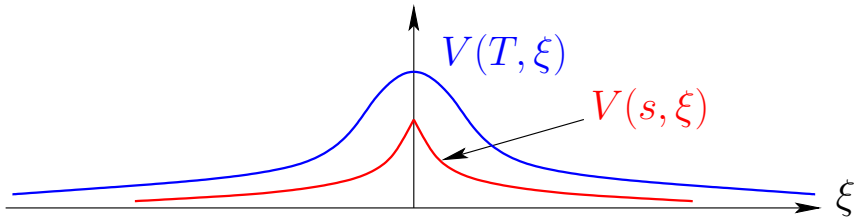

FIGURE 8.1 – Illustration de la remarque 8.2.

— Si $\xi > 0$, le contrôle optimal est $\overline{u} \equiv 1$ et $V(s,\xi) = h(\xi + T - s)$;
— si $\xi < 0$, le contrôle optimal est $\overline{u} \equiv -1$ et $V(s,\xi) = h(\xi - T + s)$;
— si $\xi = 0$, il y a deux contrôles optimaux $\overline{u} \equiv \pm 1$ et $V(s,0) = h(\pm(T-s))$.

En conclusion, on a $V(s,\xi) = h(T - s + |\xi|)$, qui est une fonction régulière sauf en $\xi = 0$ (cf. la figure 8.1). Cet exemple illustre un phénomène important et général : la perte de régularité de la fonction valeur en les points où existent plusieurs contrôles optimaux. □

Passons maintenant au principe d'optimalité de Bellman. Celui-ci constitue la clé de voûte de la programmation dynamique. Heuristiquement, le principe d'optimalité de Bellman nous dit que un contrôle $\overline{u}$ est optimal si à tout instant $s \in [0,T]$ sur la trajectoire associée $\overline{x} = x_{\overline{u}}$, le contrôle restreint aux instants ultérieurs $\overline{u}|_{[s,T]}$ est optimal pour le nouveau problème ayant l'état courant $\overline{x}(s)$ comme état initial.

Théorème 8.3 (Principe d'optimalité de Bellman)**.** *Soit $u \in \mathcal{U}$ un contrôle optimal. Soit $(s,\xi) \in [0,T[\times \mathbb{R}^d$. Alors, pour tout $s' \in [s,T]$, on a*

$$V(s,\xi) = \inf_{u \in \mathcal{U}_s^{s'}} \left\{ \int_{I_s^{s'}} g(t, x_u(t), u(t))\,\mathrm{d}t + V(s', x_u(s')) \right\}, \tag{8.7}$$

avec $I_s^{s'} = [s,s']$ et $\mathcal{U}_s^{s'} = L^1(I_s^{s'}; U)$.

Démonstration. Nous nous contenterons d'esquisser la preuve. On observe que toute fonction $u \in \mathcal{U}_s$ peut être identifiée à un couple de fonctions $(u_1, u_2) \in \mathcal{U}_s^{s'} \times \mathcal{U}_{s'}$ en posant $u_1 = u|_{I_s^{s'}}$ et $u_2 = u|_{I_{s'}}$. D'une part, le contrôle u_1 conduit à la trajectoire $x_1 \equiv x$ sur $I_s^{s'}$ telle que $x_1(s) = \xi$. D'autre part, le contrôle u_2 conduit à la trajectoire $x_2 \equiv x$ sur $I_{s'}$ telle que $x_2(s') = x_1(s')$. En utilisant l'additivité du critère le long de la trajectoire, on constate que

$$\begin{aligned}
V(s,\xi) &= \inf_{u \in \mathcal{U}_s} \left\{ \int_{I_s} g(t, x_u(t), u(t))\,\mathrm{d}t + h(x_u(T)) \right\} \\
&= \inf_{(u_1,u_2) \in \mathcal{U}_s^{s'} \times \mathcal{U}_{s'}} \left\{ \int_{I_s^{s'}} g(t, x_1(t), u_1(t))\,\mathrm{d}t + \left\{ \int_{I_{s'}} g(t, x_2(t), u_2(t))\,\mathrm{d}t + h(x_2(T)) \right\} \right\} \\
&= \inf_{u_1 \in \mathcal{U}_s^{s'}} \left\{ \int_{I_s^{s'}} g(t, x_1(t), u_1(t))\,\mathrm{d}t + \inf_{u_2 \in \mathcal{U}_{s'}} \left\{ \int_{I_{s'}} g(t, x_2(t), u_2(t))\,\mathrm{d}t + h(x_2(T)) \right\} \right\} \\
&= \inf_{u_1 \in \mathcal{U}_s^{s'}} \left\{ \int_{I_s^{s'}} g(t, x_1(t), u_1(t))\,\mathrm{d}t + V(s', x_1(s')) \right\},
\end{aligned}$$

ce qui conclut la preuve. □

On rappelle que le Hamiltonien associé au système de contrôle non-linéaire (8.1) et au critère (8.3) est l'application $H : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times U \to \mathbb{R}$ telle que

$$H(t,x,p,u) = p^\dagger f(t,x,u) + g(t,x,u). \tag{8.8}$$

En outre, on définit le Hamiltonien minimisé comme l'application $H_\flat : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ telle que

$$H_\flat(t,x,p) = \inf_{u \in U} H(t,x,p,u). \tag{8.9}$$

Théorème 8.4 (Équation HJB)**.** *En tout point $(s,\xi) \in [0,T[\times \mathbb{R}^d$ où la fonction valeur est différentiable, elle satisfait l'équation HJB*

$$\frac{\partial V}{\partial s}(s,\xi) + H_\flat\left(s, \xi, \frac{\partial V}{\partial \xi}(s,\xi)\right) = 0, \tag{8.10}$$

et elle vérifie en outre la condition à l'instant final

$$V(T,\xi) = h(\xi), \quad \forall \xi \in \mathbb{R}^d. \tag{8.11}$$

Exemple 8.5. Dans l'exemple considéré à la remarque 8.2, on a $H(x,p,u) = pu$ et $U = [-1,1]$, si bien que le Hamiltonien minimisé vaut $H_\flat(x,p) = -|p|$. On avait vu que la fonction valeur est égale à $V(s,\xi) = h(T - s - |\xi|)$. On vérifie facilement qu'en tout (s,ξ) où la fonction V est régulière (i.e., si $\xi \neq 0$), elle satisfait l'équation HJB $\frac{\partial V}{\partial s} - |\frac{\partial V}{\partial \xi}| = 0$. □

Démonstration. La preuve repose sur le principe d'optimalité de Bellman (cf. le théorème 8.3). Pour simplifier, on se restreint au cas où le sous-ensemble U est borné. Soit $(s,\xi) \in [0,T[\times \mathbb{R}^d$ un point où la fonction valeur est différentiable. On applique le principe d'optimalité de Bellman avec $s' = s + \delta$ et $0 < \delta < T - s$. On obtient ainsi

$$V(s,\xi) = \inf_{u \in \mathcal{U}_s^{s+\delta}} \left\{ \int_{I_s^{s+\delta}} g(t, x_u(t), u(t))\,\mathrm{d}t + V(s+\delta, x_u(s+\delta)) \right\}.$$

Puisque le sous-ensemble U est borné, les hypothèses sur f impliquent que

$$\sup_{t \in I_s^{s+\delta}} |x_u(t) - \xi|_{\mathbb{R}^d} \leq C\delta,$$

uniformément en δ et en u. On en déduit que

$$x_u(s+\delta) = \xi + \int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t + o(\delta),$$

uniformément en u. Comme la fonction V est supposée différentiable en (s,ξ), on obtient

$$V(s+\delta, x(s+\delta)) = V(s,\xi) + \delta\left\{ \frac{\partial V}{\partial s}(s,\xi) + \frac{\partial V}{\partial \xi}(s,\xi)^\dagger \left(\frac{1}{\delta} \int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t \right) \right\} + o(\delta).$$

De plus, les hypothèses sur g impliquent que

$$\int_{I_s^{s+\delta}} g(t, x_u(t), u(t))\,\mathrm{d}t = \int_{I_s^{s+\delta}} g(s, \xi, u(t))\,\mathrm{d}t + o(\delta).$$

En reportant dans le principe d'optimalité de Bellman, en simplifiant par $V(s,\xi)$ et en divisant par δ, et comme les termes en $o(\delta)$ sont uniformes en u, il vient

$$\frac{\partial V}{\partial s}(s,\xi) + \inf_{u\in\mathcal{U}_s^{s+\delta}} \left\{ \frac{1}{\delta}\int_{I_s^{s+\delta}} g(s,\xi,u(t))\,\mathrm{d}t + \frac{\partial V}{\partial \xi}(s,\xi)^\dagger \Big(\frac{1}{\delta}\int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t\Big)\right\} = o(1).$$

On conclut en remarquant que (s et ξ sont ici fixés)

$$\begin{aligned}\inf_{u\in\mathcal{U}_s^{s+\delta}} &\left\{ \frac{1}{\delta}\int_{I_s^{s+\delta}} g(s,\xi,u(t))\,\mathrm{d}t + \frac{\partial V}{\partial \xi}(s,\xi)^\dagger \Big(\frac{1}{\delta}\int_{I_s^{s+\delta}} f(s,\xi,u(t))\,\mathrm{d}t\Big)\right\} \\ &= \inf_{v\in U}\left\{ g(s,\xi,v) + \frac{\partial V}{\partial \xi}(s,\xi)^\dagger f(s,\xi,v)\right\} = \inf_{v\in U} H\bigg(s,\xi,\frac{\partial V}{\partial \xi}(s,\xi),v\bigg).\end{aligned}$$

Cela repose sur l'observation élémentaire que l'on a $\inf_{u\in\mathcal{U}_s^{s+\delta}} \frac{1}{\delta}\int_{I_s^{s+\delta}} \Phi(u(t))\,\mathrm{d}t = \inf_{v\in U}\Phi(v)$ où $\Phi : \mathbb{R}^k \to \mathbb{R}$. Pour montrer ce résultat, notons I_1 le premier infimum et I_2 le deuxième. Pour tout $u \in \mathcal{U}_s^{s+\delta}$, on a $\Phi(u(t)) \geq \inf_{v\in U}\Phi(v)$, pour tout $t \in I_s^{s+\delta}$, si bien que $I_1 \geq I_2$. De plus, pour tout $v \in U$, en considérant la fonction constante égale à v, on obtient

$$\Phi(v) \geq \inf_{u\in\mathcal{U}_s^{s+\delta}} \frac{1}{\delta}\int_{I_s^{s+\delta}} \Phi(u(t))\,\mathrm{d}t,$$

ce qui implique que $I_2 \geq I_1$ et complète la preuve. □

Remarque 8.6. [Unicité de la solution régulière] On peut montrer que si une fonction W suffisamment régulière (à savoir, $W \in C^0([0,T]\times\mathbb{R}^d)\cap C^1([0,T[\,\times\mathbb{R}^d)$) satisfait l'équation HJB (8.10) et la condition en temps final (8.11) et si le sous-ensemble U est borné, alors on a $W \equiv V$. En d'autres termes, l'équation HJB a au plus une solution régulière. Ce résultat d'unicité s'étend au cas où le sous-ensemble U est non-borné sous hypothèse de décroissance de W quand $|\xi| \to +\infty$ uniformément en t (voir par exemple la preuve du théorème 5.2 dans la référence [8]). Le lecteur désireux d'en savoir plus sur l'équation HJB pourra également consulter le livre [2]. □

Proposition 8.7 (Synthèse d'un feedback optimal)**.** *On suppose que la fonction valeur V est suffisamment régulière, i.e.,*

$$V \in C^0([0,T]\times\mathbb{R}^d)\cap C^1([0,T[\,\times\mathbb{R}^d). \tag{8.12}$$

On suppose que pour tout $(s,\xi)\in[0,T]\times\mathbb{R}^d$, on peut trouver un feedback optimal

$$\tilde{u}(s,\xi) \in \arg\min_{v\in U} H\bigg(s,\xi,\frac{\partial V}{\partial \xi}(s,\xi),v\bigg). \tag{8.13}$$

(L'existence d'un tel feedback optimal est assurée par les hypothèses sur f et g ; en général, on n'a pas unicité, ni dépendance continue en (s,ξ).) On suppose enfin que l'on peut choisir le feedback $\tilde{u}(s,\xi)$ de sorte à ce que le système différentiel

$$\frac{d\overline{x}}{dt}(t) = f(t,\overline{x}(t),\tilde{u}(t,\overline{x}(t))), \quad \forall t \in [0,T], \qquad \overline{x}(0) = x_0, \tag{8.14}$$

admette une solution $\overline{x} \in AC([0,T];\mathbb{R}^d)$. Dans ces conditions,

$$\overline{u}(t) = \tilde{u}(t,\overline{x}(t)) \tag{8.15}$$

est un contrôle optimal sur $[0,T]$.

Démonstration. On a

$$\frac{d}{dt}V(t,\overline{x}(t)) = \frac{\partial V}{\partial s}(t,\overline{x}(t)) + \frac{\partial V}{\partial \xi}(t,\overline{x}(t))^\dagger f(t,\overline{x}(t),\overline{u}(t)), \quad \text{p.p. } t \in [0,T].$$

Comme la fonction V satisfait l'équation HJB, on a

$$\frac{\partial V}{\partial s}(t,\overline{x}(t)) + H_\flat\left(t,\overline{x}(t),\frac{\partial V}{\partial \xi}(t,\overline{x}(t))\right) = 0.$$

Par définition de $\overline{u}$, on a

$$H_\flat\left(t,\overline{x}(t),\frac{\partial V}{\partial \xi}(t,\overline{x}(t))\right) = H\left(t,\overline{x}(t),\frac{\partial V}{\partial \xi}(t,\overline{x}(t)),\overline{u}(t)\right).$$

Comme $H(t,x,p,u) = p^\dagger f(t,x,u) + g(t,x,u)$, on obtient

$$\frac{d}{dt}V(t,\overline{x}(t)) = -g(t,\overline{x}(t),\overline{u}(t)), \quad \text{p.p. } t \in [0,T].$$

On en déduit que

$$\begin{aligned} V(0,x_0) = V(0,\overline{x}(0)) &= V(T,\overline{x}(T)) - \int_0^T \frac{d}{dt}V(t,\overline{x}(t))\,\mathrm{d}t \\ &= h(\overline{x}(T)) + \int_0^T g(t,\overline{x}(t),\overline{u}(t))\,\mathrm{d}t = J(0,x_0;\overline{u}), \end{aligned}$$

ce qui montre que $\overline{u}$ est bien un contrôle optimal. □

Proposition 8.8 (Fonction valeur et état adjoint)**.** *On suppose qu'il existe un contrôle optimal $\overline{u} : [0,T] \to U$ On note $\overline{x} = x_{\overline{u}} : [0,T] \to \mathbb{R}^d$ la trajectoire correspondante. Soit $\overline{p} : [0,T] \to \mathbb{R}^d$ l'état adjoint introduit dans le PMP, i.e., tel que $\frac{d\overline{p}}{dt}(t) = -\frac{\partial f}{\partial x}(t,\overline{x}(t),\overline{u}(t))^\dagger \overline{p}(t) - \frac{\partial g}{\partial x}(t,\overline{x}(t),\overline{u}(t))$, pour tout $t \in [0,T]$, et $\overline{p}(T) = \frac{\partial h}{\partial x}(\overline{x}(T))$. On suppose que la fonction valeur V est différentiable en $(s,\overline{x}(s))$ pour tout $s \in [0,T]$. Dans ces conditions, on a*

$$\overline{p}(s) = \frac{\partial V}{\partial \xi}(s,\overline{x}(s)), \quad \forall s \in [0,T]. \tag{8.16}$$

Démonstration. Pour $s < T$, on a $V(s, \overline{x}(s)) = J(s, \overline{x}(s); \overline{u}|_{I_s})$ de par le principe d'optimalité de Bellman. Pour tout $\xi \in \mathbb{R}^d$, on a $V(s,\xi) = \inf_{u \in \mathcal{U}_s} J(s,\xi;u) \leq J(s,\xi;\overline{u}|_{I_s})$. La fonction $\xi \mapsto V(s,\xi) - J(s,\xi,\overline{u}|_{I_s})$ est donc maximale en $\xi = \overline{x}(s)$. Par suite, on a

$$\frac{\partial}{\partial \xi_i} V(s, \overline{x}(s)) = \frac{\partial}{\partial \xi_i} J(s, \overline{x}(s), \overline{u}|_{I_s}), \quad \forall i \in \{1{:}d\}.$$

On vérifie aisément que

$$\frac{\partial}{\partial \xi_i} J(s, \overline{x}(s), \overline{u}|_{I_s}) = \int_{I_s} \frac{\partial g}{\partial x}(t, \overline{x}(t), \overline{u}(t))^\dagger y_i(t)\, \mathrm{d}t + \frac{\partial h}{\partial x}(\overline{x}(T))^\dagger y_i(T),$$

où $\dot{y}_i(t) = \frac{\partial f}{\partial x}(t, \overline{x}(t), \overline{u}(t)) y_i(t)$, pour tout $t \in I_s$, et $y_i(s) = e_i = (\delta_{ij})_{j \in \{1:d\}}$. En introduisant l'état adjoint et en intégrant par parties en temps, il vient

$$\frac{\partial}{\partial \xi_i} J(s, \overline{x}(s), \overline{u}|_{I_s}) = -\int_{I_s} \frac{d}{dt}\big(\overline{p}(t)^\dagger y_i(t)\big)\, \mathrm{d}t + \overline{p}(T)^\dagger y_i(T) = \overline{p}(s)^\dagger e_i.$$

Enfin, en $s = T$, il vient $\overline{p}(T) = \frac{\partial h}{\partial \xi}(\overline{x}(T)) = \frac{\partial V}{\partial \xi}(T, \overline{x}(T))$ puique $V(T,\xi) = h(\xi)$. □

8.2 Application au système LQ

On considère la famille de systèmes LQ (avec cible nulle pour simplifier)

$$\dot{x}_u(t) = A x_u(t) + B u(t), \quad \forall t \in I_s, \qquad x_u(s) = \xi, \qquad u \in L^2(I_s; \mathbb{R}^k), \tag{8.17a}$$

$$J(s,\xi;u) = \int_{I_s} \Big\{ \frac{1}{2} u(t)^\dagger R u(t) + \frac{1}{2} x_u(t)^\dagger Q x_u(t) \Big\}\, \mathrm{d}t + \frac{1}{2} x_u(T)^\dagger D x_u(T). \tag{8.17b}$$

Le Hamiltonien est

$$H(x,p,u) = p^\dagger (Ax + Bu) + \frac{1}{2} u^\dagger R u + \frac{1}{2} x^\dagger Q x, \tag{8.18}$$

et le Hamiltonien minimisé est

$$H_\flat(x,p) = \min_{u \in \mathbb{R}^k} H(x,p,u) = p^\dagger A x - \frac{1}{2} p^\dagger B R^{-1} B^\dagger p + \frac{1}{2} x^\dagger Q x, \tag{8.19}$$

l'unique minimiseur étant $\tilde{u} = -R^{-1} B^\dagger p$. La fonction valeur satisfait la condition finale $V(T,\xi) = \frac{1}{2}\xi^\dagger D \xi$ et l'équation HJB

$$\frac{\partial V}{\partial s} + \left(\frac{\partial V}{\partial \xi}\right)^\dagger A\xi - \frac{1}{2}\left(\frac{\partial V}{\partial \xi}\right)^\dagger B R^{-1} B^\dagger \frac{\partial V}{\partial \xi} + \frac{1}{2}\xi^\dagger Q \xi = 0. \tag{8.20}$$

On peut vérifier que la solution de l'équation HJB est de la forme

$$V(s,\xi) = \frac{1}{2}\xi^\dagger P(s) \xi, \tag{8.21}$$

où $P : [0,T] \to \mathbb{R}^{d\times d}$ est solution de l'équation de Riccati. En effet, en reportant l'expression (8.21) dans (8.20) et en utilisant la symétrie de $P(s)$ pour tout $s \in [0,T]$, il vient

$$\xi^\dagger\left(\frac{1}{2}\dot{P}(s) + P(s)^\dagger A - \frac{1}{2}P(s)^\dagger BR^{-1}B^\dagger P(s) + \frac{1}{2}Q\right)\xi = 0, \tag{8.22}$$

ce qui implique que la partie symétrique de la matrice entre parenthèses est nulle, ce qui n'est rien d'autre que l'équation de Riccati. On notera au passage que pour le système LQ, la fonction V est régulière sur $[0,T] \times \mathbb{R}^d$ (on notera également l'unicité du contrôle optimal).

Exemple 8.9. [Mouvement d'un point matériel] On considère le mouvement d'un point matériel avec un critère quadratique :

$$\dot{x}_u(t) = u(t), \quad x(s) = \xi, \quad J(s,\xi;u) = \int_s^T \frac{1}{2}(u(t)^2 + x_u(t)^2)\,\mathrm{d}t + \frac{1}{2}x_u(T)^2.$$

Le Hamiltonien est $H(x,p,u) = pu + \frac{1}{2}(x^2+u^2)$, et le Hamiltonien minimisé est $H_\flat(x,p) = \frac{1}{2}(x^2-p^2)$ avec $\tilde{u} = -p$ comme unique minimiseur. On obtient l'équation HJB et la condition finale

$$\frac{\partial V}{\partial s} + \frac{1}{2}\left(\xi^2 - \left(\frac{\partial V}{\partial \xi}\right)^2\right) = 0, \quad V(T,\xi) = \frac{1}{2}\xi^2.$$

En cherchant une solution de la forme $V(s,\xi) = \frac{1}{2}\mu(s)\xi^2$, il vient $\mu'(s) = \mu(s)^2 - 1$ et $\mu(T) = 1$, d'où $\mu \equiv 1$. Le contrôle optimal (sous forme de feedback) est

$$\tilde{u}(s,\xi) = -\frac{\partial V}{\partial \xi}(s,\xi) = -\mu(s)\xi = -\xi,$$

si bien que $\frac{d\overline{x}}{dt}(t) = -\overline{x}(t)$; d'où $\overline{x}(t) = x_0e^{-t}$ et $\overline{u}(t) = -x_0e^{-t}$. En guise de variante, on peut considérer le critère $J(s,\xi;u) = \int_s^T \frac{1}{2}(u(t)^2 + x_u(t)^2)\,\mathrm{d}t$ comme dans l'exemple 4.9. L'équation HJB est inchangée, mais la condition finale devient $V(T,\xi) = 0$. Il vient $\mu'(s) = \mu(s)^2 - 1$ et $\mu(T) = 0$, d'où $\mu(s) = \tanh(T-s)$. Le feedback optimal est $\tilde{u}(s,\xi) = -\tanh(T-s)\xi$, si bien que $\frac{d\overline{x}}{dt}(t) = -\tanh(T-t)\overline{x}(t)$; d'où $\overline{x}(t) = \frac{x_0}{\cosh(T)}\cosh(T-t)$ et $\overline{u}(t) = -\frac{x_0}{\cosh(T)}\sinh(T-t)$. □

8.3 Bilan : PMP ou HJB ?

Pour résumer les principaux résultats que nous avons vus sur le PMP et l'équation HJB, nous pouvons conclure avec les commentaires suivants. Le **PMP**

- fournit une condition nécessaire d'optimalité ;
- fournit le contrôle optimal en boucle ouverte (fonction du temps) ;
- repose sur la résolution d'équations différentielles ordinaires ;
- ne s'applique (sauf rares exceptions) qu'aux systèmes déterministes.

En revanche, la programmation dynamique via la résolution de l'équation **HJB**

- fournit une condition suffisante d'optimalité ;
- fournit le contrôle optimal en boucle fermée (fonction de l'état) ;
- repose sur la résolution d'une équation aux dérivées partielles (ce qui devient rapidement intractable lorsque la dimension d de l'espace des états croit) ;
- s'applique aux systèmes déterministes et stochastiques.

Annexe A

Stabilité des systèmes dynamiques

On se place ici en horizon de temps infini.

A.1 Notions de stabilité

On considère une fonction $f \in C^1(\mathbb{R}^d;\mathbb{R}^d)$, une condition initiale $x_0 \in \mathbb{R}^d$ et le système dynamique **autonome**

$$\dot{x}(t) = f(x(t)), \quad \forall t \geq 0, \qquad x(0) = x_0. \tag{A.1}$$

Définition A.1 (Point d'équilibre)**.** *On dit que le vecteur $\overline{x} \in \mathbb{R}^d$ est un **point d'équilibre** de* (A.1) *si*

$$f(\overline{x}) = 0 \ (\in \mathbb{R}^d). \tag{A.2}$$

Définition A.2 (Stabilité des points d'équilibre)**.** *Soit $\overline{x} \in \mathbb{R}^d$ un point d'équilibre de* (A.1)*.*
(i) *On dit que le point d'équilibre $\overline{x} \in \mathbb{R}^d$ est **stable** (on parle également de **stabilité orbitale**) si*

$$\exists \epsilon_0 > 0, \quad \forall \epsilon \in]0, \epsilon_0], \quad \exists \delta > 0, \quad \forall x_0 \in \overline{B}(\overline{x}, \delta), \quad x(t) \in \overline{B}(\overline{x}, \epsilon), \ \forall t \geq 0, \tag{A.3}$$

où $\overline{B}(\overline{x}, \delta)$ désigne la boule fermée de centre $\overline{x}$ et de rayon δ.
(ii) *On dit que le point d'équilibre $\overline{x} \in \mathbb{R}^d$ est **localement asymptotiquement stable (LAS)** si il est stable et de plus*

$$\lim_{t \to +\infty} x(t) = \overline{x}, \quad \forall x_0 \in \overline{B}(\overline{x}, \delta). \tag{A.4}$$

*On dit que le point d'équilibre $\overline{x} \in \mathbb{R}^d$ est **globalement asymptotiquement stable (GAS)** si il est stable et de plus*

$$\lim_{t \to +\infty} x(t) = \overline{x}, \quad \forall x_0 \in \mathbb{R}^d. \tag{A.5}$$

Une illustration est présentée à la figure A.1.

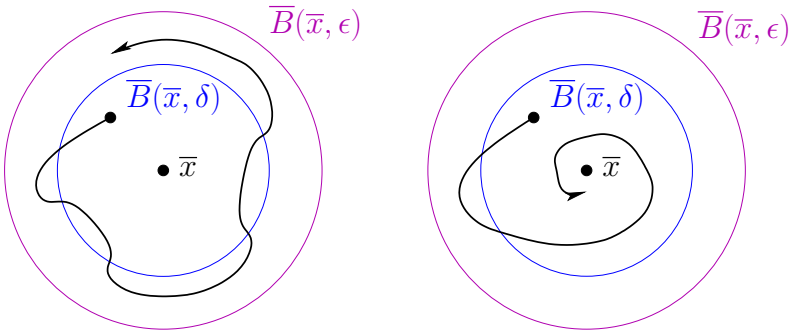

FIGURE A.1 – Point d'équilibre stable (à gauche) et point d'équilibre localement asymptotiquement stable (à droite).

La stabilité des points d'équilibre s'analyse très facilement dans le cas linéaire. Soit $A \in \mathbb{R}^{d\times d}$. On considère le système dynamique linéaire

$$\dot{x}(t) = Ax(t), \quad \forall t \geq 0, \qquad x(0) = x_0. \tag{A.6}$$

Il est clair que $\overline{x} = 0$ est point d'équilibre et que c'est le seul point d'équilibre si la matrice A est inversible. L'étude de la stabilité du point d'équilibre $\overline{x} = 0$ repose sur l'étude du spectre de la matrice A que l'on note $\sigma(A) \subset \mathbb{C}$. Pour une valeur propre $\lambda \in \sigma(A)$, on désigne par $\Re(\lambda)$ sa partie réelle.

Lemme A.3 (Stabilité, cas linaire)**.** *On considère le point d'équilibre $\overline{x} = 0 \in \mathbb{R}^d$.*
(i) *Si $\Re(\lambda) \leq 0$, $\forall \lambda \in \sigma(A)$, et toutes les valeurs propres à partie réelle nulle sont simples, alors $\overline{x} = 0$ est un point d'équilibre* ***stable****.*
(ii) *Si $\Re(\lambda) < 0$, $\forall \lambda \in \sigma(A)$ (on dit que la matrice A est* ***Hurwitz****), alors $\overline{x} = 0$ est un point d'équilibre* ***GAS****.*
(iii) *S'il existe une valeur propre $\lambda \in \sigma(A)$ telle que $\Re(\lambda) > 0$, alors le point d'équilibre $\overline{x} = 0$ est* ***instable****.*

L'analyse de la stabilité des points d'équilibre dans le cas non-linéaire est plus délicate. Une première analyse, locale, peut se faire par linéarisation.

Lemme A.4 (Stabilité locale, cas non-linéaire)**.** *Soit $\overline{x} \in \mathbb{R}^d$ un point d'équilibre du système dynamique* (A.1)*. On introduit la matrice*

$$A = \frac{\partial f}{\partial x}(\overline{x}) \in \mathbb{R}^{d\times d}. \tag{A.7}$$

(i) *Si la matrice A est* ***Hurwitz****, alors le point d'équilibre $\overline{x}$ est* ***LAS****.*
(ii) *S'il existe une valeur propre $\lambda \in \sigma(A)$ telle que $\Re(\lambda) > 0$, alors le point d'équilibre $\overline{x}$ est* ***instable****.*

Démonstration. Ces résultats se montrent en posant $\delta x(t) = x(t) - \overline{x}$ de sorte que $\dot{\delta x}(t) = f(\overline{x} + \delta x(t)) = A\delta x(t) + o(\delta x)$. □

Exemple A.5. [Pendule inversé] On considère un pendule inversé, i.e., avec la masse vers le haut et la tige vers le bas. On considère pour simplifier une masse et une longueur unités ($m = 1$, $l = 1$). 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). Le système dynamique s'écrit

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

En posant $x = (x_1, x_2)^\dagger = (\theta, \dot{\theta})^\dagger \in \mathbb{R}^2$, on obtient

$$\dot{x}(t) = f(x(t)), \qquad f(x) = \begin{pmatrix} x_2 \\ \sin(x_1) \end{pmatrix}.$$

On constate que $\overline{x} = (0,0)^\dagger$ est point d'équilibre. De plus, en évaluant la matrice

$$A = \frac{\partial f}{\partial x}(\overline{x}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

on constate que $\sigma(A) = \{-1, 1\}$, si bien que ce point d'équilibre est **instable**. □

Exemple A.6. [Oscillateur anharmonique amorti] Soit une fonction $g \in C^1(\mathbb{R}; \mathbb{R})$ vérifiant $g(0) = 0$, $g'(0) > 0$ et $xg(x) > 0$, pour tout $x \neq 0$. On s'intéresse au mouvement d'un point matériel (de masse unité) sous le champ de force décrit par la fonction g et d'un terme d'amortissement (dû par exemple au frottement). Ce mouvement est régi par le système différentiel d'ordre deux en temps

$$\ddot{x}(t) + \eta \dot{x}(t) + g(x(t)) = 0, \quad \forall t \geq 0,$$

où le paramètre réel $\eta \geq 0$ quantifie le terme d'amortissement. En posant $X(t) = (x(t), \dot{x}(t))^\dagger$, on obtient

$$\dot{X}(t) = F(X(t)), \quad F(X) = \begin{pmatrix} X_2 \\ -\eta X_2 - g(X_1) \end{pmatrix}.$$

On constate que $\overline{X} = (0,0)^\dagger$ est point d'équilibre. De plus, on obtient

$$A = \frac{\partial F}{\partial X}(X) = \begin{pmatrix} 0 & 1 \\ -g'(X_1) & -\eta \end{pmatrix}, \qquad A = \frac{\partial F}{\partial X}(\overline{X}) = \begin{pmatrix} 0 & 1 \\ -g'(0) & -\eta \end{pmatrix}.$$

Si $\eta < 2\sqrt{g'(0)}$, les 2 valeurs propres de A sont imaginaires pures et simples ; par conséquent, $\overline{X} = (0,0)^\dagger$ est point d'équilibre **stable** du système linéarisé. En revanche, si $\eta \geq 2\sqrt{g'(0)}$, alors la matrice A est Hurwitz, et le point d'équilibre $\overline{X} = (0,0)^\dagger$ est **GAS** pour le système linéarisé et **LAS** pour le système non-linéaire. □

A.2 Fonction de Lyapunov et principe d'invariance de LaSalle

Un outil puissant pour aller plus loin dans l'étude de la stabilité d'un point d'équilibre d'un système dynamique non-linéaire est la notion de fonction de Lyapunov. Soit $\overline{x} \in \mathbb{R}^d$ un point d'équilibre du système dynamique (A.1) et soit Ω un ouvert de $\mathbb{R}^d$ contenant $\overline{x}$.

Définition A.7 (Fonction de Lyapunov)**.** *On dit que la fonction $V : \Omega \to \mathbb{R}$ est une* ***fonction de Lyapunov*** *en $\overline{x}$ sur Ω si* (i) *V est de classe C^1 sur Ω ;* (ii) *$V(\overline{x}) < V(x)$, pour tout $x \in \Omega \setminus \{\overline{x}\}$;* (iii) *on a*

$$(\nabla V(x), f(x))_{\mathbb{R}^d} \leq 0, \quad \forall x \in \Omega, \tag{A.8}$$

où $(\cdot, \cdot)_{\mathbb{R}^d}$ désigne le produit scalaire usuel sur $\mathbb{R}^d$. On a donc

$$\frac{d}{dt} V(x(t)) = (\nabla V(x(t)), f(x(t)))_{\mathbb{R}^d} \leq 0, \quad \forall t \geq 0, \tag{A.9}$$

ce qui signifie que la fonction $t \mapsto V(x(t))$ décroît le long des trajectoires. Si l'inégalité (A.8) *est stricte sur $\Omega \setminus \{\overline{x}\}$, on dit que la fonction de Lyapunov est* ***stricte****. Enfin, on dit que la fonction de Lyapunov est* ***propre*** *sur Ω si l'image réciproque de tout compact dans $V(\Omega)$ est un compact. Une illustration est présentée à la figure A.2.*

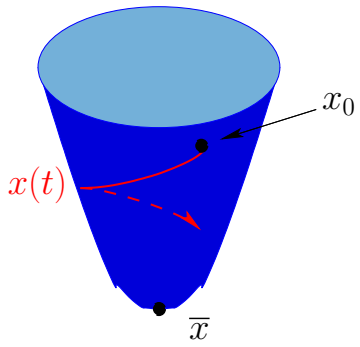

FIGURE A.2 – Graphe d'une fonction de Lyapunov en un point d'équilibre $\overline{x}$ et trajectoire $t \mapsto x(t)$.

Théorème A.8 (Lyapunov)**.** *Soit $\overline{x} \in \mathbb{R}^d$ un point d'équilibre du système dynamique* (A.1) *et soit Ω un ouvert de $\mathbb{R}^d$ contenant $\overline{x}$. On suppose qu'il existe une fonction de Lyapunov en $\overline{x}$ sur Ω. Alors, le point d'équilibre $\overline{x}$ est* ***stable****. De plus, si la fonction de Lyapunov est* ***stricte****, alors le point d'équilibre $\overline{x}$ est* ***LAS****. Enfin, si la fonction de Lyapunov est propre sur Ω, alors le point d'équilibre $\overline{x}$ est* ***GAS*** *sur Ω, i.e.,*

$$\lim_{t \to +\infty} x(t) = \overline{x}, \quad \forall x_0 \in \Omega. \tag{A.10}$$

Démonstration. Voir par exemple la section 5.7 de l'ouvrage [10]. □

Exemple A.9. [Oscillateur anharmonique amorti] On reprend l'exemple A.6 de l'oscillateur harmonique amorti, i.e., le système dynamique

$$\dot{X}(t) = F(X(t)), \qquad F(X) = \begin{pmatrix} X_2 \\ -\eta X_2 - g(X_1) \end{pmatrix}.$$

On a vu que $\overline{X} = (0,0)$ est point d'équilibre du système. On constate que

$$V(X) = \frac{1}{2}X_2^2 + \int_0^{X_1} g(x)\,\mathrm{d}x$$

est une fonction de Lyapunov en $\overline{X} = (0,0)^\dagger$ sur $\mathbb{R}^2$. En effet, on a $V(X) \geq 0$ et $V(X) = 0$ si et seulement si $X = (0,0)$; en outre,

$$(\nabla V(X), F(X))_{\mathbb{R}^2} = (g(X_1)\ X_2)\begin{pmatrix} X_2 \\ -\eta X_2 - g(X_1)\end{pmatrix} = -\eta X_2^2 \leq 0.$$

On en déduit que le point d'équilibre $\overline{X} = (0,0)^\dagger$ est **stable**. Comme la fonction de Lyapunov n'est pas stricte, on ne peut, à ce stade, aller plus loin dans l'application du théorème de Lyapunov A.8. □

Théorème A.10 (Principe d'invariance de LaSalle)**.** *Soit $\overline{x} \in \mathbb{R}^d$ un point d'équilibre du système dynamique* (A.1) *et soit Ω un ouvert de $\mathbb{R}^d$ contenant $\overline{x}$. On suppose qu'il existe une fonction de Lyapunov en $\overline{x}$ sur Ω et que celle-ci est propre. On note $\mathcal{S}$ le plus grand sous-ensemble de $\{x \in \Omega \mid (\nabla V(x), f(x))_{\mathbb{R}^d} = 0\}$ invariant par la dynamique. Alors, on a*

$$\lim_{t\to+\infty} d(x(t), \mathcal{S}) = \lim_{t\to+\infty} \inf_{y\in\mathcal{S}} |x(t) - y|_{\mathbb{R}^d} = 0, \quad \forall x_0 \in \Omega. \tag{A.11}$$

Démonstration. Voir [5]. □

Le principe d'invariance de LaSalle est utile si on sait montrer que $\mathcal{S} = \{\overline{x}\}$, i.e., que le sous-ensemble $\mathcal{S}$ est réduit au seul point d'équilibre $\overline{x}$. Dans ce cas, on peut déduire du théorème A.10 que le point d'équilibre $\overline{x}$ est **GAS** même si la fonction de Lyapunov n'est pas stricte.

Exemple A.11. [Oscillateur anharmonique amorti] On reprend à nouveau l'exemple A.6 de l'oscillateur harmonique amorti. On a vu ci-dessus que $(\nabla V(X), F(X))_{\mathbb{R}^2} = -\eta X_2^2 \leq 0$ (rappelons que $\eta > 0$). On cherche le plus grand sous-ensemble $\mathcal{S} \subset \{X \in \mathbb{R}^2 \mid X_2 = 0\}$ invariant par la dynamique. Comme on doit avoir $\dot{X}_1(t) = X_2(t) = 0$, il vient $X_1(t) = X_{1,0}$, et donc $0 = \dot{X}_2(t) = g(X_{1,0})$, d'où $X_{1,0} = 0$. En conclusion, $\mathcal{S} = \{\overline{X}\} = \{(0,0)^\dagger\}$. Par le principe d'invariance de LaSalle, le point d'équilibre $\overline{X} = (0,0)^\dagger$ est **GAS**. □

A.3 Stabilisation par retour d'état

On considère le système de contrôle linéaire autonome

$$\dot{x}(t) = Ax(t) + Bu(t), \quad \forall t \geq 0, \qquad x(0) = x_0, \tag{A.12}$$

avec $A \in \mathbb{R}^{d\times d}$ et $B \in \mathbb{R}^{d\times k}$.

Définition A.12 (Boucle par retour d'état)**.** *On dit que le système de contrôle linéaire autonome* (A.12) *est* ***bouclé par retour d'état*** *(on dit aussi* ***bouclé par feedback****) s'il existe une matrice $K \in \mathbb{R}^{k\times d}$ telle que*

$$u(t) = Kx(t), \quad \forall t \geq 0. \tag{A.13}$$

La matrice K est appelée ***matrice de feedback****. Dans ces conditions, le système linéaire de contrôle bouclé par retour d'état s'écrit*

$$\dot{x}(t) = (A + BK)x(t), \quad \forall t \geq 0. \tag{A.14}$$

Définition A.13 (Stabilisation asymptotique)**.** *On dit que le système de contrôle linéaire est* ***stabilisable asymptotiquement*** *s'il existe une matrice de feedback $K \in \mathbb{R}^{k\times d}$ telle que la matrice $A + BK$ soit Hurwitz, i.e.,*

$$\Re(\lambda) < 0, \quad \forall \lambda \in \sigma(A + BK). \tag{A.15}$$

Lorsque le système de contrôle linéaire est stabilisable asymptotiquement, toute trajectoire du système bouclé (A.14) tend vers le point d'équilibre $\overline{x} = 0$, i.e., le point d'équilibre $\overline{x} = 0$ est GAS pour le système bouclé.

Proposition A.14 (Contrôlabilité $\Longrightarrow$ Stabilisable asymptotiquement)**.** *Si le système de contrôle linéaire* (A.12) *est* ***contrôlable****, il est* ***stabilisable asymptotiquement****.*

Démonstration. La preuve de ce résultat repose sur un résultat d'algèbre linéaire, le théorème de placement des pôles A.15, qui est rappelé ci-dessous (pour la preuve, on pourra se référer à celle du théorème 13.34 dans [11]). Le système de contrôle linéaire (A.12) étant contrôlable, les matrices A et B vérifient la condition de Kalman (1.16). Grâce au théorème A.15, on déduit l'existence d'une matrice $K \in \mathbb{R}^{k\times d}$ telle que le polynôme caractéristique de $A + BK$ soit tel que $\chi_{A+BK}(\lambda) = (\lambda+1)^d$. Ceci montre que la matrice $A+BK$ est Hurwitz. Par suite, le point d'équilibre $\overline{x} = 0$ est GAS pour le système bouclé. □

Théorème A.15 (Placement des pôles)**.** *On note χ_M le polynôme caractéristique d'une matrice carrée M, i.e., $\chi_M(\lambda) = \det(\lambda I - M)$. Soit $A \in \mathbb{R}^{d\times d}$ et $B \in \mathbb{R}^{d\times k}$. On suppose que les matrices A et B vérifient la condition de Kalman* (1.16)*. Alors, pour tout polynôme π unitaire de degré d, il existe une matrice $K \in \mathbb{R}^{k\times d}$ telle que*

$$\chi_{A+BK}(\lambda) = \pi(\lambda). \tag{A.16}$$

On considère maintenant le système de contrôle non-linéaire autonome

$$\dot{x}(t) = f(x(t), u(t)), \quad \forall t \geq 0, \qquad x(0) = x_0, \tag{A.17}$$

où la fonction $f : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ est de classe C^1. On suppose que la paire $(\overline{x}, \overline{u}) \in \mathbb{R}^d \times \mathbb{R}^k$ est telle que $f(\overline{x}, \overline{u}) = 0$. Ainsi, $\overline{x}$ est point d'équilibre pour le système dynamique $\dot{x}(t) = f(x(t), \overline{u})$. Le système dynamique linéarisé en ce point est $\dot{y}(t) = Ay(t) + Bv(t)$ avec

$$A = \frac{\partial f}{\partial x}(\overline{x}, \overline{u}), \qquad B = \frac{\partial f}{\partial u}(\overline{x}, \overline{u}). \tag{A.18}$$

Corollaire A.16 (Système non-linéaire bouclé)**.** *On suppose que le système linéarisé est stabilisable asymptotiquement par le feedback $v(t) = Ky(t)$. Alors, le point d'équilibre $(\overline{x}, \overline{u})$ est* ***LAS*** *pour le système non-linéaire bouclé*

$$\dot{x}(t) = f(x(t), \overline{u} + K(x(t) - \overline{x})),\ \forall t \geq 0. \tag{A.19}$$

Exemple A.17. [Pendule inversé] On rappelle que la dynamique du pendule inversé est décrite par l'équation différentielle d'ordre deux

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

On considère la paire $(\overline{x}, \overline{u}) = (0, 0)$. On obtient

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Il vient

$$C = (B, AB) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

ce qui montre que la condition de Kalman est vérifiée, et donc la contrôlabilité locale du système non-linéaire. On peut choisir la matrice de feedback $K = (2\ 2)$, ce qui donne

$$A + BK = \begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix},$$

et par suite $\chi_{A+BK}(\lambda) = (\lambda + 1)^2$. La commande en boucle fermée s'écrit $u(t) = 2(\theta(t) + \dot{\theta}(t))$. □

Bibliographie

[1] Jean-Pierre Aubin. *Mathematical methods of game and economic theory*, volume 7 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam-New York, 1979.

[2] Martino Bardi and Italo Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Systems & Control : Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1997. With appendices by Maurizio Falcone and Pierpaolo Soravia.

[3] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[4] R. Fletcher. *Practical methods of optimization. Vol. 1.* John Wiley & Sons, Ltd., Chichester, 1980. Unconstrained optimization, A Wiley-Interscience Publication.

[5] Henry Hermes and Joseph P. LaSalle. *Functional analysis and time optimal control*. Academic Press, New York-London, 1969. Mathematics in Science and Engineering, Vol. 56.

[6] Alberto Isidori. *Nonlinear control systems*. Communications and Control Engineering Series. Springer-Verlag, Berlin, third edition, 1995.

[7] E. B. Lee and L. Markus. *Foundations of optimal control theory*. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, second edition, 1986.

[8] Pierre-Louis Lions. *Contrôle de modèles dynamiques*. Cours polycopié. École Polytechnique, 2016.

[9] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[10] Eduardo D. Sontag. *Mathematical control theory*, volume 6 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 1998. Deterministic finite-dimensional systems.

[11] Emmanuel Trélat. *Contrôle optimal*. Mathématiques Concrètes. Vuibert, Paris, 2005. Théorie & applications.

[12] Richard Vinter. *Optimal control*. Systems & Control : Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 2000.