

MAP 561

Introduction à l'automatique

Ugo Boscain et Yacine Chitour

Notes de cours

Édition 2010/2011

Table des matières

Avant-propos	7
1 Introduction	9
1.1 Mise sous forme d'état et définition du système commandé	10
1.2 Commandabilité	11
1.3 Bouclage	12
1.4 Stabilisation	12
1.5 Observabilité	14
1.6 Commande optimale	15
1.7 Plan du cours	15
2 Equations différentielles ordinaires et stabilité	17
2.1 Théorie générale des équations différentielles	17
2.1.1 Existence et unicité	18
2.1.2 Solutions maximales et durée de vie	20
2.1.3 Flots, portraits de phase	23
2.1.4 Équations différentielles linéaires	27
2.1.5 Linéarisation et perturbation du flot	32
2.2 Équations différentielles linéaires autonomes	37
2.2.1 Approche élémentaire	37
2.2.2 Exponentielle de matrices	39
2.2.3 Calcul de l'exponentielle de matrices	41
2.2.4 Forme des solutions	47
2.3 Stabilité	52
2.3.1 Équilibres et stabilité	52
2.3.2 La stabilité par la linéarisation	54
2.3.3 Fonctions de Lyapunov	57
3 Commandabilité et observabilité des systèmes linéaires	65
3.1 Systèmes de commande	65
3.2 Commandabilité	68
3.3 Planification de trajectoires	73

3.3.1	Exemple	73
3.3.2	Forme de Brunovsky	73
3.3.3	Application à la planification de trajectoires	74
3.3.4	Preuve du théorème 3.7 pour le cas mono-entrée	74
3.4	Stabilisation	75
3.5	Observabilité	78
3.5.1	Définition et critère d'observabilité de Kalman	78
3.5.2	Stabilisation par retour d'état statique	82
3.5.3	Observateur asymptotique de Luenberger	82
3.5.4	Stabilisation par retour dynamique de sortie	83
4	Commandabilité des systèmes non linéaires	85
4.1	Commandabilité locale et globale	85
4.2	Crochets et algèbres de Lie	86
4.3	Accessibilité locale et conditions suffisantes pour la commandabilité globale	88
4.4	Champs compatibles	90
4.5	Orbites et conditions nécessaires pour la commandabilité	94
5	Théorie linéaire-quadratique	97
5.1	Existence de trajectoires optimales	98
5.2	Condition nécessaire et suffisante d'optimalité : principe du maximum dans le cas LQ	101
5.3	Fonction valeur et équation de Riccati	104
5.3.1	Définition de la fonction valeur	104
5.3.2	Equation de Riccati	104
5.3.3	Représentation linéaire de l'équation de Riccati	108
5.4	Applications de la théorie LQ	109
5.4.1	Problèmes de régulation	109
5.4.2	Filtre de Kalman déterministe	114
5.4.3	Régulation sur un intervalle infini et rapport avec la stabilisation .	117
6	Temps-optimalité pour les systèmes linéaires	121
6.1	Existence de trajectoires temps-optimales	121
6.2	Condition nécessaire d'optimalité : principe du maximum dans le cas linéaire	125
6.3	Exemple : Synthèse optimale pour le problème de l'oscillateur harmonique	128
6.3.1	Contrôlabilité du système	129
6.3.2	Interprétation physique	129
6.3.3	Calcul du contrôle optimal	129
7	Contrôle optimal non-linéaire	135
7.1	Enoncé général du Principe du maximum de Pontryagin	135
7.2	Le problème sous-riemannien	136
7.3	Temps minimum pour un système affine bidimensionnel	137
7.3.1	Trajectoires singulières et détermination des commutations	138

7.4	Principe du maximum de Pontryagin faible	141
7.4.1	Régularité de l'application entrée-sortie	142
7.4.2	Caractérisation hamiltonienne des contrôles singuliers	143
7.4.3	Démonstration du Théorème 7.7	144

Avant-propos

Le cours « Introduction à l'automatique » a pour objectif de présenter les concepts de base de l'automatique linéaire. On utilise l'approche par représentation d'état, qui repose sur les équations différentielles ordinaires.

Ces notes de cours se composent de six parties. La première est une rapide présentation de l'automatique à travers l'étude d'un exemple classique, celui de la commande d'un bras de robot et fait l'objet du chapitre 1. En particulier, on expliquera pourquoi les équations différentielles ordinaires (EDO) sont utilisées et la nécessité d'une bonne connaissance de leurs propriétés fondamentales avant d'aborder la résolution de problèmes en automatique. Il est à souligner que cette partie est très complète et que certains points abordés ont déjà été vus. Dans ce cas, on pourra passer rapidement sur ces points.

La seconde partie est développée dans le chapitre 2 et est consacrée à l'étude des équations différentielles ordinaires (EDO) ainsi qu'à leur utilisation pour la modélisation en physique, mécanique, économie, biologie... L'accent est principalement mis sur deux points : tout d'abord la notion de stabilité dont l'importance, pour de nombreux problèmes pratiques, est comparable à celle de la connaissance effective des solutions ; et ensuite une description détaillée des solutions des EDO linéaires à coefficients constants.

Dans la troisième partie (chapitre 3), sont abordées les notions essentielles de l'automatique telles que la commandabilité, l'observabilité et la stabilisation des systèmes linéaires commandés. On établira en détail comment se caractérisent ces propriétés sous la forme de critères classiques tels que celui de Kalman ou le théorème de placement de pôles. On présentera aussi une solution effective à la question de commandabilité grâce à la sortie de Brunovski ainsi que l'observateur de Luenberger et le principe de séparation qui en résulte.

Dans le chapitre 4, on abordera l'étude de la commandabilité pour les systèmes non linéaires, en se contentant de ceux qui sont affines en la commande. On mettra en évidence la notion de crochet de Lie, centrale pour décrire l'ensemble atteignable. On donnera les principaux résultats généraux comme le théorème de Krener pour l'*accessibilité* et les théorèmes de Chow-Rashevski, Nagano et Sussman pour ce qui est de la commandabilité et la description de l'orbite.

Les deux dernières parties sont à des questions de commande optimale. La cinquième et dernière traite plus particulièrement de deux aspects classiques de la commande optimale : (a) la commande linéaire quadratique et son application la plus fameuse, le filtre

de Kalman ; (b) le principe du maximum de Pontryagin (PMP) appliqué aux systèmes linéaires et à la minimisation du temps.

Enfin, la sixième et dernière partie traite de questions de contrôle optimal pour des systèmes non linéaires avec un énoncé général du PMP et ensuite son application pour caractériser les trajectoires optimales dans le cadre sous-riemannien et pour l'étude de la synthèse optimale pour des systèmes en dimension deux.

Les résultats sont parfois accompagnés de leur preuve. Lorsque celle-ci n'est pas utile à la compréhension du cours, elle est écrite en petits caractères (petits comme ceci) et est précédée du symbole *. Le même traitement (petits caractères et symbole *) est appliqué aux parties les plus avancées du document, qui ne seront pas traitées en cours. Il n'est cependant pas interdit de les lire. . . Enfin, le symbole " $:=$ " signifie que ce qui est à gauche du symbole est défini par ce qui est à droite.

Enfin, le présent cours a été enseigné par Pierre Rouchon et Frédéric Bonnans entre 1994 et 2004 sous le titre « Commande et Optimisation de systèmes dynamiques » et fait l'objet de l'ouvrage [Rou – Bo] extrêmement riche en exemples et qu'il est fortement conseillé de consulter. De plus les auteurs du présent document remercient Frédéric Jean et Emmanuel Trélat pour avoir autorisé de nombreux emprunts au très beau cours "Equations différentielles et fondements de l'automatique" ainsi qu'à l'excellent ouvrage "Contrôle optimal : théorie et applications" chez Vuibert.

Par ailleurs, ces notes de cours sont loin d'être parfaites et les auteurs seront grés à toute personne leur signalant des corrections à effectuer.

Chapitre 1

Introduction

L'automatique (ou théorie du contrôle) est la science qui traite des lois de régulation des systèmes commandés. Commander un objet (on dit aussi le contrôler ou l'asservir) signifie influencer son comportement pour lui faire effectuer une tâche définie à l'avance. Afin de réaliser en pratique cette "influence", les ingénieurs ont mis au point des mécanismes appropriés faisant appel à des principes théoriques généraux, eux-mêmes s'exprimant à l'aide de divers outils mathématiques. Ainsi, ces mécanismes vont du régulateur à boules de Watt (pour les moteurs à vapeur) aux microprocesseurs les plus sophistiqués que l'on peut trouver dans les CD players, les automobiles, ou encore dans les robots industriels ou les pilotes automatiques des avions. L'étude de ces mécanismes et leur interaction avec l'objet à commander est l'objet de ce cours.

Nous allons illustrer plus en détail notre propos à l'aide d'un exemple simple issu de la robotique. Nous reprenons la description qui en est faite dans [Rou – Bo]. Il s'agit d'un bras rigide tournant dans un plan vertical autour d'un axe horizontal (Ox) et ce dernier est équipé d'un moteur délivrant un couple variable $u \in \mathbb{R}$ que l'on peut choisir *arbitrairement* à chaque instant : u est la *commande* du système (ou encore *entrée*). La position géométrique du système est décrite par un angle $\theta \in S^1$ où l'on a noté le cercle unité par S^1 (c'est-à-dire que S^1 est la "sphère" de dimension 1). La dynamique du système est obtenue à partir de la conservation du moment cinétique autour de l'axe (Ox) :

$$J\ddot{\theta}(t) + mlg \sin \theta(t) = u(t), \quad (1.1)$$

avec m la masse du bras, J son moment d'inertie par rapport à l'axe (Ox), l la distance du centre de gravité à (Ox) et g l'accélération due à la pesanteur. On aura reconnu l'équation d'un pendule pesant sans frottement. Bien que les constantes qui interviennent dans ce problème jouent un rôle prépondérant en pratique, on supposera dans la suite que $mlg = J = 1$. La dynamique du bras est donc

$$\ddot{\theta}(t) + \sin \theta(t) = u(t). \quad (1.2)$$

Pour un angle θ_r fixé, un objectif possible de commande est d'amener le bras à l'angle θ_r et de l'y maintenir ensuite. On dira alors que le bras est en position d'équilibre à $\theta = \theta_r$ et l'objectif de commande est de *stabiliser* le système en θ_r . Plus généralement, on pourra se donner comme objectif de *suivre une trajectoire* de référence $\theta_r(\cdot)$ qui vérifie la dynamique (1.2) pour une commande de référence $u_r(\cdot)$ avec les fonctions du temps θ_r, u_r définies sur un intervalle de temps $[0, T_r]$ (T_r pouvant être fini ou non).

Remarque. Remarquons que l'objectif initial est un cas particulier du suivi de trajectoire puisque amener le bras à un angle $\theta = \theta_1$ l'y maintenir ensuite est équivalent à suivre la trajectoire de (1.2) associée à $\theta_r(\cdot) \equiv \theta_r$ et $u_r(\cdot) \equiv \sin \theta_r$.

Nous allons maintenant décrire les étapes successives que suit l'automaticien pour résoudre le problème.

1.1 Mise sous forme d'état et définition du système commandé

Un intervalle de temps $[0, T]$ doit être fixé pour toutes les fonctions temporelles que nous envisagerons. Ici, il est naturel de prendre $T = T_r$, temps auquel on veut amener le bras en position d'équilibre. Afin de connaître l'évolution du mouvement lorsque l'on applique un couple $u(\cdot) : [0, T] \rightarrow \mathbb{R}$, il faut intégrer l'équation différentielle (1.2), qui est du second ordre donc. Pour cela, il faut connaître, à l'instant $t = 0$, la position angulaire $\theta(0) = \theta_0$ et la vitesse angulaire $\dot{\theta}(0) = \omega_0$. La paire (θ_0, ω_0) représente les conditions initiales du système différentiel du premier ordre suivant, obtenu à partir de (1.2) :

$$(S) \quad \begin{cases} \dot{\theta}(t) &= \omega, \\ \dot{\omega}(t) &= -\sin \theta(t) + u. \end{cases} \quad (1.3)$$

À l'instant t , l'état du système est donc uniquement déterminé par la donnée de $(\theta(t), \omega(t))$ et $u(t)$. La variable $x = (\theta, \omega)$ forme l'état du système, qui est donc un point de $S^1 \times \mathbb{R}$, et le couple de fonctions $t \mapsto (x(t), u(t))$ avec $t \in [0, T]$ est appelé trajectoire du système.

De manière plus classique, une trajectoire est la *réponse* du système à la commande $u(\cdot)$ (appelée aussi entrée).

La dynamique (S) peut être réécrite en termes de x comme suit :

$$(S) \quad \dot{x}(t) = F(x(t), u(t)), \quad (1.4)$$

avec $F : S^1 \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^2$ est le *champ de vecteurs* qui, à tout point $(x, u) = (\theta, \omega) \in S^1 \times \mathbb{R} \times \mathbb{R}$ associe le vecteur $(\omega, -\sin \theta + u) \in \mathbb{R}^2$.

Enfin, on appellera système commandé (associé au bras de robot) (Σ), l'ensemble des données suivantes : l'espace d'état $S^1 \times \mathbb{R}$, l'espace de commande \mathbb{R} , la dynamique (1.4) et

la classe des commandes admissibles Ad , c'est-à-dire l'ensemble des fonctions $u : [O, T] \rightarrow \mathbb{R}$ (continues, continues par morceaux, polynômes, fonctions bornées etc.).

Comme étude préliminaire à ce système (et aux objectifs de commande que l'on veut atteindre), il est impératif de comprendre ce qui se passe si "l'on ne fait rien", c'est-à-dire en mettant $u = 0$ dans (1.4). On est alors conduit à étudier l'équation différentielle ordinaire (EDO) définie sur $S^1 \times \mathbb{R}$ par

$$(S)_0 \begin{cases} \dot{\theta}(t) = \omega, \\ \dot{\omega}(t) = -\sin \theta(t), \end{cases} \quad (1.5)$$

ou encore $\dot{x} = F(x(t), 0)$. En automatique, cette étape est appelée *étude de la réponse libre*. Il est particulièrement important de savoir quel est le comportement des trajectoires libres lorsque t tend vers l'infini. On parle alors d'étude de la stabilité asymptotique de l'EDO (1.5). Par exemple, est-ce que les trajectoires peuvent converger vers un point de $S^1 \times \mathbb{R}^2$? De tels points sont appelés *points d'équilibre* du système libre et correspondent aux trajectoires constantes. Un simple calcul montre que les seuls points d'équilibre du bras sont $(0, 0)$ et $(\pi, 0)$. (Une trajectoire constante correspond à annuler le membre de droite.)

1.2 Commandabilité

Étant donnés deux états $x_0 = (\theta_0, \omega_0)$ et $x_1 = (\theta_1, \omega_1)$ dans l'espace d'états, le problème de commandabilité entre ces deux états consiste à trouver une trajectoire de (Σ) (c'est-à-dire un triplé $t \mapsto (\theta(t), \omega(t), u(t))$) partant de x_0 en $t = 0$ et arrivant en x_1 en $t = T$. En d'autres termes, il s'agit de trouver la commande u pour amener le système d'un état à un autre. Si cela est possible, on dira que (Σ) est commandable entre x_0 et x_1 et qu'il est complètement commandable s'il est commandable pour toutes paires d'états. Il faut remarquer que la question de commandabilité peut se scinder en deux :

- (Q1) étant donné (Σ) , peut-on montrer qu'il est ou non complètement commandable, et ce sans explicitation des commandes?
- (Q2) Étant donnée une paire (x_0, x_1) d'états commandables par (Σ) , donner une procédure effective pour déterminer une commande qui amène le système de x_0 à x_1 .

La question (Q1) est de nature théorique et est loin d'être résolue à l'heure actuelle. Nous l'étudierons dans le cas particulier des systèmes linéaires. On peut exprimer le problème de commandabilité en termes d'algèbre linéaire et on peut alors donner une condition nécessaire et suffisante sur la dynamique du système commandé qui caractérise la commandabilité.

La question (Q2), appelée aussi *planification de trajectoires*, est encore plus difficile à résoudre que (Q1). Nous proposerons, toujours pour les systèmes linéaires, une solution.

1.3 Bouclage

Pour le bras rigide, une manière assez naturelle de déterminer la commande $u(\cdot)$ qui doit réaliser notre objectif est de procéder comme suit : on sait d'où on part (du point x_0) et où on doit arriver (au point $x_r = (\theta_r, 0)$) en temps T . Supposons que, uniquement à partir de la connaissance de x_0, x_r et T , on soit maintenant capable de calculer une commande $u(\cdot)$ amenant le bras de x_0 à x_r . On dit alors que l'on commande *en boucle ouverte*. Cette façon de faire présente au moins deux défauts :

- à l'instant $t = 0$, on est censé calculer toute la loi de commande $u : t \in [0, T]$ puis, l'implémenter dans le système physique, pratiquement à l'instant $t = 0$. Cela suppose que le temps de calcul de la commande $u(\cdot)$ est négligeable par rapport à celui du système. Pour certaines applications, cela est irréaliste ;
- supposons qu'il arrive un incident sur l'intervalle $[0, T]$ qui ne soit pas pris en compte par la dynamique du système. La loi de commande ayant déjà été calculée à l'avance, le système ne pourra pas réagir à l'incident imprévu. (Il faut remarquer que cet "imprévu" peut survenir constamment si le modèle qui est utilisé n'est qu'approximatif!)

C'est pour cela qu'il convient parfois de calculer $u(\cdot)$ de manière plus simple et de l'ajuster en temps réel de façon à compenser les écarts instantanés à la trajectoire de référence, $\theta - \theta_r$ et $\omega - \omega_r$ qui peuvent apparaître. Par exemple, on peut choisir, pour $t \in [0, T]$, $u(t)$ en terme de $\theta(t) - \theta_r$ et $\omega(t) - \omega_r$. L'utilisation de ce type de termes correspond à un *bouclage ou retour d'état*, appelée aussi *feedback*. Remarquons qu'une commande de ce type suppose la connaissance à tout instant $t \in [0, T]$ des quantités $\theta(t) - \theta_r$ et $\omega(t) - \omega_r$.

1.4 Stabilisation

Rappelons que notre objectif est de stabiliser le bras au point $x_r = (\theta_r, 0)$. Supposons que l'on atteigne x_r , on y reste avec la commande constante $u_r := \sin \theta_r$. De plus, il est clair que l'on peut s'approcher de θ_r en temps fini et rester dans un voisinage de cet angle avec une vitesse angulaire ω , elle aussi pas trop grande. Être capable alors de faire diminuer **en même temps** $\theta(t) - \theta_r$ et $\omega(t)$ vers zéro est un peu moins évident. Une façon de faire est alors de linéariser (Σ) au voisinage de x_r c'est-à-dire d'écrire

$$x = x_r + \delta x \text{ avec } \delta x = (\delta\theta, \delta\omega) \quad u = u_r + \delta u.$$

Ici, δx et δu sont petits. Pour écrire la dynamique de δx à partir de (1.3), on effectue un développement limité des seconds membres de (1.3) en ne retenant que les termes d'ordre 1 en $\delta x, \delta u$. On obtient alors *le linéarisé tangent* $(S)_L$ de (S) le long de la trajectoire constante x_r .

$$(S)_L \begin{cases} \dot{\delta\theta}(t) &= \delta\omega, \\ \dot{\delta\omega}(t) &= \delta u(t) - \cos(\theta_r)\delta\theta(t). \end{cases} \quad (1.6)$$

On peut aussi écrire ce système

$$(S)_L \quad \dot{\delta x} = A\delta x + b\delta u, \quad (1.7)$$

avec

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ et } b := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

L'objectif de stabilisation consiste maintenant à amener tout point de \mathbb{R}^2 en $(0,0)$ le long de $(S)_L$. Remarquons que le membre de droite de $(S)_L$ est linéaire en $(\delta\theta, \delta\omega, \delta u)$. On dira alors que l'on a affaire à un système commandé linéaire stationnaire (c'est-à-dire ne dépendant pas explicitement du temps) et à coefficients constants. Comme nous le verrons, ces caractéristiques de linéarité permettent d'exprimer les trajectoires de ce système avec de manière explicite. Afin de stabiliser le système tout en conservant un caractère linéaire, il est naturel de choisir δu linéaire en δx ,

$$\delta u = -k^T \delta x = -k_1 \delta\theta - k_2 \delta\omega, \quad (1.8)$$

avec $k = (k_1, k_2)^T$ vecteur constant appelé *gain du contrôleur*. La loi de feedback précédente est dénommée retour d'état ou feedback *statique*. Avec ce choix pour $u(\cdot)$, le système bouclé s'écrit

$$(Lin)_k \quad \dot{\delta x} = (A + bk^T)\delta x.$$

On pose

$$A(k) := A + bk^T.$$

Il s'agit donc de déterminer k de telle sorte que l'EDO linéaire $(Lin)_k$ soit *asymptotiquement stable* : pour toute condition initiale $\delta x_0 \in \mathbb{R}^2$, ses solutions tendent vers zéro lorsque t tend vers l'infini. Cela est équivalent au problème d'algèbre linéaire suivant : trouver un vecteur k tel que les valeurs propres de $A(k)$ soient de partie réelle strictement négative. Ces valeurs propres sont appelés les *pôles* du système bouclé $(Lin)_k$. On verra qu'il suffit de choisir $k_1, k_2 > 0$ et la convergence des solutions (vers zéro) est alors exponentielle. Par exemple, si les valeurs propres λ_1, λ_2 de $A(k)$ sont réelles, alors toute solution est combinaison linéaire de $\exp(\lambda_1 t)$ et $\exp(\lambda_2 t)$.

Enfin, on peut vouloir aussi contrôler la vitesse à laquelle on stabilise le bras autour de x_r ou encore demander à ce qu'il n'y ait pas d'oscillation. Cela revient dans le premier cas à contrôler la vitesse de convergence de $(Lin)_k$ et dans le second à n'avoir que des valeurs propres réelles pour $A(k)$. On a deux instances particulières du problème plus général qui consiste à déterminer le vecteur k afin que $A(k)$ ait des valeurs propres vérifiant certaines conditions. On arrive donc à la question purement algébrique suivante : "étant donné une paire (A, B) avec A matrice 2×2 et b vecteur colonne, caractériser l'ensemble de toutes les valeurs propres de $A(k) = A + bk^T$ lorsque k est un vecteur quelconque. Le *théorème de*

placement de pôles dit essentiellement que si la paire (A, B) est commandable (c'est-à-dire que la matrice ayant pour colonnes b, Ab est inversible), alors l'ensemble précédent est \mathbb{C} tout entier.

1.5 Observabilité

La loi de feedback donnée en (1.8) suppose que l'on mesure **à tout instant** $t \in [0, T]$ l'état complet du système $x = (\theta, \omega)$. En pratique, les capteurs de vitesse sont très onéreux. Il est donc souvent raisonnable de supposer que l'on ne mesure que la position et ici cela signifie que l'on ne connaît de manière instantanée que θ . Les quantités mesurées constituent la *sortie* d'un système commandé. Celle-ci représente une information sur l'état, instantanée mais souvent partielle. Il est clair qu'une loi de feedback n'a de réalité pratique que si cette dernière ne s'obtient qu'à l'aide de la sortie.

Revenons au système commandé avec comme unique sortie la position angulaire θ . Si l'on essaye un feedback statique avec seulement une fonction de θ , on peut montrer qu'il est impossible de stabiliser le bras. Par exemple, si l'on prend un feedback linéaire en θ , on aboutit à une équation du type

$$\delta\ddot{\theta} + k\delta\dot{\theta} = 0,$$

qui n'est pas asymptotiquement stable quelle que soit la valeur de k .

Cependant, on remarque que l'on peut obtenir $\omega(t)$ en dérivant $\theta(t)$. On dit alors que l'état x du système est *observable* à partir de la sortie θ . Plus généralement, on verra que l'état est observable à partir d'une sortie y si l'on peut reconstruire x à partir d'un nombre fini de dérivées de y .

Pour le bras, nous pouvons dériver numériquement le signal mesuré θ pour en déduire ω et ainsi construire une loi de feedback qui stabilise. Cette solution fonctionne si la mesure de θ n'est pas trop bruitée. Dans le cas contraire, l'opération numérique de la dérivation est à éviter. Après linéarisation, l'idée est alors d'*estimer* l'état δx à partir de la seule connaissance de l'angle $\delta\theta$ **sans dériver** $\delta\theta$. Pour cela, il faut utiliser une autre information sur δx : il vérifie la dynamique $(S)_L$! On cherche alors à construire un état artificiel $\tilde{\delta x}$ tel que $\delta x - \tilde{\delta x}$ tende vers zéro lorsque t tend vers l'infini. Un tel $\tilde{\delta x}$ est appelé observateur asymptotique. Notons $\delta\theta = C\delta x$ avec C le vecteur ligne égal à $(1 \ 0)$. On choisit $\tilde{\delta x}$ comme trajectoire de

$$(S)_L \quad \dot{\tilde{\delta x}} = A\tilde{\delta x} + b\delta u + LC(\tilde{\delta x} - \delta x), \quad (1.9)$$

avec L un vecteur colonne à déterminer. Ici, on a bien sur $C(\tilde{\delta x} - \delta x) = (\tilde{\delta\theta} - \delta\theta)$. Remarquons que la dynamique (1.9) est obtenue en ajoutant, à la dynamique linéarisée du bras, le terme $LC(\tilde{\delta x} - \delta x)$ qui ne fait intervenir, mis à part des termes "artificiels", que la sortie $\delta\theta$. Lorsque l'on considère la dynamique de l'*erreur* $e := \delta x - \tilde{\delta x}$, on a

$$\dot{e} = (A + LC)e.$$

Posons $A(L) := A + LC$. Faire tendre e vers zéro lorsque t tend vers l'infini devient encore une fois un problème d'algèbre linéaire que l'on résout simplement. En appliquant alors la loi de feedback (1.8) obtenue pour $\tilde{\delta x}$, on obtiendra une loi de feedback qui stabilise (localement) le bras. Ce type de loi est appelée feedback dynamique puisque l'on stabilise le bras à l'aide de $\tilde{\delta x}$ qui est obtenu, à partir de la sortie $\delta\theta$, de manière dynamique (c'est-à-dire via une équation différentielle). On remarquera aussi que l'action qui consiste à estimer δx est découplée de celle qui consiste à le stabiliser : les choix de L (estimation) et de k (stabilisation) sont indépendants l'un de l'autre. C'est le principe de séparation.

1.6 Commande optimale

Une fois la question de commandabilité comprise, on a vu que la détermination d'une loi de commande est effectuée en fonction de l'objectif de commande que l'on s'impose. Celui-ci peut être un but de stabilisation comme on l'a vu précédemment. On peut aussi vouloir minimiser la commande nécessaire à la réalisation de l'objectif fixé. Par exemple, pour le bras de robot décrit par (1.1), on peut vouloir minimiser le travail de la force qui est nécessaire pour amener le système (1.1) d'un état x_0 à un autre x_1 en temps $T = 1$. Cela signifie qu'il faut minimiser $\int_0^1 |u|$ parmi toutes les lois de commandes qui amènent le système (1.1) de x_0 à x_1 . De même, un autre type de minimisation est celui du temps lorsque l'amplitude de la commande est uniformément bornée : supposons que la commande u prenne ses valeurs dans $[-1, 1]$. Pour tous paires d'états x_0 et x_1 , il s'agit de minimiser le temps nécessaire pour amener le système (1.1) de x_0 à x_1 .

Ainsi, lorsque l'on cherche à "optimiser" la commande, il faut se donner un critère d'optimisation qui sera appelé le coût. Le but est alors de montrer qu'il existe (ou non) une commande minimisante, appelée aussi commande optimale, et surtout de caractériser cette (ou ces) commande(s) optimale(s). Il existe plusieurs manières d'attaquer les questions de commande optimale, suivant le type de dynamiques que l'on considère. Notons d'ores et déjà que ces méthodes peuvent être toutes regroupées sous un principe général, dit de Pontryagin.

1.7 Plan du cours

On va maintenant reprendre de manière plus systématique et rigoureuse les divers points évoqués dans l'exemple ci-dessus.

Le chapitre 2 est consacré aux équations différentielles ordinaires (EDO) autonomes en dimension finie. On définira les notions de champ de vecteurs, trajectoires et on présentera quelques résultats fondamentaux qui s'y rattachent : problème de Cauchy (existence et unicité de solutions), stabilité au sens de Lyapunov pour des équilibres, etc... Le cas des EDO linéaires à coefficients constants sera soigneusement analysé.

Dans le chapitre 3, est entamée l'étude des systèmes linéaires stationnaires. Dans un premier temps, on mettra l'accent sur la forme de Brunovsky et son application à la planification de trajectoire ainsi que sur la stabilisation par placement de pôles. Nous aborderons ensuite l'observabilité comme problème dual de la commandabilité, la construction d'observateurs asymptotiques (ou de Luenberger) et la synthèse d'un bouclage dynamique de sortie (on dit aussi observateur-contrôleur). Le chapitre 4 constitue une brève introduction à la commandabilité des systèmes non linéaire.

Les chapitres 5 et 6 font l'objet de (a) l'étude détaillée de la commande LQ avec son application au filtre de Kalman ; (b) la minimisation du temps pour les systèmes linéaires ; (c) le principe du maximum de Pontryagin pour les systèmes affines en la commande ; (d) l'introduction à la synthèse optimale en dimension deux.

Il faudra consulter [Rou-Bo] pour les nombreux exemples représentatifs de questions préoccupant les ingénieurs ainsi que pour les exercices qui permettent d'assimiler le cours.

Les parties écrites en petits caractères peuvent être ignorées sauf mention spéciale : il s'agit souvent de démonstrations.

Chapitre 2

Equations différentielles ordinaires et stabilité

2.1 Théorie générale des équations différentielles

Dans cette section, nous présentons la théorie générale des équations différentielles *autonomes*, qui sont de la forme

$$x'(t) = f(x(t)). \quad (2.1)$$

Cette théorie permet de modéliser et d'étudier de nombreux processus d'évolution déterministes, finis et *différentiables*.

Dans la formulation (2.1), les *données* sont :

- un ensemble ouvert $\Omega \subset \mathbb{R}^n$; x et Ω sont respectivement appelés l'état du système et *l'espace d'état* du système : à chaque instant le système est caractérisé par la donnée de x qui vit dans Ω .
- une application continue $f : \Omega \rightarrow \mathbb{R}^n$, (parfois appelée "membre de droite de l'équation différentielle").

(Les résultats que nous allons présenter restent valables quand on remplace \mathbb{R}^n par n'importe quel espace vectoriel de dimension finie, par exemple \mathbb{C}^n , $M_n(\mathbb{R})$, ...) Une telle application $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ est appelée un *champ de vecteurs* : à tout point x dans Ω , elle associe un vecteur $f(x)$ dans \mathbb{R}^n . En mécanique, f est aussi appelé *champ de vitesse*.

Exemple. En revenant à l'exemple du bras rigide (cf. chapitre précédent), la réponse libre correspond à l'EDO $x'(t) = f(x(t))$ avec $\Omega = S^1 \times \mathbb{R}$ et le champ de vecteur $f : S^1 \times \mathbb{R} \rightarrow \mathbb{R}^2$ défini par $f(x) = (w, -\sin \theta)^T$ si $x = (\theta, w)$.

Une *solution* de l'équation différentielle est une fonction dérivable $x(\cdot) : I \rightarrow \mathbb{R}^n$ telle que :

- I est un intervalle de \mathbb{R} ;
- $x(\cdot)$ prend ses valeurs dans Ω , *i.e.* $x(I) \subset \Omega$;
- pour tout $t \in I$, $x'(t) = f(x(t))$.

Une solution est donc en fait un couple $(x(\cdot), I)$: l'intervalle de définition I fait partie des inconnues. Nous verrons comment caractériser cet intervalle dans la section 2.1.2.

Notons enfin que, comme l'application f est supposée continue, toute solution $x(\cdot)$ de l'équation différentielle est automatiquement de classe C^1 .

Remarque. Il peut sembler réducteur de ne considérer que les équations différentielles autonomes, alors que le cadre le plus général est celui des équations de la forme

$$x'(t) = f(t, x(t)), \quad t \in J \subset \mathbb{R}, \quad (2.2)$$

qui dépendent explicitement du temps, et qui sont dites *non-autonomes*. Ce n'est en fait pas vraiment une restriction : toute équation non-autonome dans \mathbb{R}^n peut être vue comme une équation autonome dans \mathbb{R}^{n+1} . En effet, définissons un champ de vecteur $F : J \times \Omega \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$ par $F(t, x) = (1, f(t, x))$. Il est alors clair que l'équation non-autonome (2.2) est équivalente à l'équation autonome

$$\begin{pmatrix} t \\ x \end{pmatrix}' = \begin{pmatrix} 1 \\ f(t, x) \end{pmatrix} = F(t, x(t)).$$

2.1.1 Existence et unicité

Toute équation différentielle n'a pas de solution. Pour s'en convaincre, considérons l'EDO définie sur \mathbb{R} par

$$\begin{cases} x'(t) = -\text{sign}(x(t)), \\ x(t_0) = 0, \end{cases}$$

avec $\text{sign}(x) = x/|x|$ si x est non nul et $\text{sign}(0) = 0$. Le lecteur essaiera de donner un argument (simple) montrant qu'il n'y a pas de solution pour l'EDO précédente sur n'importe quel voisinage de 0. A la lumière de cet exemple, il est nécessaire d'effectuer une hypothèse de régularité sur le membre de droite d'une EDO pour espérer avoir un "bout" de solution dans un voisinage ouvert du temps de départ.

Définition 2.1. On appelle problème de Cauchy, le système

$$\begin{cases} x'(t) = f(x(t)), \\ x(t_0) = x_0, \end{cases} \quad (2.3)$$

c'est-à-dire le système formé d'une EDO et d'une condition initiale (valeur de l'état donnée à un instant fixé). Rappelons que l'un des objectifs des EDO est de modéliser des processus physiques qui sont souvent déterministes : si on connaît la dynamique d'un système et une condition initiale à $t = t_0$, alors l'évolution de ce système est unique pour $t \geq t_0$. Cette notion de déterminisme se traduit en termes mathématiques par le fait que tout problème de Cauchy a une solution et une seule pour $t \geq 0$. Avoir unicité des solutions d'une EDO est donc une nécessité pour un modèle réaliste.

Le propos du théorème ci-dessous est de répondre aux questions précédentes.

Théorème 2.1 (Théorème de Cauchy-Lipschitz). *Supposons f de classe C^1 sur Ω . Alors, pour tout point $x_0 \in \Omega$ et tout $t_0 \in \mathbb{R}$, il existe $\delta > 0$ tel que le problème de Cauchy défini en (refeq :syscl) possède une unique solution définie sur $]t_0 - \delta, t_0 + \delta[$.*

*PREUVE.

▷ La démonstration de ce théorème repose sur le théorème du point fixe de Picard. Fixons un réel $\alpha > 0$ tel que la boule fermée $\overline{B}(x_0, \alpha)$ soit contenue dans Ω . Puisque f est C^1 , il existe des constantes M et $K > 0$ telles que, sur $\overline{B}(x_0, \alpha)$, f est bornée en norme par M et est K -lipschitzienne (pourquoi?) Posons en outre

$$\delta = \min\left(\frac{\alpha}{M}, \frac{1}{2K}\right).$$

▷ Définissons \mathcal{E} comme étant l'ensemble des fonctions $x(\cdot)$ continues sur $]t_0 - \delta, t_0 + \delta[$ à valeurs dans $\overline{B}(x_0, \alpha)$ et telles que $x(t_0) = x_0$. Muni de la norme de la convergence uniforme $\|\cdot\|_0$, c'est un espace complet. L'application

$$\Phi(x(\cdot)) = x_0 + \int_{t_0}^{\cdot} f(x(s)) ds,$$

est une application de \mathcal{E} dans lui-même : en effet, pour $|t - t_0| \leq \delta$,

$$\|\Phi(x(t)) - x_0\| = \left\| \int_{t_0}^t f(x(s)) ds \right\| \leq \delta M \leq \alpha.$$

Cette application est en outre $\frac{1}{2}$ -lipschitzienne puisque, pour $t \in]t_0 - \delta, t_0 + \delta[$,

$$\begin{aligned} \|\Phi(x(\cdot)) - \Phi(y(\cdot))\|_0 &\leq \sup_{|t-t_0| < \delta} \left(\int_{t_0}^t \|f(x(s)) - f(y(s))\| ds \right) \\ &\leq \sup_{|t-t_0| < \delta} \left(\int_{t_0}^t K \|x(s) - y(s)\| ds \right) \\ &\leq \delta K \|x(\cdot) - y(\cdot)\|_0 \leq \frac{1}{2} \|x(\cdot) - y(\cdot)\|_0. \end{aligned}$$

Le théorème du point fixe de Picard s'applique et montre que l'application Φ admet un unique point fixe dans \mathcal{E} , c'est-à-dire que le système (2.3) admet une unique solution $x(\cdot) :]t_0 - \delta, t_0 + \delta[\rightarrow \mathbb{R}^n$ à valeurs dans la boule $\overline{B}(x_0, \alpha)$.

▷ Il ne reste plus qu'à montrer que toute solution $x(\cdot) :]t_0 - \delta, t_0 + \delta[\rightarrow \mathbb{R}^n$ du système (2.3) est à valeurs dans la boule $\overline{B}(x_0, \alpha)$. Par l'absurde, supposons qu'une solution $x(\cdot)$ de (2.3) sorte de $\overline{B}(x_0, \alpha)$ en temps inférieur à δ , et notons t_1 le premier instant où $x(\cdot)$ sort de la boule ouverte $B(x_0, \alpha)$. D'après le théorème des accroissements finis,

$$\alpha = \|x(t_1) - x_0\| \leq \left(\sup_{t \in [t_0, t_1]} \|x'(t)\| \right) |t_1 - t_0| < M\delta,$$

ce qui contredit $\delta \leq \alpha/M$. Toute solution de (2.3) sur $]t_0 - \delta, t_0 + \delta[$ est donc à valeurs dans $\overline{B}(x_0, \alpha)$, ce qui montre le théorème.

□

Hypothèses plus faibles sur f

Nous avons énoncé le théorème de Cauchy-Lipschitz avec l'hypothèse que f est C^1 sur Ω car elle est simple à utiliser et fréquemment satisfaite dans les applications. Remarquons cependant que, dans la preuve, nous avons seulement besoin que f soit *localement lipschitzienne*, c'est-à-dire que pour tout $x_0 \in \Omega$, il existe un voisinage U_0 de x_0 dans Ω et une constante K tels que f soit K -lipschitzienne sur U_0 . La conclusion du théorème de Cauchy-Lipschitz reste donc valable sous l'hypothèse que f est localement lipschitzienne. En particulier, elle est valable si f est (globalement) lipschitzienne sur Ω .

Que se passe-t-il si on affaiblit encore les hypothèses et que l'on suppose f seulement continue ? Un théorème de Peano affirme que, dans ce cas, le système (2.3) admet toujours une solution. En revanche, l'unicité n'est pas garantie. Par exemple le problème de Cauchy

$$\begin{cases} y'(t) &= \sqrt{|y(t)|} \\ y(0) &= 0 \end{cases}, \quad y \in \mathbb{R}$$

admet comme solutions les fonctions $y_1(t) = 0$ et $y_2(t) = \frac{t|t|}{4}$. Il en admet même une infinité puisque, pour tout $a \geq 0$, la fonction $y^a(\cdot)$ définie par

$$y^a(t) = 0 \quad \text{pour } t \leq a, \quad y^a(t) = y_2(t - a) \quad \text{pour } t > a$$

est également solution.

Remarque. D'après la remarque faite en introduction, le théorème de Cauchy-Lipschitz est également valable pour une équation non-autonome $x' = f(t, x)$: si f est C^1 sur $J \times \Omega$, alors, pour $(t_0, x_0) \in J \times \Omega$, l'équation a une unique solution définie sur $]t_0 - \delta, t_0 + \delta[$ et valant x_0 en t_0 .

On peut dans ce cas affaiblir nettement les hypothèses sur f : en effet, la conclusion du théorème restera valable si on suppose seulement que f est *localement lipschitzienne en la seconde variable* x , c'est-à-dire que, pour tout $(t_0, x_0) \in J \times \Omega$, il existe un voisinage J_0 de t_0 dans J , un voisinage U_0 de x_0 dans Ω et une constante K tels que, pour tout $t \in J_0$, l'application $f(t, \cdot)$ est K -lipschitzienne sur U_0 . La preuve est une simple adaptation de celle que nous avons donnée ici.

2.1.2 Solutions maximales et durée de vie

Considérons l'équation différentielle

$$x'(t) = f(x(t)), \tag{2.4}$$

où le champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ est supposé de classe C^1 .

Nous avons défini au début de ce chapitre une solution de cette équation comme une fonction $x(\cdot)$ définie sur un certain intervalle I de \mathbb{R} . Cette section est consacrée à l'étude de cet intervalle de définition I . Nous rencontrerons deux types de problèmes.

- Étant donné un couple $(t_0, x_0) \in \mathbb{R} \times \Omega$, il existe une infinité de solutions de (2.4) satisfaisant la condition initiale $x(t_0) = x_0$: par exemple si $x(\cdot)$ est solution sur l'intervalle I , $t_0 \in I$, toute restriction de $x(\cdot)$ à un sous-intervalle de I contenant t_0 est une solution différente. Pour éviter de considérer comme solutions différentes la même fonction prise sur des sous-intervalles, nous chercherons à associer à une fonction un unique intervalle, le plus grand, sur lequel elle est solution : c'est la notion de *solution maximale*.
- Si on choisit l'intervalle I le plus grand possible, peut-on le prendre égal à \mathbb{R} tout entier ? Si ce n'est pas le cas, que se passe-t-il pour la solution ? et quelle est la forme de I ? C'est le problème de la *durée de vie* des solutions.

Solutions maximales

Définition 2.2. On dit qu'une solution $x(\cdot) : I \rightarrow \Omega$ de (2.4) est une *solution maximale* si elle n'a pas de prolongement à un intervalle strictement plus grand, c'est-à-dire si elle n'est pas la restriction à I d'une solution définie sur un intervalle $I' \supsetneq I$.

Nous allons montrer qu'il existe une unique solution maximale de l'équation (2.4) satisfaisant une condition initiale donnée. Nous avons besoin pour cela d'un résultat d'unicité global.

Proposition 2.2. Si $x(\cdot)$ et $y(\cdot) : I \rightarrow \Omega$ sont deux solutions de (2.4) définies sur le même intervalle I qui coïncident en un point $t_0 \in I$, alors elles sont égales.

PREUVE.

▷ Considérons d'abord le dernier instant $> t_0$ où les solutions coïncident :

$$t_+ = \sup\{t \in I : t > t_0, x(s) = y(s) \text{ pour tout } s \in [t_0, t]\}.$$

Par l'absurde, supposons $t_+ < \sup I$. Les solutions étant continues, on a $x(t_+) = y(t_+)$. En appliquant le théorème de Cauchy-Lipschitz au couple $(t_+, x(t_+))$, on obtient que les deux solutions sont encore égales sur un intervalle $[t_+, t_+ + \delta]$, ce qui contredit la définition même de t_+ . Donc $t_+ = \sup I$. Le même argument vaut pour l'infimum des instants où les deux solutions coïncident. □

Théorème 2.3. Pour toute donnée initiale $(t_0, x_0) \in \mathbb{R} \times \Omega$, il existe une unique solution maximale $x(\cdot) :]t_-, t_+[\rightarrow \Omega$ de (2.4) satisfaisant $x(t_0) = x_0$. Tout autre solution satisfaisant cette condition initiale est une restriction de $x(\cdot)$ à un sous-intervalle de $]t_-, t_+[$.

Remarque. Insistons sur le fait que l'intervalle de définition d'une solution maximale est toujours un intervalle ouvert $]t_-, t_+[$. Les bornes t_+ et t_- de l'intervalle maximal sont des fonctions de (t_0, x_0) qui prennent leurs valeurs dans $\overline{\mathbb{R}}$: t_+ peut être soit un réel, soit $+\infty$, alors que t_- peut être soit un réel soit $-\infty$. Dans tous les cas, $t_- < t_0 < t_+$.

PREUVE.

▷ Soit I la réunion de tous les intervalles contenant t_0 sur lesquels le système

$$\begin{cases} x'(t) = f(x(t)), \\ x(t_0) = x_0, \end{cases} \quad (2.3)$$

admet une solution. D'après le théorème de Cauchy-Lipschitz, cette réunion est un intervalle ouvert, c'est-à-dire de la forme $I =]t_-, t_+[$. Pour tout $t \in]t_-, t_+[$, définissons $x(t)$ comme la valeur en t de n'importe quelle solution de (2.3) définie sur $[t_0, t]$. La proposition précédente montre que la fonction $x(\cdot) :]t_-, t_+[\rightarrow \Omega$ ainsi définie est bien solution de (2.3). De plus, par construction, c'est un prolongement de toute autre solution. □

Durée de vie

On s'intéresse maintenant à l'intervalle de définition $]t_-, t_+[$ d'une solution maximale $x(\cdot)$ de (2.4). Cet intervalle peut être différent de \mathbb{R} , même pour les équations les plus simples.

Exemple. Considérons l'équation $y'(t) = y^2(t)$ dans \mathbb{R} , dont la solution valant y_0 en t_0 est

$$y(t) = \frac{y_0}{(t - t_0)y_0 + 1}.$$

L'intervalle maximal de définition de cette solution est $]t_0 - \frac{1}{y_0}, +\infty[$ si $y_0 > 0$, $] -\infty, t_0 - \frac{1}{y_0}[$ si $y_0 < 0$, et \mathbb{R} tout entier si $y_0 = 0$.

L'idée générale est que, si une solution ne peut être prolongée sur tout \mathbb{R} , c'est qu'elle s'approche en temps fini du bord de l'ensemble Ω . Formalisons cette idée pour la borne supérieure t_+ de l'intervalle (les résultats pour t_- sont similaires).

Proposition 2.4. *Soit $x(\cdot) :]t_-, t_+[\rightarrow \Omega$ une solution maximale de (2.4). Alors, si $t_+ < +\infty$, $x(t)$ sort définitivement de tout compact contenu dans Ω quand $t \rightarrow t_+$.*

*PREUVE.

▷ À faire. □

On rencontrera le cas $t_+ < +\infty$ essentiellement dans les deux situations suivantes :

- quand $\Omega = \mathbb{R}^n$ et $\lim_{t \rightarrow t_+} \|x(t)\| = +\infty$: c'est le phénomène *d'explosion en temps fini* dont nous avons donné un exemple ci-dessus ;
 - quand le bord de Ω est borné et $x(t)$ converge vers un point du bord quand $t \rightarrow t_+$.
- Inversement, retenons une condition suffisante pour que $]t_-, t_+[= \mathbb{R}$.

Corollaire 2.5. *Si toutes les valeurs d'une solution maximale $x(\cdot)$ sont contenues dans un compact inclus dans Ω , alors $x(\cdot)$ est définie sur tout \mathbb{R} .*

Champs de vecteurs complets

Définition 2.3. On dit que le champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ est *complet* (ou que l'équation associée est *complète*) si toute solution maximale est définie sur \mathbb{R} tout entier.

D'après le corollaire précédent, si toutes les solutions maximales sont contenues dans des compacts, le champ est complet.

Exemple. Si $\Omega = \mathbb{R}^n$, les champs linéaires $f(x) = Ax$ sont complets et, de façon plus générale, tous les champs admettant une majoration linéaire $\|f(x)\| \leq \alpha\|x\| + \beta$, avec $\alpha, \beta \geq 0$, sont complets (donc en particulier les champs bornés). C'est une conséquence du lemme de Gronwall.

2.1.3 Flots, portraits de phase

Considérons à nouveau une équation différentielle autonome

$$x'(t) = f(x(t)), \quad (2.4)$$

où le champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ est supposé de classe C^1 . Une des spécificités de cette équation est qu'elle ne dépend pas explicitement du temps (d'où le qualificatif autonome). En particulier, les solutions sont invariantes par translation du temps : si $x(\cdot)$ est solution, $x(t_0 + \cdot)$ aussi.

Proposition 2.6. *Soit $x(\cdot) :]t_-, t_+[\rightarrow \Omega$ une solution maximale de (2.4) et $t_0 \in \mathbb{R}$. Alors $\bar{x} : t \mapsto x(t + t_0)$, définie sur $]t_- - t_0, t_+ - t_0[$, est également une solution maximale de (2.4).*

Ainsi le temps n'a pas de rôle intrinsèque ici et on pourra se limiter aux données initiales en $t = 0$. Pour un point $x \in \Omega$, notons $\phi(\cdot, x)$ la solution maximale de (2.4) valant x en $t = 0$ et $I_x =]t_-, t_+[$ son intervalle de définition. Autrement dit, $\phi(\cdot, x)$ est la solution du système

$$\begin{cases} \frac{\partial}{\partial t} \phi(t, x) = f(\phi(t, x)) \\ \phi(0, x) = x \end{cases}, \quad \forall t \in I_x.$$

Définition 2.4. L'application $(t, x) \mapsto \phi(t, x)$ est appelée le *flot* du champ de vecteurs f (ou de l'équation $x' = f(x)$).

Par définition, l'application partielle à x fixé, $t \mapsto \phi(t, x)$, est une solution maximale de l'équation. Pour une étude qualitative de l'équation différentielle, il est important d'étudier plutôt l'autre application partielle, $\phi_t : x \mapsto \phi(t, x)$, pour t fixé. De façon imagée, $\phi_t(x)$ est la position à l'instant t d'un corps transporté par l'équation différentielle qui se trouvait à la position x en $t = 0$.

Exemple. Si f est linéaire, i.e. $f(x) = Ax$, $A \in M_n(\mathbb{R})$, le flot est donné par l'exponentielle de A :

$$\phi_t(x) = e^{tA}x, \quad \forall (t, x) \in \mathbb{R} \times \mathbb{R}^n.$$

Ainsi le flot est une généralisation de l'exponentielle de matrice. Il possède des propriétés similaires.

Proposition 2.7 (formule du flot). Si $t_1 \in I_x$ et $t_2 \in I_{\phi_{t_1}(x)}$, alors $t_1 + t_2 \in I_x$ et

$$\phi_{t_1+t_2}(x) = \phi_{t_2}(\phi_{t_1}(x)).$$

En particulier, si $t \in I_x$,

$$\phi_{-t}(\phi_t(x)) = x$$

PREUVE.

▷ D'après la proposition sur l'invariance par translation du temps, $t \mapsto \phi_{t_1+t}(x)$ est la solution maximale valant $\phi_{t_1}(x)$ en $t = 0$, ce qui est la définition de $t \mapsto \phi_t(\phi_{t_1}(x))$. \square

Remarque. La formule du flot peut aussi se lire de la façon suivante : si $x(\cdot)$ est une solution de (2.4), alors

$$x(t) = \phi_{t-t_0}(x(t_0))$$

pour tous t_0 et t dans l'intervalle de définition de $x(\cdot)$.

Le domaine de définition du flot est l'ensemble

$$\mathcal{D} = \{(t, x) \in \mathbb{R} \times \Omega : t \in I_x\}.$$

Pour obtenir des propriétés intéressantes sur le flot (continuité, différentiabilité), il est nécessaire de montrer d'abord que son domaine de définition \mathcal{D} est un ouvert de $\mathbb{R} \times \Omega$. Nous le verrons dans la section suivante.

Il y a cependant déjà un cas où cela est évident : si f est un champ de vecteurs complet sur Ω , c'est-à-dire si $I_x = \mathbb{R}$ pour tout $x \in \Omega$, le domaine de définition du flot est $\mathcal{D} = \mathbb{R} \times \Omega$. On peut alors réécrire les propriétés du flot de façon globale : pour tous $t, s \in \mathbb{R}$,

1. $\phi_t \circ \phi_s = \phi_{t+s}$;
2. $\phi_{-t} \circ \phi_t = \text{id}$;
3. $\phi_0 = \text{id}$;
4. $\frac{\partial}{\partial t} \phi_t = f \circ \phi_t$.

Les trois premières propriétés montrent en particulier que ϕ_t obéit à une loi de groupe.

Orbites et portraits de phase

Définition 2.5. On appelle *orbite* d'un point $x_0 \in \Omega$ (ou trajectoire passant par x_0) l'ensemble

$$\mathcal{O}_{x_0} = \{\phi_t(x_0) : t \in I_{x_0}\}.$$

Autrement dit, l'orbite de x_0 est la courbe tracée sur \mathbb{R}^n par la solution maximale de l'équation (2.4) passant par x_0 en $t = 0$.

La propriété d'invariance par translation du temps implique que, pour tout point $x \in \mathcal{O}_{x_0}$, on a $\mathcal{O}_x = \mathcal{O}_{x_0}$. En effet, dans ce cas, il existe un instant t_0 tel que $x = \phi_{t_0}(x_0)$. Tout point y de \mathcal{O}_x s'écrit alors $y = \phi_t(x) = \phi_{t+t_0}(x_0)$, c'est-à-dire $y \in \mathcal{O}_{x_0}$. En particulier, ceci implique que *deux orbites distinctes ne peuvent pas se croiser*. Chaque point de Ω appartient donc à une et une seule orbite.

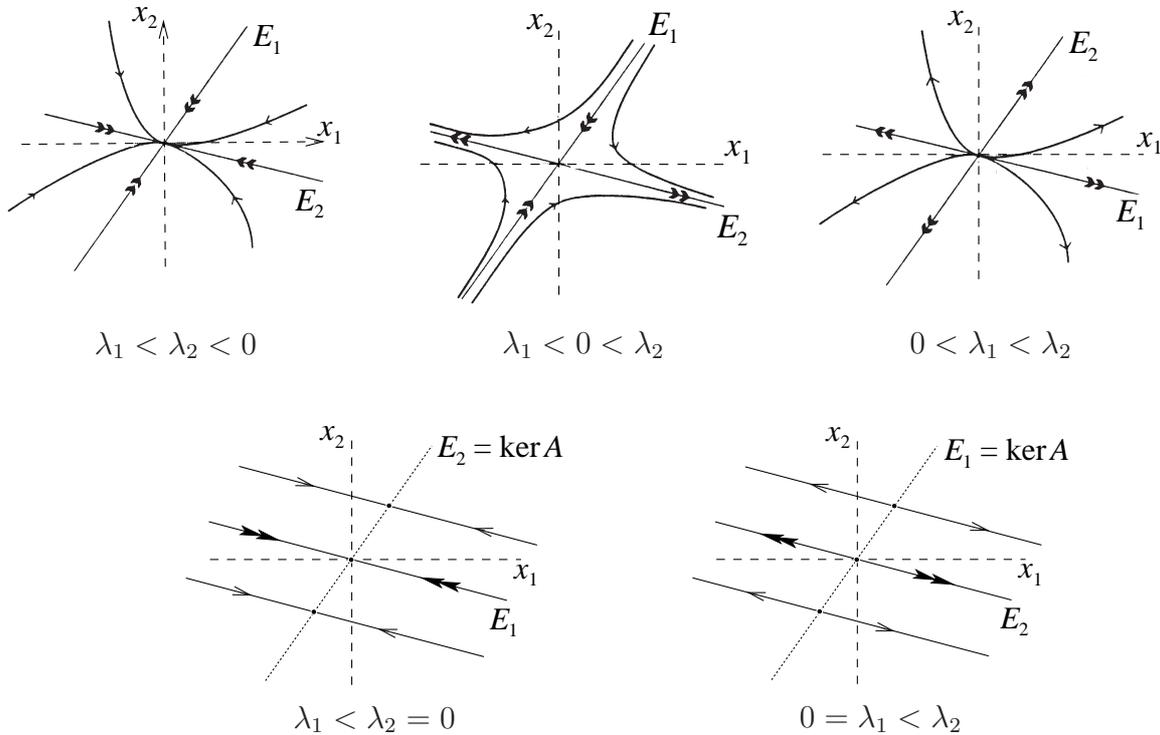
La partition de Ω en orbites s'appelle le *portrait de phase* du champ de vecteurs. On y trouve trois sortes d'orbites :

- des points, *i.e.* $\mathcal{O}_{x_0} = \{x_0\}$: un tel point vérifie nécessairement $f(x_0) = 0$. C'est ce que l'on appelle un *point d'équilibre* (voir Définition 2.12). Remarquer qu'un point d'équilibre correspond à un point fixe de ϕ_t pour tout t : $\phi_t(x_0) = x_0$.
- des courbes fermées : il existe alors un point x dans l'orbite et un temps $T > 0$ tels que $\phi_T(x) = x$. Ceci implique $\phi_{t+T}(x) = \phi_t(x)$ pour tout $t \in \mathbb{R}$, c'est-à-dire que la solution maximale $\phi(\cdot, x)$ est T -périodique. On parlera dans ce cas d'*orbite périodique*.
- des courbes ouvertes : il n'y a alors aucun point double, *i.e.* si $t \neq s$, $\phi_t(x) \neq \phi_s(x)$.

On porte habituellement sur le dessin d'un portrait de phase le sens de parcours des orbites.

Exemple. Considérons le champ de vecteurs linéaire $f(x) = Ax$ dans \mathbb{R}^2 , et supposons que la matrice $A \in M_2(\mathbb{R})$ a deux valeurs propres réelles et distinctes $\lambda_1 < \lambda_2$. L'étude réalisée dans la section 2.2.4 permet de déterminer la forme du portrait de phase en fonction de λ_1 et λ_2 . Nous avons représenté les différentes possibilités dans la figure 2.1, où nous avons noté E_1 et E_2 les sous-espaces propres associés à λ_1 et λ_2 .

Les points d'équilibre et les orbites périodiques sont des exemples de sous-ensembles invariants dont la définition est donnée ci-dessous.

FIGURE 2.1 – Exemples de portraits de phase pour $f(x) = Ax$ dans \mathbb{R}^2 .

Définition 2.6. Soit A un sous-ensemble de l'espace d'état Ω . On dit que A est *invariant* (respectivement *positivement invariant*) par le flot ϕ_t si, pour tout $t \in \mathbb{R}$ (respectivement dans \mathbb{R}_+), $\phi_t(A)$ est inclus dans A .

D'autres exemples d'ensembles invariants sont fournis par les hypersurfaces de niveau d'une fonction réelle de l'espace d'état qui reste constante le long des trajectoires, i.e. une intégrale première.

Définition 2.7. On appelle *intégrale première* d'une EDO $\dot{x} = f(x)$, une fonction $h : \Omega \rightarrow \mathbb{R}$, de classe C^1 , qui reste constante le long des trajectoires de l'EDO. Cela est vrai en particulier si, pour tout $x \in \Omega$ et t , $\frac{d}{dt}(h(\phi_t(x)))$, condition qui est équivalente à

$$D_x h(x) \cdot f(x) = 0, \quad \text{pour tout } x \in \Omega.$$

(Noter que cette dernière condition ne demande pas une connaissance explicite du flot.) Ainsi, les hypersurfaces de niveau $H_c := \{x \in \Omega : h(x) = c\}$, $c \in \mathbb{R}$ sont invariantes par le flot.

Exemple. Montrer que (1.5) admet comme intégrale première $h := 1/2\omega^2 - \cos\theta$. En déduire l'équation des orbites. Dessiner l'allure du portrait de phases sur le cylindre $S^1 \times \mathbb{R}$ suivant la valeur (constante) de h . Afin d'avoir une représentation plane, on identifiera S^1 avec $\mathbb{R}/[0, 2\pi]$ et on montrera l'existence de trajectoires périodiques.

Pour d'autres exemples, voir [Rou-Bo] où est expliqué le fait que l'énergie totale (énergie cinétique plus énergie potentielle) d'un système mécanique holonome parfait (sans frottement) est une intégrale première du système.

2.1.4 Équations différentielles linéaires

On suppose que le membre de droite de (2.1) est linéaire par rapport à l'état x c'est-à-dire qu'il prend la forme

$$x'(t) = A(t)x(t), \quad t \in I. \quad (2.5)$$

Précisons les notations. Les *données* sont :

- un intervalle I de \mathbb{R} ;
- une application $A : I \rightarrow M_n(\mathbb{K})$ de classe C^k (k est un entier positif ou $k = \infty$) ; chaque valeur $A(t)$ est donc une matrice ($n \times n$) à coefficients dans le corps $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .

Une *solution* de (2.5) est une application dérivable $x : I \rightarrow \mathbb{K}^n$ telle que, pour tout $t \in I$, sa dérivée $x'(t) = \frac{dx}{dt}(t)$ vérifie $x'(t) = A(t)x(t)$. Noter qu'une solution est automatiquement de classe C^{k+1} .

Nous traiterons aussi le cas un peu plus général des équations différentielles *affines*,

$$x'(t) = A(t)x(t) + b(t), \quad t \in I, \quad (2.6)$$

la donnée $b(\cdot)$ étant une application de I dans \mathbb{K}^n de classe C^k . Nous verrons que l'étude de ces équations se déduit de celle des équations linéaires.

Remarque. Il est fréquent dans la littérature que l'expression « équation linéaire » soit utilisée pour les équations affines, les équations (2.5) étant alors appelées équations linéaires *homogènes*.

Existence et Unicité globales

Théorème 2.8 (Existence et Unicité globales). *Soient $t_0 \in I$ et $x_0 \in \mathbb{K}^n$. Il existe une unique solution $x(\cdot)$ de l'équation (2.6) satisfaisant à la condition initiale*

$$x(t_0) = x_0.$$

Insistons sur le fait que ce théorème garantit l'existence de $x(\cdot)$ sur *tout* l'intervalle I . Ce phénomène est propre aux équations linéaires (on comparera avec le théorème 2.1 de Cauchy-Lipschitz concernant les équations non-linéaires).

La preuve de ce théorème repose sur la remarque suivante : $x(\cdot)$ est solution de (2.6) avec $x(t_0) = x_0$ si et seulement si $x(\cdot)$ est continue et vérifie pour tout $t \in I$,

$$x(t) = x_0 + \int_{t_0}^t (A(s)x(s) + b(s)) ds, \quad (2.7)$$

autrement dit, si $x(\cdot)$ est un point fixe de l'application

$$x(\cdot) \mapsto x_0 + \int_{t_0}^{\cdot} (A(s)x(s) + b(s)) ds.$$

Ainsi le théorème 2.8 est un résultat de point fixe et nous utiliserons donc le théorème du point fixe de Picard pour le montrer.

*PREUVE.

▷ Supposons pour simplifier que I soit un intervalle compact de la forme $I = [a, b]$ et introduisons Φ l'application affine qui à toute fonction $x(\cdot) \in C^0(I, \mathbb{R}^n)$ associe la fonction

$$\Phi(x(\cdot)) = x_0 + \int_{t_0}^{\cdot} (A(s)x(s) + b(s)) ds,$$

qui est visiblement continue et à valeurs dans \mathbb{R}^n . Comme nous l'avons remarqué ci-dessus, il s'agit de montrer que Φ a un unique point fixe. Nous allons pour cela vérifier qu'un *itéré* de Φ est contractant et appliquer le théorème du point fixe de Picard à cet itéré. On vérifiera alors que les points fixes de Φ sont ceux de cet itéré.

▷ Pour $x(\cdot), y(\cdot) \in C^0(I, \mathbb{R}^n)$ et $t \in I$, on a

$$\begin{aligned} \|(\Phi \circ x)(t) - (\Phi \circ y)(t)\| &= \left\| \int_{t_0}^t (A(s)(x(s) - y(s)) ds \right\| \\ &\leq \|A\|_{C^0} \int_{t_0}^t \|x - y\| ds \leq |t - t_0| \|A\|_{C^0} \|x - y\|_{C^0}, \end{aligned}$$

où $\|f\|_{C^0} = \max_{t \in I} \|f(t)\|$ désigne la norme C^0 . On voit en particulier que

$$\|\Phi \circ x - \Phi \circ y\|_{C^0} \leq (b - a) \|A\|_{C^0} \|x - y\|_{C^0},$$

ce qui prouve que Φ est continue de $C^0(I, \mathbb{R}^n)$ dans lui-même.

▷ Par le même calcul, on a, en notant $\Phi^2 = \Phi \circ \Phi$,

$$\begin{aligned} \|(\Phi^2 \circ x)(t) - (\Phi^2 \circ y)(t)\| &\leq \|A\|_{C^0} \int_{t_0}^t \|(\Phi \circ x - \Phi \circ y)(s)\| ds \\ &\leq \|A\|_{C^0} \|A\|_{C^0} \|x - y\|_{C^0} \int_{t_0}^t |s - t_0| ds \leq \|A\|_{C^0}^2 \frac{1}{2} |t - t_0|^2 \|x - y\|_{C^0}, \end{aligned}$$

et on montre par récurrence que

$$\|(\Phi^N \circ x)(t) - (\Phi^N \circ y)(t)\| \leq \|A\|_{C^0}^N \frac{|t - t_0|^N}{N!} \|x - y\|_{C^0},$$

c'est-à-dire,

$$\|\Phi^N \circ x - \Phi^N \circ y\|_{C^0} \leq \frac{((b - a) \|A\|_{C^0})^N}{N!} \|x - y\|_{C^0}.$$

▷ Choisissons N suffisamment grand pour que $\frac{((b-a)\|A\|_{C^0})^N}{N!} \leq \frac{1}{2}$ (ce qui est possible puisque le membre de gauche de cette inégalité tend vers 0 avec N). L'application continue $\Phi^N : C^0(I, \mathbb{R}^n) \rightarrow C^0(I, \mathbb{R}^n)$ est alors $\frac{1}{2}$ -contractante et, d'après le théorème du point fixe de Picard, admet un unique point fixe $x(\cdot) \in C^0(I, \mathbb{R}^n)$.

Mais il n'est pas difficile de voir que les points fixes de Φ sont les points fixes de Φ^N : en effet si $x(\cdot)$ est point fixe de Φ on a évidemment

$$\Phi^N(x(\cdot)) = \Phi^{N-1} \circ \Phi(x(\cdot)) = \Phi^{N-1}(x(\cdot)),$$

et donc $\Phi^N(x(\cdot)) = x(\cdot)$; réciproquement, si $\Phi^N(x(\cdot)) = x(\cdot)$, alors

$$\Phi^N(\Phi(x(\cdot))) = \Phi^{N+1}(x(\cdot)) = \Phi(\Phi^N(x(\cdot))) = \Phi(x(\cdot)),$$

et donc $\Phi(x(\cdot))$ et $x(\cdot)$ sont points fixes de Φ^N qui est *contractante* et admet de ce fait un *unique* point fixe : ils sont donc égaux. Ceci prouve le théorème pour I de la forme $I = [a, b]$. \square

La résolvante

Revenons à l'étude des équations linéaires dans \mathbb{K}^n

$$x'(t) = A(t)x(t), \tag{2.5}$$

et notons \mathcal{E} l'ensemble des solutions de cette équation.

Proposition 2.9. *L'ensemble \mathcal{E} est un espace vectoriel de dimension n .*

PREUVE.

▷ Il est immédiat que \mathcal{E} est un \mathbb{K} -espace vectoriel. Introduisons alors $L_{t_0} : \mathbb{K}^n \rightarrow \mathcal{E}$ l'application qui à $x_0 \in \mathbb{K}^n$ associe la solution $x(\cdot)$ de (2.5) telle que $x(t_0) = x_0$. C'est clairement une application linéaire et il résulte directement de l'existence et de l'unicité des solutions que L_{t_0} est un isomorphisme de \mathbb{K}^n sur \mathcal{E} , ce qui prouve le résultat. \square

Définition 2.8. On appelle *résolvante* de l'équation (2.5) l'application $R_A(t, s) : \mathbb{K}^n \rightarrow \mathbb{K}^n$ qui à $x_0 \in \mathbb{K}^n$ associe $x(t)$, où $x(\cdot)$ est la solution de (2.5) qui satisfait $x(s) = x_0$.

Il résulte des théorèmes d'existence et d'unicité que la résolvante est linéaire et bijective; $R_A(t, s)$ est donc une matrice *inversible* de $M_n(\mathbb{K})$. Elle permet d'exprimer toute solution $x(\cdot)$ de l'équation (2.5) en fonction d'une condition initiale :

$$x(t) = R_A(t, t_0)x(t_0).$$

En particulier, dans le cas autonome, c'est-à-dire quand $A(\cdot) \equiv A$ est constante, la résolvante est l'exponentielle de A : $R_A(t, s) = e^{(t-s)A}$.

La résolvante peut également être caractérisée par une équation différentielle.

Proposition 2.10.

1. Pour tout $t_0 \in I$, $R_A(\cdot, t_0)$ est la solution de l'équation différentielle matricielle

$$\begin{cases} \frac{\partial}{\partial t} R_A(t, t_0) = A(t)R_A(t, t_0), \\ R_A(t_0, t_0) = I, \end{cases} \quad (2.9)$$

2. Pour tous t_0, t_1, t_2 dans I ,

$$R_A(t_2, t_0) = R_A(t_2, t_1) \times R_A(t_1, t_0).$$

3. Si $A(\cdot)$ est de classe C^k , l'application $t \mapsto R_A(t, t_0)$ est de classe C^{k+1} .

PREUVE.

▷ Le premier point résulte du fait que, pour tout $x_0 \in \mathbb{K}^n$,

$$\frac{\partial}{\partial t} (R_A(t, t_0)x_0) = A(t)(R_A(t, t_0)x_0).$$

Le deuxième point résulte de la définition même de la résolvante. Quant à la dernière assertion, elle résulte de la première. □

Remarques.

- L'équation différentielle satisfaite par la résolvante a la même forme que l'équation (2.5) mais a lieu dans $M_n(\mathbb{K})$ (au lieu de \mathbb{K}^n).
- Le point 2 (ou directement la définition) implique en particulier :

$$R_A(t, s)^{-1} = R_A(s, t).$$

- Les colonnes $x_1(t), \dots, x_n(t)$ de la résolvante $R_A(t, t_0)$ sont les valeurs à l'instant t des solutions valant e_1, \dots, e_n (les vecteurs de la base canonique) à $t = t_0$, puisque par définition $x_j(t) = R_A(t, t_0)e_j$. Les fonctions $x_1(\cdot), \dots, x_n(\cdot)$ forment donc une base de l'ensemble \mathcal{E} des solutions.

Comme nous venons de le voir, pour résoudre une équation différentielle linéaire, il suffit de savoir calculer la résolvante (de même que pour les équations autonomes il suffit de savoir calculer l'exponentielle). Malheureusement, en dehors du cas autonome, **il est très rare de pouvoir donner une expression explicite de la résolvante.** Nous allons voir en revanche que l'on peut obtenir des informations *qualitatives* sur les solutions de l'équation grâce à l'étude de la résolvante.

Quelques propriétés de la résolvante

Proposition 2.11. Soit $t_0 \in \mathbb{R}$. La fonction $\Delta(t) = \det R_A(t, t_0)$ vérifie l'équation différentielle

$$\begin{cases} \Delta'(t) = \operatorname{tr}(A(t)) \Delta(t) \\ \Delta(t_0) = 1 \end{cases},$$

ce qui implique

$$\det R_A(t, t_0) = \exp\left(\int_{t_0}^t \operatorname{tr}(A(u)) du\right).$$

PREUVE.

▷ Rappelons (voir TD1) que, si $A \in GL_n(\mathbb{K})$, $D \det(A) \cdot H = (\det A) \operatorname{tr}(A^{-1}H)$. On a donc

$$\begin{aligned} \Delta'(t) &= (\det R_A(t, t_0)) \operatorname{tr}(R_A^{-1}(t, t_0)R'_A(t, t_0)) \\ &= \Delta(t) \operatorname{tr}(A(t)). \end{aligned}$$

La conclusion de la proposition suit. □

Corollaire 2.12 (Liouville). Si pour tout $t \in \mathbb{R}$, $A(t)$ est de trace nulle, alors le déterminant de $R_A(t, s)$ est identiquement égal à 1.

Il résulte de ce corollaire que, si $A(t)$ est de trace nulle, l'équation différentielle (2.5) préserve les volumes. En effet, si Γ est un domaine de \mathbb{R}^n , notons Γ_t son transport de t_0 à t par l'équation (2.5), c'est-à-dire

$$\Gamma_t = \{x(t) : x(\cdot) \text{ solution de (2.5) t.q. } x(t_0) \in \Gamma\},$$

ou encore $\Gamma_t = R_A(t, t_0)\Gamma$. Alors, en utilisant la formule de changement de variable dans les intégrales multiples, on obtient

$$\operatorname{vol}(\Gamma_t) = |\det R_A(t, t_0)| \operatorname{vol}(\Gamma),$$

et donc $\operatorname{vol}(\Gamma_t) = \operatorname{vol}(\Gamma)$ si $\operatorname{tr} A(t) \equiv 0$.

La préservation du volume a une conséquence sur le comportement asymptotique des solutions : il est en effet impossible dans ce cas que toutes les solutions de (2.5) tendent vers 0 quand $t \rightarrow \pm\infty$ (de même qu'il est impossible que $\|x(t)\|$ tende vers l'infini pour toute solution $x(\cdot)$).

Une classe particulièrement intéressante de matrices de trace nulle est l'ensemble des matrices antisymétriques, qui interviennent fréquemment dans les problèmes issus de la physique.

Proposition 2.13. *Si, pour tout $t \in \mathbb{R}$, $A(t)$ est une matrice réelle antisymétrique, la résolvante $R_A(t, s)$ est une rotation pour tous $t, s \in \mathbb{R}$.*

Rappelons qu'une rotation est une matrice $R \in M_n(\mathbb{R})$ orthogonale (c'est-à-dire $R^T R = I$) et de déterminant 1.

PREUVE.

▷ On sait déjà d'après le corollaire précédent que $\det R_A(t, s) \equiv 1$. D'autre part

$$\begin{aligned} \frac{\partial}{\partial t} (R_A(t, s)^T R_A(t, s)) &= \frac{\partial}{\partial t} R_A(t, s)^T R_A(t, s) + R_A(t, s)^T \frac{\partial}{\partial t} R_A(t, s) \\ &= R_A(t, s)^T A(t)^T R_A(t, s) + R_A(t, s)^T A(t) R_A(t, s) \\ &= R_A(t, s)^T (A(t)^T + A(t)) R_A(t, s) = 0. \end{aligned}$$

Ainsi, $R_A(t, s)^T R_A(t, s)$ est constant. Comme $R_A(s, s) = I$, on a la conclusion. □

Une conséquence de ce résultat est que, si $A(t)$ est une matrice réelle antisymétrique pour tout t , l'équation différentielle (2.5) préserve la norme. En effet, si $x(\cdot)$ est une solution de l'équation,

$$\|x(t)\| = \|R_A(t, t_0)x(t_0)\| = \|x(t_0)\|,$$

puisque $R_A(t, t_0)$ est une rotation. En particulier, toute solution est bornée. En revanche il est impossible qu'une solution tende vers 0 (sauf si $x(\cdot) \equiv 0$ bien sûr...). Nous verrons à la section 2.3.1 que l'on dit alors que 0 est un équilibre stable mais non asymptotiquement stable.

2.1.5 Linéarisation et perturbation du flot

Dans la pratique, on n'a quasiment jamais une connaissance exacte des conditions initiales (ni de l'équation elle-même, en fait). Il est donc primordial de savoir ce qui se passe pour la solution d'une équation différentielle lorsque la condition initiale est perturbée (ou quand l'équation elle-même est perturbée) : comment varie l'intervalle de définition et les valeurs de la solution, peut-on donner un ordre de grandeur de ces variations... ? Les réponses à ces questions sont contenues dans le théorème ci-dessous : donnons d'abord le théorème et sa preuve, nous expliquerons ensuite pourquoi il permet de répondre aux questions précédentes.

Rappelons que nous considérons une équation différentielle autonome

$$x'(t) = f(x(t)), \quad (2.4)$$

où le champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ est supposé de classe C^1 .

Théorème 2.14. *Soit $\bar{x}(\cdot)$ une solution de l'équation (2.4) définie sur un intervalle $[a, b]$ contenant 0. Il existe alors un voisinage $\mathcal{V} \subset \Omega$ de $v_0 = \bar{x}(0)$ tel que, pour tout $v \in \mathcal{V}$, l'équation (2.4) admet une unique solution $x_v(\cdot)$ définie sur $[a, b]$ et vérifiant $x_v(0) = v$.*

De plus, l'application $v \mapsto x_v(\cdot)$ est de classe C^1 sur \mathcal{V} et sa différentielle en v_0 est l'application qui à Δv associe la solution de

$$\begin{cases} y'(t) = Df(\bar{x}(t)) \cdot y(t) \\ y(0) = \Delta v \end{cases}, \quad t \in [a, b].$$

Remarque. La solution $x_v(\cdot)$ est simplement la restriction de la solution maximale $\phi(\cdot, v)$ à l'intervalle $[a, b]$.

Conséquences et signification du théorème 2.14

a. Domaine de définition du flot La première conséquence est que, pour toute condition initiale v dans le voisinage \mathcal{V} de $\bar{x}(0)$, la solution maximale $\phi(\cdot, v)$ est définie sur tout l'intervalle $[a, b]$, *i.e.* $[a, b] \subset I_v$. Autrement dit, de façon informelle, si une solution est définie sur un temps « long », les solutions voisines sont également définies sur un temps « long ». Ceci se traduit par une propriété du domaine de définition du flot.

Corollaire 2.15. *Le flot ϕ est défini sur un ouvert \mathcal{D} de $\mathbb{R} \times \Omega$.*

En particulier, si $(t, v_0) \in \mathcal{D}$, alors l'application ϕ_t est définie sur un voisinage de v_0 .

Cette propriété est très importante pour l'étude du flot et de sa dépendance par rapport aux conditions initiales : en effet, ϕ et ϕ_t étant définies sur des ouverts, il est maintenant possible d'étudier leur continuité et leur différentiabilité.

PREUVE.

▷ Rappelons que le domaine de définition du flot est

$$\mathcal{D} = \{(t, v) \in \mathbb{R} \times \Omega : t \in I_v\}.$$

Soit $(t_0, v_0) \in \mathcal{D}$. Puisque l'intervalle maximal I_{v_0} est ouvert (théorème 2.3), la solution maximale $\phi(\cdot, v_0)$ est définie sur un intervalle $[a, b] \subset I_{v_0}$ contenant t_0 . Le théorème 2.14

implique alors que, pour tout v dans un voisinage \mathcal{V} de v_0 , on a encore $[a, b] \subset I_v$, c'est-à-dire que l'ensemble $]a, b[\times \mathcal{V}$, qui est un voisinage de (t_0, v_0) dans $\mathbb{R} \times \Omega$, est inclus dans \mathcal{D} .

□

b. Dépendance continue L'application $v \mapsto x_v(\cdot)$ définie dans le théorème 2.14 est l'application qui à une condition initiale dans \mathcal{V} associe la solution correspondante de l'équation différentielle sur $[a, b]$. Cette application étant C^1 , elle est en particulier continue, c'est-à-dire que

les solutions de l'équation différentielle (2.4) dépendent de façon continue de leur condition initiale.

C'est une propriété essentielle pour les applications (et d'un point de vue numérique) : en effet, elle signifie, grosso modo, que la solution calculée à partir d'une approximation de la condition initiale est une approximation de la vraie solution. Ceci justifie l'utilisation d'équations différentielles dans la modélisation de phénomènes réels, où on n'a qu'une connaissance approximative des données.

c. Équation linéarisée La dernière partie du théorème 2.14 affirme que les valeurs de la différentielle de l'application $\psi : v \mapsto x_v(\cdot)$ sont les solutions d'une certaine équation linéaire. Cette équation linéaire joue un rôle important dans la suite.

Définition 2.9. Soit $\bar{x}(\cdot) : [a, b] \rightarrow \Omega \subset \mathbb{R}^n$ une solution de (2.4). L'équation linéaire dans \mathbb{R}^n

$$y'(t) = Df(\bar{x}(t)) \cdot y(t), \quad t \in [a, b],$$

est appelée *équation linéarisée de (2.4) autour de $\bar{x}(\cdot)$* .

Pour tout $\delta v \in \mathbb{R}^n$, $D\psi(\bar{x}(0)) \cdot \delta v$ est la solution de l'équation linéarisée autour de $\bar{x}(\cdot)$ valant δv en $t = 0$. En notant $R(t, s)$ la résolvante de l'équation linéarisée autour de $\bar{x}(\cdot)$, on obtient, pour tout $t \in [a, b]$,

$$(D\psi(\bar{x}(0)) \cdot \delta v)(t) = R(t, 0)\delta v.$$

Intéressons-nous maintenant à l'application ϕ_t . Fixons un point v_0 de Ω et un instant $t \in I_{v_0}$. D'après le théorème 2.14, l'application ϕ_t est définie et de classe C^1 sur un voisinage \mathcal{V} de v_0 dans Ω . Avec les notations ci-dessus, on a clairement $\phi_t(v) = x_v(t) = (\psi(v))(t)$ et donc

$$D\phi_t(v_0) \cdot \delta v = (D\psi(v_0) \cdot \delta v)(t).$$

On en déduit le résultat suivant.

Corollaire 2.16. Soient $v_0 \in \Omega$ et $t \in I_{v_0}$. L'application ϕ_t est de classe C^1 sur un voisinage \mathcal{V} de v_0 et

$$D\phi_t(v_0) = R(t, 0),$$

où R est la résolvante de l'équation linéarisée

$$y'(s) = Df(\phi_s(v_0)) \cdot y(s), \quad s \in [0, t].$$

On peut donner une explication plus intuitive du rôle de l'équation linéarisée. On choisit une solution $\bar{x}(\cdot) : [a, b] \rightarrow \Omega$ de l'équation différentielle, de condition initiale $v_0 = \bar{x}(0)$. Considérons maintenant une perturbation $v_0 + \delta v$ de la condition initiale et écrivons la solution correspondante (*i.e.* $x_{v_0 + \delta v}(\cdot)$) sous la forme d'une perturbation $\bar{x}(\cdot) + \delta x(\cdot)$ de la solution d'origine. Cette perturbation étant une solution, elle doit satisfaire l'équation différentielle :

$$\bar{x}'(t) + (\delta x)'(t) = f(\bar{x}(t) + \delta x(t)), \quad t \in [a, b].$$

En utilisant un développement limité de f en $\bar{x}(t)$ (à t fixé) :

$$f(\bar{x}(t) + \delta x(t)) = f(\bar{x}(t)) + Df(\bar{x}(t)) \cdot \delta x(t) + \text{reste},$$

et en tenant compte du fait que $\bar{x}'(t) = f(\bar{x}(t))$, on obtient

$$(\delta x)'(t) = Df(\bar{x}(t)) \cdot \delta x(t) + \text{reste}.$$

En ne conservant que les « termes du premier ordre » on retrouve l'équation linéarisée $(\delta x)'(t) = Df(\bar{x}(t)) \cdot \delta x(t)$. Autrement dit, la solution perturbée s'écrit $\bar{x}(\cdot) + \delta x(\cdot) + \text{reste}$, où le « terme du premier ordre » $\delta x(\cdot)$ est la solution de

$$\begin{cases} (\delta x)'(t) = Df(\bar{x}(t)) \cdot (\delta x)(t) \\ (\delta x)(0) = \delta v \end{cases}.$$

L'équation linéarisée indique donc comment se propage au cours du temps une perturbation de la condition initiale.

Bien entendu ce qui précède n'est pas un raisonnement rigoureux (les restes posent évidemment quelques problèmes!), seulement une heuristique.

Remarque. L'équation linéarisée est en général une équation linéaire non-autonome, on ne sait donc pas a priori calculer ses solutions. Cependant, si v_0 est un point d'équilibre, la solution maximale $\bar{x}(\cdot) = \phi(\cdot, v_0)$ est la fonction constante $\bar{x}(\cdot) \equiv v_0$, définie sur tout \mathbb{R} , et dans ce cas l'équation linéarisée est autonome :

$$y'(t) = Df(v_0) \cdot y(t), \quad t \in \mathbb{R}.$$

Exemple d'application : champs de vecteurs à divergence nulle. Rappelons d'abord que la divergence d'un champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, $f(x) = (f_1(x), \dots, f_n(x))$, est définie comme

$$\operatorname{div} f(x) = \frac{\partial f_1}{\partial x_1}(x) + \dots + \frac{\partial f_n}{\partial x_n}(x) = \operatorname{tr} Df(x).$$

Considérons alors un temps t et un domaine Γ de \mathbb{R}^n , supposé inclus dans le domaine de définition de ϕ_t . Notons $\Gamma_t = \phi_t(\Gamma)$ le transport de Γ de 0 à t par l'équation (2.4). En utilisant la formule de changement de variable dans les intégrales multiples, on obtient

$$\operatorname{vol}(\Gamma_t) = \int_{\phi_t(\Gamma)} d\mu = \int_{\Gamma} |\det D\phi_t(x)| d\mu,$$

où, d'après le corollaire 2.16, $D\phi_t(x)$ est la résolvante de l'équation linéarisée.

Supposons maintenant que $\operatorname{div} f(x) = \operatorname{tr} Df(x) \equiv 0$. Le théorème de Liouville (corollaire 2.12) implique que le déterminant de la résolvante du linéarisé est égal à 1, et donc que $\det D\phi_t(x) = 1$. On a alors $\operatorname{vol}(\Gamma_t) = \operatorname{vol}(\Gamma)$, c'est-à-dire *si f est un champ de divergence nulle, le flot de f préserve le volume.*

Dépendance par rapport à un paramètre

Considérons maintenant une famille d'équations différentielles dépendant d'un paramètre $\lambda \in \mathbb{R}^p$:

$$x'(t) = f_\lambda(x(t)), \quad (2.10)$$

où chaque f_λ est un champ de vecteurs sur $\Omega \subset \mathbb{R}^n$. Supposons également que $f(x, \lambda) = f_\lambda(x)$ est une application de classe C^1 . On s'intéresse à la dépendance des solutions de ces équations différentielles par rapport au paramètre λ .

Remarquons d'abord que l'équation (2.10) est équivalente à

$$\begin{cases} x'(t) &= f(x(t), \lambda) \\ \lambda'(t) &= 0 \end{cases},$$

c'est-à-dire à l'équation différentielle dans \mathbb{R}^{n+p} associée au champ de vecteurs $F(x, \lambda) = (f(x, \lambda), 0)$. Ainsi, les solutions de (2.10) dépendent du paramètre λ de la même façon que les solutions de l'équation différentielle $(x, \lambda)'(t) = F((x, \lambda)(t))$ dépendent de leur condition initiale. D'après ce que nous avons vu précédemment dans cette section, nous avons donc les propriétés suivantes.

- Les solutions $\phi^\lambda(\cdot, v)$ de (2.10) dépendent de façon C^1 , donc continue, du paramètre λ (et de la condition initiale v);
- La différentielle de l'application $(v, \lambda) \mapsto \phi^\lambda(\cdot, v)$ en un point (v_0, λ_0) est l'application qui à $(\delta v, \delta \lambda)$ associe la solution $y(\cdot)$ de l'équation différentielle affine

$$\begin{cases} y'(t) &= D_x f(\bar{x}(t), \lambda_0) \cdot y(t) + D_\lambda f(\bar{x}(t), \lambda_0) \cdot \delta \lambda \\ y(0) &= \delta v \end{cases},$$

où $\bar{x}(\cdot) = \phi^{\lambda_0}(\cdot, v_0)$. Autrement dit, en utilisant la formule de variation de la constante,

$$y(t) = R(t, 0)\delta v + \int_0^t R(t, s)D_\lambda f(\bar{x}(s), \lambda_0) \cdot \delta \lambda ds,$$

où $R(t, s)$ est la résolvante associée au système linéarisé $y'(t) = D_x f(\bar{x}(t), \lambda_0) \cdot y(t)$.

***Application de Poincaré**

Supposons que l'équation (2.4) admette une solution T -périodique non triviale $x(\cdot)$ et notons $x(0) = x_0$. Traçons un hyperplan affine Σ passant par x_0 et *transverse* à $x(\cdot)$ en x_0 , c'est-à-dire que $f(x_0)$ n'est pas parallèle à Σ (notez que $f(x_0) \neq 0$ puisque $x(\cdot)$ est non triviale). Nous ferons pour simplifier l'hypothèse que Σ est en fait un hyperplan affine parallèle à \mathbb{R}^{n-1} et que $f(x_0) \in \mathbb{R}e_1$, où e_1 vérifie $\mathbb{R}^n = \mathbb{R}e_1 \oplus \mathbb{R}^{n-1}$: c'est une situation à laquelle on peut toujours se ramener par un changement linéaire de coordonnées.

Proposition 2.17. *Il existe un voisinage $\mathcal{V} \subset \mathbb{R}^n$ de x_0 et une fonction η de classe C^1 tels que, pour tout z dans $\Sigma \cap \mathcal{V}$, $\phi_{T+\eta(z)}(z)$ appartient à $\Sigma \cap \mathcal{V}$. L'application $G : \Sigma \cap \mathcal{V} \rightarrow \Sigma \cap \mathcal{V}$ ainsi définie est un C^1 -difféomorphisme : c'est l'application de premier retour de Poincaré. On a en outre, pour tout $(\delta s, \delta z) \in \mathbb{R} \times \mathbb{R}^{n-1}$,*

$$D\phi_T(x_0) \cdot \begin{pmatrix} \delta s \\ \delta z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & D_z G(x_0) \end{pmatrix} \cdot \begin{pmatrix} \delta s \\ \delta z \end{pmatrix}.$$

Remarque. Dans le cas des flots, l'application de Poincaré ne dépend que des sections transverses à l'orbite, Σ_0, Σ_1 , et encore à conjugaison près : elle a donc un sens géométrique très fort. D'autre part cette construction est très utile dans l'étude du voisinage d'une orbite périodique. Enfin notez que la construction précédente permet de passer de l'étude d'un flot en dimension n à celui d'un difféomorphisme local en dimension $n - 1$.

2.2 Équations différentielles linéaires autonomes

Nous abordons dans ce chapitre l'étude des équations différentielles les plus simples, les équations linéaires *autonomes* – aussi appelées équations linéaires à *coefficients constants* –, c'est-à-dire les équations de la forme

$$x'(t) = Ax(t). \quad (2.11)$$

Précisons les notations. La *donnée* $A \in M_n(\mathbb{K})$ est une matrice carrée ($n \times n$) à coefficients dans \mathbb{K} , où $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} . L'*inconnue* est une application dérivable $x(\cdot) : \mathbb{R} \rightarrow \mathbb{K}^n$. Résoudre l'équation (2.11) signifie trouver une application $x(\cdot)$ telle que, pour tout $t \in \mathbb{R}$, la dérivée $x'(t) = \frac{dx}{dt}(t)$ vérifie $x'(t) = Ax(t)$.

Cette équation est dite autonome parce que la donnée $A \in M_n(\mathbb{R})$ ne dépend pas du temps.

2.2.1 Approche élémentaire

Commençons par un cas connu, celui d'une équation scalaire

$$x'(t) = \alpha x(t),$$

où α est un réel et x une fonction de \mathbb{R} dans \mathbb{R} . Une solution $x(\cdot)$ de cette équation décrit l'évolution en fonction du temps d'une quantité dont le taux de variation α est constant.

Rappelons (cela résulte également du théorème 2.19) que la seule solution de cette équation valant x_0 à l'instant t_0 est

$$x(t) = x_0 e^{\alpha(t-t_0)}.$$

Cette expression nous fournit toutes les informations que l'on peut souhaiter sur l'équation différentielle. Par exemple le comportement asymptotique de $x(t)$ quand $t \rightarrow +\infty$ est caractérisé par le signe de α :

- si $\alpha < 0$, $\lim_{t \rightarrow +\infty} x(t) = 0$,
- si $\alpha = 0$, $x(t)$ est constant,
- si $\alpha > 0$, $\lim_{t \rightarrow +\infty} |x(t)| = \begin{cases} +\infty & \text{si } x(t_0) \neq 0 \\ 0 & \text{si } x(t_0) = 0 \end{cases}$.

Considérons maintenant un système de deux équations différentielles

$$\begin{cases} x_1' &= \alpha_1 x_1 \\ x_2' &= \alpha_2 x_2 \end{cases}.$$

C'est un système très simple puisque les fonctions $x_1(t)$ et $x_2(t)$ sont découplées. La solution de ce système est bien évidemment

$$x_1(t) = x_1(t_0) e^{\alpha_1(t-t_0)}, \quad x_2(t) = x_2(t_0) e^{\alpha_2(t-t_0)}.$$

Comme dans le cas scalaire, on a une connaissance complète du comportement des solutions. Par exemple, si α_1 et α_2 sont strictement négatifs, toute solution $(x_1(t), x_2(t))$ du système d'équations tend vers l'origine quand $t \rightarrow +\infty$; si $\alpha_1 > 0$ et $x_1(t_0) \neq 0$, la norme de $(x_1(t), x_2(t))$ tend vers l'infini quand $t \rightarrow +\infty$...

Adoptons maintenant une écriture matricielle. En posant $x = (x_1, x_2)$, le système de deux équations ci-dessus apparaît comme le cas particulier $n = 2$ de l'équation différentielle dans \mathbb{R}^n suivante :

$$x'(t) = \Delta x(t), \quad \text{où } \Delta = \begin{pmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_n \end{pmatrix} \text{ est diagonale.} \quad (2.12)$$

Cette équation étant en fait un système de n équations scalaires découplées, la solution est donnée par

$$x(t) = \begin{pmatrix} x_1(t_0) e^{\alpha_1(t-t_0)} \\ \vdots \\ x_n(t_0) e^{\alpha_n(t-t_0)} \end{pmatrix} = \begin{pmatrix} e^{\alpha_1(t-t_0)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\alpha_n(t-t_0)} \end{pmatrix} x(t_0),$$

avec $x = (x_1, \dots, x_n)$. Nous verrons dans la section suivante que la matrice diagonale ci-dessus est l'exponentielle de la matrice Δ et nous la noterons donc $e^{(t-t_0)\Delta}$.

Ainsi, pour les équations différentielles de la forme (2.12), nous avons une connaissance parfaite des solutions. Nous sommes par exemple en mesure d'analyser le comportement asymptotique des solutions en fonction de $\alpha_1, \dots, \alpha_n$ et de la condition initiale $x(t_0)$:

- si tous les α_i sont strictement négatifs, toute solution $x(t)$ converge vers l'origine quand $t \rightarrow +\infty$;
- si tous les α_i sont négatifs ou nuls, toute solution $x(t)$ est bornée quand $t \rightarrow +\infty$;
- si au moins un des α_i est strictement positif, alors $\lim_{t \rightarrow +\infty} \|x(t)\| = +\infty$ pour toute solution vérifiant $x_i(t_0) \neq 0$;
- etc. . .

L'équation différentielle $x'(t) = \Delta x(t)$ que nous venons de traiter est un cas très particulier, puisqu'il correspond à un système de n équations scalaires découplées. Beaucoup d'équations différentielles linéaires peuvent cependant s'y ramener. Considérons en effet le système $x'(t) = Ax(t)$ avec A diagonalisable dans \mathbb{R} : il existe donc une matrice inversible $P \in GL_n(\mathbb{R})$ et une matrice diagonale $\Delta \in M_n(\mathbb{R})$ telles que $A = P\Delta P^{-1}$.

Remarquons maintenant que, si $x(t)$ vérifie $x'(t) = Ax(t)$, alors $y(t) = P^{-1}x(t)$ vérifie $y'(t) = \Delta y(t)$. Autrement dit, à un *changement de coordonnées* près, l'équation différentielle est un système de n équations scalaires découplées. Connaissant $y(t_0) = P^{-1}x(t_0)$, on obtient alors $y(t) = e^{(t-t_0)\Delta}y(t_0)$, et

$$x(t) = Py(t) = Pe^{(t-t_0)\Delta}y(t_0) = Pe^{(t-t_0)\Delta}P^{-1}x(t_0).$$

On est donc encore capable de calculer les solutions de l'équation différentielle dans ce cas. Plus important, on voit que le comportement asymptotique des solutions est caractérisé par les éléments diagonaux de Δ , c'est-à-dire par les *valeurs propres* de A .

En résumé, cette première approche élémentaire fait apparaître les points clés que nous allons développer maintenant :

- les solutions se calculent à l'aide de l'exponentielle de matrice ;
- le comportement asymptotique des solutions est caractérisé par les valeurs propres de A .

2.2.2 Exponentielle de matrices

Définition 2.10. On appelle *exponentielle de matrice* l'application

$$\begin{aligned} \exp : M_n(\mathbb{K}) &\longrightarrow M_n(\mathbb{K}) \\ A &\longmapsto \exp A = e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} \end{aligned}$$

Notons que la série $\sum_{k=0}^{\infty} \frac{A^k}{k!}$ converge normalement. En effet

$$\sum_{k=0}^{\infty} \frac{\|A^k\|}{k!} \leq \sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} = e^{\|A\|} < \infty,$$

où on a choisi pour $\|\cdot\|$ une norme multiplicative sur $M_n(\mathbb{K})$ (par exemple une norme d'opérateurs). L'application exponentielle est donc continue (elle est en fait C^∞). Rappelons ses propriétés principales (sans démonstrations).

Proposition 2.18.

1. Pour $A \in M_n(\mathbb{K})$, l'application $t \mapsto e^{tA}$ est dérivable et

$$\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A$$

2. Pour tout $A \in M_n(\mathbb{K})$, $\exp(A)$ est inversible et $\exp(A)^{-1} = \exp(-A)$.

3. Si A et $B \in M_n(\mathbb{K})$ commutent, i.e. $AB = BA$, on a

$$\exp(A + B) = \exp(A)\exp(B).$$

4. Si $P \in GL_n(\mathbb{K})$, alors $Pe^AP^{-1} = e^{PAP^{-1}}$.

5. Si Δ est une matrice diagonale d'éléments diagonaux $\lambda_1, \dots, \lambda_n$, alors e^Δ est diagonale d'éléments diagonaux $e^{\lambda_1}, \dots, e^{\lambda_n}$.

Nous pouvons maintenant résoudre l'équation (2.11).

Théorème 2.19. Soient $t_0 \in \mathbb{R}$ et $x_0 \in \mathbb{K}^n$. L'unique solution de l'équation $x'(t) = Ax(t)$ valant x_0 en t_0 est l'application $x(\cdot)$ définie par

$$x(t) = e^{(t-t_0)A}x_0, \quad \forall t \in \mathbb{R}.$$

PREUVE.

▷ La première propriété de la proposition précédente implique

$$\frac{d}{dt} \left(e^{(t-t_0)A}x_0 \right) = \left(\frac{d}{dt} e^{(t-t_0)A} \right) x_0 = Ae^{(t-t_0)A}x_0.$$

La fonction $x(t) = e^{(t-t_0)A}x_0$ est donc solution de $x'(t) = Ax(t)$, et $x(t_0) = x_0$ puisque $e^{0A} = I$.

Pour montrer l'unicité de la solution, considérons une autre solution $y(t)$ valant x_0 en t_0 et formons le produit $z(t) = e^{-(t-t_0)A}y(t)$. En utilisant encore la proposition 2.18, on obtient

$$z'(t) = -Ae^{-(t-t_0)A}y(t) + e^{-(t-t_0)A}y'(t) = -Ae^{-(t-t_0)A}y(t) + e^{-(t-t_0)A}Ay(t) = 0,$$

car A et e^{tA} commutent. Donc $z(t)$ est une constante, et puisque $z(t_0) = x_0$, on obtient $y(t) = e^{(t-t_0)A}z(t_0) = e^{(t-t_0)A}x_0$.

□

Ainsi, la résolution d'équations linéaires autonomes se ramène au calcul d'exponentielle de matrices. Remarquons en particulier que les deux derniers points de la proposition 2.18 permettent de retrouver le résultat de la section 2.2.1 : si A est diagonalisable dans \mathbb{K} , c'est-à-dire

$$A = P\Delta P^{-1}, \quad P \in GL_n(\mathbb{K}), \quad \Delta \in M_n(\mathbb{K}) \text{ diagonale,}$$

alors $e^{tA} = Pe^{t\Delta}P^{-1}$ et la solution de l'équation (2.11) est

$$x(t) = Pe^{(t-t_0)\Delta}P^{-1}x(t_0).$$

Malheureusement toutes les matrices ne sont pas diagonalisables. La théorie de la réduction des endomorphismes permet cependant de mener à bien le calcul.

2.2.3 Calcul de l'exponentielle de matrices

Le but de cette section est de calculer l'exponentielle d'une matrice $A \in M_n(\mathbb{K})$ à un changement de base près, c'est-à-dire de calculer $P^{-1}e^{tA}P$ pour un $P \in GL_n(\mathbb{K})$ bien choisi.

Nous mènerons d'abord ce calcul dans le cas $\mathbb{K} = \mathbb{C}$, pour les matrices à coefficients complexes. Nous montrerons ensuite comment en déduire le cas $\mathbb{K} = \mathbb{R}$.

a. Matrices à coefficients complexes

Polynôme caractéristique. Considérons une matrice $A \in M_n(\mathbb{C})$ et notons $\lambda_1, \dots, \lambda_r$ ses *valeurs propres*. Rappelons que ce sont les seuls nombres complexes pour lesquels l'équation

$$Av = \lambda_i v$$

admette une solution $v \in \mathbb{C}^n$ non nulle (les v_i correspondants s'appellent des *vecteurs propres*). Les valeurs propres s'obtiennent également comme les racines du *polynôme caractéristique* de A : $P_A(\lambda) = \det(\lambda I - A)$. Ce polynôme est donc de la forme

$$P_A(\lambda) = (\lambda - \lambda_1)^{p_1} \cdots (\lambda - \lambda_r)^{p_r},$$

où chaque entier p_i est strictement positif et $p_1 + \cdots + p_r = n$ (le polynôme caractéristique étant de degré n). On appelle p_i la *multiplicité algébrique* de la valeur propre λ_i .

Une propriété importante du polynôme caractéristique est qu'il s'annule en A .

Théorème 2.20 (Cayley-Hamilton). *Toute matrice annule son polynôme caractéristique :*

$$P_A(A) = (A - \lambda_1 I)^{p_1} \cdots (A - \lambda_r I)^{p_r} = 0.$$

Sous-espaces propres, sous-espaces caractéristiques. À chaque valeur propre de A sont associés deux sous-espaces vectoriels de \mathbb{C}^n . Le premier est le *sous-espace propre* :

$$\Pi_i \text{ ou } \Pi_{\lambda_i} = \ker_{\mathbb{C}}(A - \lambda_i I).$$

C'est l'ensemble des vecteurs propres associés à λ_i . L'entier $e_i = \dim \Pi_i$ est appelé la *multiplicité géométrique* de la valeur propre λ_i .

Le deuxième est le *sous-espaces caractéristique* :

$$\Gamma_i \text{ ou } \Gamma_{\lambda_i} = \ker_{\mathbb{C}}(A - \lambda_i I)^{p_i}.$$

Il est clair que $\Pi_i \subset \Gamma_i$, mais ces deux espaces peuvent être différents. Le rôle des espaces caractéristiques est précisé dans le résultat suivant, que nous donnons sans démonstration.

Théorème 2.21 (de décomposition des noyaux). *Avec les notations précédentes, on a la décomposition*

$$\mathbb{C}^n = \Gamma_1 \oplus \cdots \oplus \Gamma_r,$$

et les propriétés suivantes :

1. $\dim \Gamma_i = p_i$;
2. *chacun des espaces Γ_i est invariant par $A : x \in \Gamma_i \Rightarrow Ax \in \Gamma_i$;*
3. *la restriction $A|_{\Gamma_i}$ de A à Γ_i s'écrit*

$$A|_{\Gamma_i} = \lambda_i I_{\Gamma_i} + N_i,$$

où I_{Γ_i} désigne l'identité de Γ_i et $N_i \in \text{End}(\Gamma_i)$ est nilpotent d'ordre $\leq p_i$, i.e. $N_i^{p_i} = 0$.

Ce résultat appelle un certain nombre de commentaires.

- Rappelons que $\text{End}(\Gamma_i)$ désigne l'ensemble des *endomorphismes* de Γ_i , c'est-à-dire des applications linéaires de Γ_i dans lui-même. Dire que Γ_i est invariant par A est équivalent à dire que $A|_{\Gamma_i}$ appartient à $\text{End}(\Gamma_i)$.
- L'opérateur N_i est défini comme $N_i = (A - \lambda_i I)|_{\Gamma_i}$. Le fait que N_i soit nilpotent d'ordre $\leq p_i$ n'est donc rien d'autre que la définition de Γ_i . Il est en revanche possible que l'ordre exact de nilpotence de N_i , c'est-à-dire le plus petit entier $m_i \leq p_i$ tel que $N_i^{m_i} = 0$, soit plus petit que p_i . Dans ce cas, on a

$$\ker(A - \lambda_i I)^{p_i} = \ker(A - \lambda_i I)^{m_i} \subsetneq \ker(A - \lambda_i I)^{m_i - 1}.$$

- Une matrice est *diagonalisable* si il existe une base de \mathbb{C}^n formée de vecteurs propres, ce qui équivaut à

$$\mathbb{C}^n = \Pi_1 \oplus \cdots \oplus \Pi_r.$$

D'après le théorème de décomposition des noyaux, ceci n'est possible que si $\Pi_i = \Gamma_i$ pour tout i . Autrement dit :

A est diagonalisable si et seulement si pour toute valeur propre les multiplicités algébrique et géométrique coïncident, i.e. $\dim \Pi_i = p_i$ pour $i = 1, \dots, r$.

Réduction de Jordan dans \mathbb{C}^n . Choisissons une base \mathcal{B} de \mathbb{C}^n formée de la réunion d'une base de Γ_1 , d'une base de Γ_2 , ..., d'une base de Γ_r , et notons P la matrice de passage de cette base à la base canonique. D'après le théorème de décomposition des noyaux, l'application linéaire associée à A a pour matrice dans la base \mathcal{B} :

$$P^{-1}AP = \Delta + N,$$

où Δ est la matrice diagonale ayant pour éléments diagonaux λ_1 (p_1 fois), ..., λ_r (p_r fois), et N est la matrice nilpotente qui s'écrit par blocs

$$N = \begin{pmatrix} N_1 & & \\ & \ddots & \\ & & N_r \end{pmatrix}.$$

Il est en fait possible de choisir la base \mathcal{B} de façon à mettre la matrice nilpotente N sous une forme relativement simple. On aboutit ainsi à la réduction de Jordan.

Théorème 2.22 (de Jordan). *Pour toute matrice $A \in M_n(\mathbb{C})$, il existe $P \in GL_n(\mathbb{C})$ telle que $P^{-1}AP$ s'écrit sous forme de matrice diagonale par bloc*

$$P^{-1}AP = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{pmatrix}, \quad (2.14)$$

où chaque J_i est une matrice ($p_i \times p_i$) de la forme

$$J_i = \begin{pmatrix} J_{i,1} & & \\ & \ddots & \\ & & J_{i,e_i} \end{pmatrix}, \quad \text{avec} \quad J_{i,k} = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$$

et $e_i = \dim \ker(A - \lambda_i I)$ est la multiplicité géométrique de λ_i .

On appelle la matrice $J = P^{-1}AP$ la *forme réduite de Jordan* de A et les matrices J_i des *blocs de Jordan*.

Remarques.

- Pour chaque i , la matrice J_i représente l'application linéaire $A|_{\Gamma_i}$, ce qui implique $J_i \in \text{End}(\Gamma_i)$.

- Les matrices $J_{i,k}$ sont des matrices carrées dont la dimension $\dim(J_{i,k}) \leq p_i$ dépend de i et k . Dans le cas particulier où les multiplicités algébrique et géométrique de λ_i coïncident (*i.e.* $e_i = p_i$), la dimension $\dim(J_{i,k})$ est égale à 1 pour tout k . Chaque bloc $J_{i,k}$ est alors réduit au scalaire λ_i et $J_i = \lambda_i I_{p_i}$.
- D'après la remarque précédente, si pour chaque valeur propre les multiplicités algébrique et géométrique coïncident, la forme réduite de Jordan est diagonale. Ainsi la réduction de Jordan généralise la diagonalisation. Insistons cependant sur le fait que toute matrice de $M_n(\mathbb{C})$ admet une réduction de Jordan, mais n'est pas forcément diagonalisable.

Calcul de l'exponentielle. La réduction de Jordan permet de calculer e^{tA} à conjugaison près. En effet, commençons par le calcul de l'exponentielle d'un bloc $J_{i,k}$. Notons $n_{i,k} = \dim(J_{i,k})$ la dimension de ce bloc et écrivons-le $J_{i,k} = \lambda_i I + N_{i,k}$, où $N_{i,k}$ est la matrice ayant des 0 sur la diagonale et des 1 juste au-dessus. Comme $\lambda_i I$ et $N_{i,k}$ commutent,

$$e^{tJ_{i,k}} = e^{t\lambda_i I} e^{tN_{i,k}} = e^{t\lambda_i} e^{tN_{i,k}}.$$

De plus, $N_{i,k}$ étant nilpotente d'ordre $n_{i,k}$, on a :

$$e^{tN_{i,k}} = \sum_{l=0}^{\infty} \frac{(tN_{i,k})^l}{l!} = \sum_{l=0}^{n_{i,k}-1} \frac{(tN_{i,k})^l}{l!} = \begin{pmatrix} 1 & t & \cdots & \frac{t^{n_{i,k}-1}}{(n_{i,k}-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & t \\ & & & 1 \end{pmatrix}.$$

D'autre part, d'après les propriétés de l'exponentielle de matrice (proposition 2.18), on a $e^{tA} = e^{tPJP^{-1}} = Pe^{tJ}P^{-1}$, ce qui donne finalement l'expression de l'exponentielle de tA :

$$e^{tA} = P \begin{pmatrix} e^{tJ_{1,e_1}} & & \\ & \ddots & \\ & & e^{tJ_{r,e_r}} \end{pmatrix} P^{-1}, \quad (2.15)$$

avec $e^{tJ_{i,k}} = e^{t\lambda_i} \begin{pmatrix} 1 & t & \cdots & \frac{t^{n_{i,k}-1}}{(n_{i,k}-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & t \\ & & & 1 \end{pmatrix}.$

b. Matrices à coefficients réels

Considérons maintenant une matrice $A \in M_n(\mathbb{R})$, à coefficients réels. On peut bien entendu considérer A comme une matrice de $M_n(\mathbb{C})$; tout ce que nous venons de voir pour les matrices à coefficients complexes s'applique donc.

Désignons les valeurs propres réelles de A par $\lambda_1, \dots, \lambda_s$, et ses valeurs propres non réelles par $\lambda_{s+1}, \bar{\lambda}_{s+1}, \dots, \lambda_q, \bar{\lambda}_q$ (avec $2q - s = r$). Le polynôme caractéristique de A est donc le polynôme à coefficients réels

$$P_A(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{p_i} \prod_{i=s+1}^q [(\lambda - \lambda_i)(\lambda - \bar{\lambda}_i)]^{p_i}.$$

Les sous-espaces vectoriels $\Gamma_{\lambda_i} = \ker_{\mathbb{C}}(A - \lambda_i I)^{p_i}$ de \mathbb{C}^n sont maintenant appelés les sous-espaces caractéristiques *complexes*.

Remarque. Rappelons les liens qui existent entre les sous-espaces vectoriels de \mathbb{C}^n , qui sont des \mathbb{C} -espaces vectoriels, et ceux de \mathbb{R}^n , qui sont des \mathbb{R} -espaces vectoriels. On considère \mathbb{R}^n comme un sous-ensemble de \mathbb{C}^n et, pour un sous-espace vectoriel Γ de \mathbb{C}^n , on note $\Gamma \cap \mathbb{R}^n$ l'ensemble des vecteurs $v \in \Gamma$ qui sont réels. Il est facile de vérifier qu'un tel ensemble $\Gamma \cap \mathbb{R}^n$ est un sous-espace vectoriel de \mathbb{R}^n . De plus, si Γ est stable par conjugaison (*i.e.* $v \in \Gamma \Rightarrow \bar{v} \in \Gamma$), alors Γ et $\Gamma \cap \mathbb{R}^n$ ont même dimension en tant que sous-espaces respectivement de \mathbb{C}^n et de \mathbb{R}^n (en fait, dans ce cas, Γ est l'ensemble des combinaisons linéaires à coefficients complexes des éléments de $\Gamma \cap \mathbb{R}^n$; en conséquence, toute base du \mathbb{R} -espace vectoriel $\Gamma \cap \mathbb{R}^n$ est également une base du \mathbb{C} -espace vectoriel Γ).

Définissons maintenant les *sous-espaces caractéristiques réels* de A comme les sous-espaces vectoriels de \mathbb{R}^n :

$$\begin{aligned} E_i &= \Gamma_{\lambda_i} \cap \mathbb{R}^n, & 1 \leq i \leq s \\ E_i &= (\Gamma_{\lambda_i} \oplus \Gamma_{\bar{\lambda}_i}) \cap \mathbb{R}^n, & s+1 \leq i \leq q. \end{aligned}$$

Remarquons alors que $(A - \lambda_i I)^{p_i} v = 0$ implique $(A - \bar{\lambda}_i I)^{p_i} \bar{v} = 0$, ce qui signifie que Γ_{λ_i} pour λ_i réel et $\Gamma_{\lambda_i} \oplus \Gamma_{\bar{\lambda}_i}$ pour λ_i non réel sont stables par conjugaison. D'après la remarque précédente et le théorème de décomposition des noyaux dans \mathbb{C}^n , on a la décomposition

$$\mathbb{R}^n = E_1 \oplus \dots \oplus E_q,$$

chaque sous-espace E_i étant invariant par A . À partir de cette décomposition, nous allons donner ci-dessous une forme réduite de Jordan réelle de A . Cette forme réduite n'est cependant pas indispensable à l'étude des équations différentielles : nous verrons dans la section suivante que les solutions de l'équation (2.11) dans \mathbb{R}^n peuvent se déduire directement des solutions dans \mathbb{C}^n .

***Réduction de Jordan dans \mathbb{R}^n .** Donner une forme réduite de $A \in M_n(\mathbb{R})$ consiste à trouver pour chaque sous-espace caractéristique une base dans laquelle l'application linéaire associée à A a une expression simple (c'est ce que nous avons fait pour les matrices à coefficients complexes). Considérons donc un des sous-espaces caractéristiques réels E_k de A .

- Si λ_k est réelle, *i.e.* $1 \leq k \leq s$: dans ce cas $E_i = \ker_{\mathbb{R}}(A - \lambda_k I)^{p_k}$ et la restriction de A à E_k s'écrit simplement

$$A|_{E_k} = \lambda_k I|_{E_k} + N_k, \quad \text{où } N_k \text{ nilpotente.}$$

On peut alors montrer comme dans le cas complexe que $A|_{E_k}$ est conjuguée au bloc de Jordan J_k .

- Si $\lambda_k = \alpha_k + i\beta_k$ n'est pas réelle, i.e. $s + 1 \leq k \leq q$: choisissons une base v_1, \dots, v_{p_k} de Γ_k dans laquelle $A|_{\Gamma_k}$ se met sous forme d'un bloc de Jordan J_k , c'est-à-dire, pour $j = 1, \dots, p_k$,

$$Av_j = \lambda_k v_j + \delta_j v_{j-1}, \quad \text{où } \delta_j = (J_k)_{j-1,j} = 0 \text{ ou } 1,$$

en posant $v_0 = 0$. Par conjugaison, il est clair que $\bar{v}_1, \dots, \bar{v}_{p_k}$ est une base de $\Gamma_{\bar{\lambda}_k}$. Pour $j = 1, \dots, p_k$, posons $v_j = a_j + ib_j$, avec $a_j, b_j \in \mathbb{R}^n$. Les vecteurs $(a_1, b_1, \dots, a_{p_k}, b_{p_k})$ forment alors une base de E_k . En effet, les $a_j = \frac{1}{2}(v_j + \bar{v}_j)$ et $b_j = \frac{i}{2}(\bar{v}_j - v_j)$ appartiennent à $\Gamma_{\lambda_k} \oplus \Gamma_{\bar{\lambda}_k}$ et l'engendrent en tant que \mathbb{C} -espace vectoriel puisqu'ils engendrent la base des v_j, \bar{v}_j : ils forment donc une famille génératrice à $2p_k$ éléments du \mathbb{R} -espace vectoriel E_k de dimension $2p_k$ éléments, c'est-à-dire une base.

De plus, en identifiant les parties réelles et imaginaires dans l'expression

$$A(a_j + ib_j) = (\alpha_k + i\beta_k)(a_j + ib_j) + \delta_j(a_{j-1} + ib_{j-1}),$$

on obtient

$$A \begin{bmatrix} a_j \\ b_j \end{bmatrix} = C_k \begin{bmatrix} a_j \\ b_j \end{bmatrix} + \delta_j \begin{bmatrix} a_{j-1} \\ b_{j-1} \end{bmatrix} \quad \text{où } C_k = \begin{pmatrix} \alpha_k & -\beta_k \\ \beta_k & \alpha_k \end{pmatrix}.$$

La restriction de A à E_k est donc conjuguée dans la base $(a_1, b_1, \dots, a_{p_k}, b_{p_k})$ à la matrice

$$J'_k = \begin{pmatrix} C_k & \delta_2 I_2 & & \\ & \ddots & \ddots & \\ & & \ddots & \delta_{p_k} I_2 \\ & & & C_k \end{pmatrix}.$$

On obtient ainsi la réduction de Jordan dans \mathbb{R}^n de A .

Théorème 2.23. *Pour toute matrice $A \in M_n(\mathbb{R})$, il existe $Q \in GL_n(\mathbb{R})$ telle que $Q^{-1}AQ$ s'écrit sous forme de matrice diagonale par bloc J' , d'éléments diagonaux $J_1, \dots, J_s, J'_{s+1}, \dots, J'_q$, où, pour $i = s + 1, \dots, q$, chaque J'_i est une matrice $(2p_i \times 2p_i)$ de la forme*

$$J'_i = \begin{pmatrix} J'_{i,1} & & \\ & \ddots & \\ & & J'_{i,2e_i} \end{pmatrix}, \quad \text{avec } J'_{i,k} = \begin{pmatrix} C_i & I_2 & & \\ & \ddots & \ddots & \\ & & \ddots & I_2 \\ & & & C_i \end{pmatrix}$$

et, pour $i = 1, \dots, s$, les matrices J_i sont celles données dans le théorème 2.22.

Le calcul de l'exponentielle ne pose alors aucun problème, il suffit de savoir calculer e^{tC_i} . Écrivons C_i comme la somme de deux matrices qui commutent

$$C_i = \alpha_i I_2 + \beta_i B, \quad B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

et donc son exponentielle comme un produit

$$e^{tC_i} = e^{\alpha_i t} e^{\beta_i t B}.$$

Remarquons que la matrice B est de carré égal à $-I_2$, ce qui implique $B^{2p} = (-1)^p I_2$ et $B^{2p+1} = (-1)^p B$, soit

$$e^{sB} = \sum_{k=0}^{\infty} \frac{1}{k!} (sB)^k = \sum_{p=0}^{\infty} \frac{1}{(2p)!} (-1)^p s^{2p} I_2 + \sum_{p=0}^{\infty} \frac{1}{(2p+1)!} (-1)^p s^{2p+1} B = \cos(s) I_2 + \sin(s) B.$$

On obtient ainsi l'exponentielle de C_i :

$$e^{tC_i} = e^{\alpha_i t} \begin{pmatrix} \cos(\beta_i t) & -\sin(\beta_i t) \\ \sin(\beta_i t) & \cos(\beta_i t) \end{pmatrix}.$$

2.2.4 Forme des solutions

Grâce au calcul de l'exponentielle que nous venons d'effectuer, nous sommes maintenant en mesure de préciser le théorème 2.19. Donnons d'abord la forme de la solution générale dans \mathbb{C}^n , qui s'obtient directement à partir de l'expression (2.15) de e^{tA} .

Théorème 2.24. *Soit $A \in M_n(\mathbb{C})$. Toute solution de $x'(t) = Ax(t)$ dans \mathbb{C}^n s'écrit sous la forme*

$$x(t) = \sum_{1 \leq i \leq r} e^{t\lambda_i} \left(\sum_{0 \leq k \leq m_i - 1} t^k v_{i,k} \right), \quad \text{où } v_{i,k} \in \Gamma_i, \quad (2.17)$$

avec $m_i = \max_{1 \leq k \leq e_i} \dim(J_{i,k})$.

Remarques.

- Le terme en facteur de $e^{t\lambda_i}$ est polynômial en t quand $m_i > 1$, et constant quand $m_i = 1$. Rappelons que ce dernier cas a lieu si et seulement si les multiplicités algébrique et géométrique de λ_i coïncident.
- Il est également important de noter la dépendance de $x(t)$ par rapport à la condition initiale $x(0)$ dans la décomposition précédente. Si $x(0)$ s'écrit comme $x(0) = v_1 + \dots + v_r$ dans $\mathbb{C}^n = \Gamma_1 \oplus \dots \oplus \Gamma_r$, alors

$$v_{i,k} = \frac{1}{k!} N^k v_i,$$

où N est la matrice nilpotente de la décomposition $\Delta + N$ de $P^{-1}AP$. En particulier, $v_i = 0$ si et seulement si tous les vecteurs $v_{i,k}$ sont nuls. Ceci est en fait une conséquence de l'invariance des sous-espaces Γ_i par A .

Considérons maintenant une matrice A à coefficients réels et une solution $x(\cdot)$ dans \mathbb{R}^n de $x' = Ax$, dont la condition initiale est $x(0) \in \mathbb{R}^n$. On peut bien entendu considérer A comme une matrice à coefficients complexes et $x(\cdot) = x(\cdot) + i0$ comme la solution dans \mathbb{C}^n de $x' = Ax$ ayant pour condition initiale $x(0) + i0$. L'expression de $x(\cdot)$ est donc donnée par la formule (2.17). Puisque cette solution est réelle, elle est en fait égale à la partie réelle de la formule (2.17), la partie imaginaire devant être nulle. On obtient ainsi la forme générale suivante pour les solutions de $x' = Ax$ dans \mathbb{R}^n .

Théorème 2.25. Soit $A \in M_n(\mathbb{R})$. Toute solution de $x'(t) = Ax(t)$ dans \mathbb{R}^n s'écrit sous la forme

$$x(t) = \sum_{1 \leq i \leq q} e^{t\alpha_i} \left(\sum_{0 \leq k \leq m_i - 1} t^k (\cos(\beta_i t) a_{i,k} + \sin(\beta_i t) b_{i,k}) \right), \quad (2.19)$$

où $\alpha_i = \Re(\lambda_i)$, $\beta_i = \Im(\lambda_i)$ et les vecteurs $a_{i,k}$, $b_{i,k}$ appartiennent à E_i .

Remarque. Comme dans le théorème 2.24, les vecteurs $a_{i,k}$, $b_{i,k}$ dépendent uniquement de la condition initiale $x(0)$. Si $x(0)$ s'écrit comme $x(0) = u_1 + \dots + u_q$ dans $\mathbb{R}^n = E_1 \oplus \dots \oplus E_q$, alors $u_i = 0$ si et seulement si tous les vecteurs $a_{i,k}$ et $b_{i,k}$ sont nuls.

Comportement asymptotique.

Les théorèmes 2.24 et 2.25 donnent toutes les informations que l'on peut souhaiter sur l'équation différentielle, généralisant les résultats de la section 2.2.1 sur les équations scalaires et les systèmes d'équations découplées. On constate en particulier que le comportement quand t tend vers l'infini des solutions $x(t)$ de $x'(t) = Ax(t)$ dépend essentiellement des signes des parties réelles des valeurs propres λ_i de A . Plus précisément, on peut décomposer le comportement des composantes de $x(t)$ sur chaque sous-espace caractéristique (on suppose ici A réelle) :

- si $\Re(\lambda_i) < 0$ la projection sur E_i de $x(t)$ s'annule quand t tend vers $+\infty$ et croît de façon au moins exponentielle en $-\infty$;
- si $\Re(\lambda_i) > 0$, c'est l'inverse, la projection sur E_i de $x(t)$ croît de façon au moins exponentielle en $+\infty$ et s'annule quand t tend vers $-\infty$;
- si $\Re(\lambda_i) = 0$, la composante sur E_i de $x(t)$ croît de façon polynômiale en $\pm\infty$ quand $\dim \Pi_i < p_i$, et est bornée pour $t \in \mathbb{R}$ quand $\dim \Pi_i = p_i$.

Il est commode de regrouper les sous-espaces caractéristiques en fonction du signe de la partie réelle des valeurs propres correspondantes. Nous définissons ainsi, pour $A \in M_n(\mathbb{R})$,

$$\begin{aligned} \text{– l'espace } stable : \quad & E^s = \left[\bigoplus_{\Re(\lambda_i) < 0} \Gamma_i \right] \cap \mathbb{R}^n = \bigoplus_{\Re(\lambda_i) < 0} E_i, \\ \text{– l'espace } instable : \quad & E^u = \left[\bigoplus_{\Re(\lambda_i) > 0} \Gamma_i \right] \cap \mathbb{R}^n = \bigoplus_{\Re(\lambda_i) > 0} E_i, \\ \text{– l'espace } indifférent : \quad & E^c = \left[\bigoplus_{\Re(\lambda_i) = 0} \Gamma_i \right] \cap \mathbb{R}^n = \bigoplus_{\Re(\lambda_i) = 0} E_i, \end{aligned}$$

si bien que $\mathbb{R}^n = E^s \oplus E^u \oplus E^c$. De même, pour $A \in M_n(\mathbb{C})$, les espaces complexes *stable*, *instable* et *indifférent* sont définis respectivement comme

$$\Gamma^s = \bigoplus_{\Re(\lambda_i) < 0} \Gamma_i, \quad \Gamma^u = \bigoplus_{\Re(\lambda_i) > 0} \Gamma_i, \quad \Gamma^c = \bigoplus_{\Re(\lambda_i) = 0} \Gamma_i,$$

et on a la décomposition $\mathbb{C}^n = \Gamma^s \oplus \Gamma^u \oplus \Gamma^c$.

D'après le théorème de décomposition des noyaux, ces espaces ont la particularité d'être invariants par e^{tA} pour tout $t \in \mathbb{R}$: $e^{tA}E^s \subset E^s$, $e^{tA}E^u \subset E^u$, etc... Ce qui entraîne que, si une solution $x(\cdot)$ de $x'(t) = Ax(t)$ vérifie par exemple $x(0) \in E^s$, alors $x(t) \in E^s$ pour tout t ; si $x(0) \in E^c$, alors $x(t) \in E^c$ pour tout t , etc...

Les espaces stable, instable et indifférent correspondent chacun à un certain type de comportement asymptotique des solutions. Nous résumons ces comportements dans le théorème suivant, dont la démonstration est laissée en exercice (utiliser soit la forme générale des solutions, soit directement la réduction de Jordan).

Théorème 2.26. Soit A une matrice ($n \times n$) réelle (resp. complexe). Notons $x(\cdot)$ les solutions dans \mathbb{R}^n (resp. \mathbb{C}^n) de l'équation différentielle $x'(t) = Ax(t)$. Alors

- E^s (resp. Γ^s) est l'ensemble des $x(0) \in \mathbb{R}^n$ (resp. $x(0) \in \mathbb{C}^n$) pour lesquels

$$\lim_{t \rightarrow +\infty} \|x(t)\| = 0;$$

- E^u (resp. Γ^u) est l'ensemble des $x(0) \in \mathbb{R}^n$ (resp. $x(0) \in \mathbb{C}^n$) pour lesquels

$$\lim_{t \rightarrow -\infty} \|x(t)\| = 0;$$

- E^c (resp. Γ^c) est l'ensemble des $x(0) \in \mathbb{R}^n$ (resp. $x(0) \in \mathbb{C}^n$) pour lesquels il existe un entier $M \geq 0$ et une constante $C > 0$ tels que, pour $|t|$ suffisamment grand,

$$C^{-1} \|x(0)\| \leq \|x(t)\| \leq C |t|^M \|x(0)\|.$$

En outre, pour $0 < \alpha < \min_{\Re(\lambda_i) \neq 0} |\Re(\lambda_i)|$, il existe une constante $C > 0$ telle que :

- si $x(0) \in E^s$ (ou Γ^s), alors, pour tout $t > 0$ assez grand,

$$\|x(t)\| \leq Ce^{-\alpha t} \|x(0)\|, \quad \|x(-t)\| \geq C^{-1} e^{\alpha t} \|x(0)\|;$$

- si $x(0) \in E^u$ (ou Γ^u), alors, pour tout $t > 0$ assez grand,

$$\|x(t)\| \geq C^{-1} e^{\alpha t} \|x(0)\|, \quad \|x(-t)\| \leq Ce^{-\alpha t} \|x(0)\|.$$

Remarque. Dans la caractérisation de E^c , on peut prendre $M = 0$ si et seulement si A est diagonalisable (puisque dans ce cas, pour toute valeur propre, multiplicités algébrique et géométrique coïncident).

Pour obtenir le comportement asymptotique d'une solution particulière $x(\cdot)$, il suffit donc de décomposer sa condition initiale $x(0)$ en $x^s(0) + x^u(0) + x^c(0)$ dans $\mathbb{R}^n = E^s \oplus E^u \oplus E^c$. On sait alors que la décomposition de $x(t)$ est $x(t) = x^s(t) + x^u(t) + x^c(t)$, où

$x^s(\cdot)$ (resp. $x^u(\cdot)$, $x^c(\cdot)$) est la solution de $x'(t) = Ax(t)$ ayant $x^s(0)$ (resp. $x^u(0)$, $x^c(0)$) pour condition initiale. En particulier, si on s'intéresse aux temps positifs, on a les critères suivants :

- si $x^u(0) \neq 0$, alors $\|x(t)\|$ tend vers l'infini quand $t \rightarrow +\infty$;
- si $x^c(0) \neq 0$, alors $\|x(t)\|$ ne tend pas vers 0 quand $t \rightarrow +\infty$ (mais $\|x(t)\|$ ne tend pas forcément vers l'infini).

Définition 2.11. Une matrice carrée A est dite **Hurwitz** si toutes ses valeurs propres ont une partie réelle strictement négatives (c'est-à-dire $E^s = \mathbb{R}^n$ ou $\Gamma^s = \mathbb{C}^n$).

Corollaire 2.27. *Toutes les solutions de $x'(t) = Ax(t)$ tendent vers 0 quand $t \rightarrow +\infty$ si et seulement si A est Hurwitz.*

Quand c'est le cas, on dit que 0 est un *équilibre asymptotiquement stable* de l'équation.

Nous dirons qu'une matrice A est *hyperbolique* si elle n'a aucune valeur propre de partie réelle nulle. Pour une telle matrice, on a $\Gamma^c = \{0\}$, c'est-à-dire que $\mathbb{C}^n = \Gamma^s \oplus \Gamma^u$ et $\mathbb{R}^n = E^s \oplus E^u$ si $A \in M_n(\mathbb{R})$.

La classe des matrices hyperboliques est d'un intérêt particulier puisqu'elle est en fait stable par perturbation.

Théorème 2.28. *Si A est une matrice hyperbolique, il existe un réel $\delta > 0$ (dépendant de A) tel que pour toute matrice F vérifiant*

$$\|F\| \leq \delta,$$

la matrice $A + F$ est encore hyperbolique. L'ensemble des matrices hyperboliques complexes (resp. réelles) est donc ouvert dans $M_n(\mathbb{C})$ (resp. $M_n(\mathbb{R})$).

En outre, les espaces stable et instable dépendent continûment de F (pour $\|F\| \leq \delta$).

*PREUVE.

- ▷ Le fait que si F est assez petite $A + F$ est hyperbolique résulte de la continuité des valeurs propres des matrices : les valeurs propres de $A + F$ sont d'autant plus proches de celles de A que F est petite ; or celles-ci sont dans l'ouvert $\mathbb{C} - \{\Re e = 0\}$. Il en est donc de même de celles de F si elle est assez petite. □

Remarque. Alors que la dépendance des espaces caractéristiques de $A + F$ n'est pas en général continue par rapport à F , celle des espaces stable et instable l'est.

Cas d'une matrice diagonalisable.

Considérons le cas particulier d'une matrice $A \in M_n(\mathbb{R})$ diagonalisable dans \mathbb{C} (on dit aussi *semi-simple*). Comme nous l'avons vu dans la section 2.2.3, ceci signifie que A satisfait les conditions suivantes, qui sont équivalentes entre elles (nous utilisons les notations de la section 2.2.3) :

- il existe une base de \mathbb{C}^n formée de vecteurs propres de A ;
- $\mathbb{C}^n = \Pi_1 \oplus \cdots \oplus \Pi_r$, où $\Pi_i = \ker_{\mathbb{C}}(A - \lambda_i I) \subset \mathbb{C}^n$ est le sous-espace propre associé à λ_i et $\lambda_1, \dots, \lambda_r$ sont les valeurs propres (complexes) de A ;
- pour toute valeur propre λ_i , le sous-espace caractéristique (complexe) est égal au sous-espace propre : $\Gamma_i = \Pi_i$;
- pour toute valeur propre, les multiplicités géométrique et algébrique coïncident : $\dim \Pi_i = p_i$ pour $i = 1, \dots, r$;
- dans la forme réduite de Jordan complexe de la matrice A , tous les blocs de Jordan sont des matrices 1×1 : $J_{i,k} = (\lambda_i)$.

Ce cas est en pratique très important, puisque l'ensemble des matrices de $M_n(\mathbb{R})$ diagonalisables dans \mathbb{C} contient un ensemble ouvert et dense dans $M_n(\mathbb{R})$. Autrement dit, être diagonalisable dans \mathbb{C} est une propriété générique sur $M_n(\mathbb{R})$.

Considérons donc une telle matrice A . La forme générale des solutions de $x' = Ax$ donnée par le théorème 2.25 se simplifie : tous les termes polynômiaux en t disparaissent, seuls restent les termes exponentiels et trigonométriques.

Corollaire 2.29. *Soit $A \in M_n(\mathbb{R})$ une matrice diagonalisable dans \mathbb{C} . Toute solution de $x'(t) = Ax(t)$ dans \mathbb{R}^n s'écrit sous la forme*

$$x(t) = \sum_{1 \leq j \leq q} e^{t\alpha_j} (\cos(\beta_j t) a_j + \sin(\beta_j t) b_j),$$

où $\alpha_j = \Re(\lambda_j)$, $\beta_j = \Im(\lambda_j)$ et les vecteurs a_j et $b_j \in \mathbb{R}^n$ sont tels que $a_j + ib_j$ est un vecteur propre de A associé à λ_j (i.e. $a_j + ib_j \in \Pi_j$).

Les sous-espaces stables, instables et indifférents s'écrivent maintenant en fonction des sous-espaces propres (puisque ceux-ci sont égaux aux sous-espaces caractéristiques complexes) :

$$E^s = \left[\bigoplus_{\Re(\lambda_i) < 0} \Pi_i \right] \cap \mathbb{R}^n, \quad E^u = \left[\bigoplus_{\Re(\lambda_i) > 0} \Pi_i \right] \cap \mathbb{R}^n, \quad E^c = \left[\bigoplus_{\Re(\lambda_i) = 0} \Pi_i \right] \cap \mathbb{R}^n.$$

La caractérisation dynamique de ces espaces résulte du théorème 2.26.

Corollaire 2.30. Soit $A \in M_n(\mathbb{R})$ une matrice diagonalisable dans \mathbb{C} . Notons $x(\cdot)$ les solutions dans \mathbb{R}^n de l'équation différentielle $x'(t) = Ax(t)$. Alors

- E^s est l'ensemble des conditions initiales $x(0) \in \mathbb{R}^n$ correspondant à des solutions $x(t)$ qui tendent exponentiellement vers 0 quand $t \rightarrow +\infty$;
- E^u est l'ensemble des conditions initiales $x(0) \in \mathbb{R}^n$ correspondant à des solutions $x(t)$ qui tendent exponentiellement vers 0 quand $t \rightarrow -\infty$;
- E^c est l'ensemble des conditions initiales $x(0) \in \mathbb{R}^n$ correspondant à des solutions $x(t)$ périodiques, donc bornées sur \mathbb{R} .

La seule différence par rapport au cas général concerne l'espace E^c : alors que dans le cas général le comportement asymptotique des solutions dans E^c était indéterminé (d'où le nom d'espace « indifférent »), on sait ici que toutes les solutions dans E^c sont bornées.

2.3 Stabilité

2.3.1 Équilibres et stabilité

Considérons l'équation différentielle autonome

$$x'(t) = f(x(t)), \quad (2.20)$$

où le champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ est supposé de classe C^1 .

Définition 2.12. On dit qu'un point $x_0 \in \Omega$ est un *équilibre* de (2.20) si la fonction constante $x(\cdot) \equiv x_0$ est solution de (2.20) ou, de façon équivalente, si $f(x_0) = 0$ (vérifier que c'est bien équivalent!).

Autrement dit, $\phi_t(x_0) = x_0$ pour tout $t \in \mathbb{R}$, où ϕ est le flot du champ de vecteurs f (l'intervalle maximal associé à x_0 étant $I_{x_0} = \mathbb{R}$). L'orbite de x_0 est donc réduite à un point : $\mathcal{O}_{x_0} = \{x_0\}$.

Quand l'équation (2.20) modélise l'évolution d'un phénomène physique (mécanique, biologique, écologique, ...), un équilibre correspond bien à la notion habituelle « d'état d'équilibre » : si le système est dans l'état x_0 , alors il y reste (et il y a toujours été). En pratique on sait cependant que seuls les états d'équilibre ayant certaines propriétés de stabilité sont significatifs.

Définition 2.13. Nous dirons qu'un équilibre x_0 est *stable* si, pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que

$$\|x - x_0\| < \delta \quad \text{et} \quad t > 0 \quad \implies \quad \|\phi_t(x) - x_0\| < \epsilon.$$

Ainsi, toute solution proche de x_0 en reste proche.

Remarque. Toute solution dont la condition initiale est dans une boule $B(x_0, \delta)$ reste dans la boule $B(x_0, \epsilon)$, et donc dans un compact de Ω , pour $t > 0$ (on suppose ϵ suffisamment petit pour que $\bar{B}(x_0, \epsilon) \subset \Omega$). D'après la proposition 2.4, ces solutions sont donc définies pour tout $t > 0$.

Définition 2.14. Nous dirons qu'un équilibre x_0 est localement *asymptotiquement stable* (LAS) si il est stable et si il existe un voisinage V de x_0 tel que, pour tout $x \in V$,

$$\lim_{t \rightarrow \infty} \phi_t(x) = x_0.$$

Si V est égal à tout l'espace d'état, on dira que x_0 est globalement *asymptotiquement stable* (GAS).

Dans le cas (LAS), toute solution proche de l'équilibre en reste proche et en plus converge vers lui.

Le cas linéaire

Considérons le cas particulier d'une équation différentielle autonome linéaire

$$x'(t) = Ax(t), \quad x \in \mathbb{R}^n.$$

L'origine est toujours un équilibre de cette équation (mais il peut y en avoir d'autres : tout élément de $\ker A$ est un équilibre). L'étude réalisée dans la section 2.2.4 permet de caractériser la stabilité de cet équilibre. Par ailleurs, due à la linéarité du système (son homogénéité suffit), il n'y a pas de distinction entre local ou global. En conséquence, dans les énoncés qui suivent, la stabilité, quand elle a lieu, est globale.

Proposition 2.31.

- L'origine est un équilibre asymptotiquement stable de $x' = Ax$ si et seulement si toutes les valeurs propres de A sont de partie réelle strictement négative, i.e. $\mathbb{R}^n = E^s$.
- Si A a au moins une valeur propre de partie réelle strictement positive, alors l'origine n'est pas un équilibre stable de $x' = Ax$.

Notons que l'origine peut être un équilibre stable mais non asymptotiquement stable. C'est une situation que l'on rencontre quand A a des valeurs propres de partie réelle nulle, par exemple quand A est antisymétrique (voir proposition 2.13 et la discussion qui suit). On a représenté figure 2.2 des portraits de phase dans \mathbb{R}^2 correspondant à une matrice antisymétrique (cas a) et une autre dont les valeurs propres ont une partie réelle < 0 (cas b).

Remarquons enfin que l'on peut donner une condition nécessaire et suffisante de stabilité :

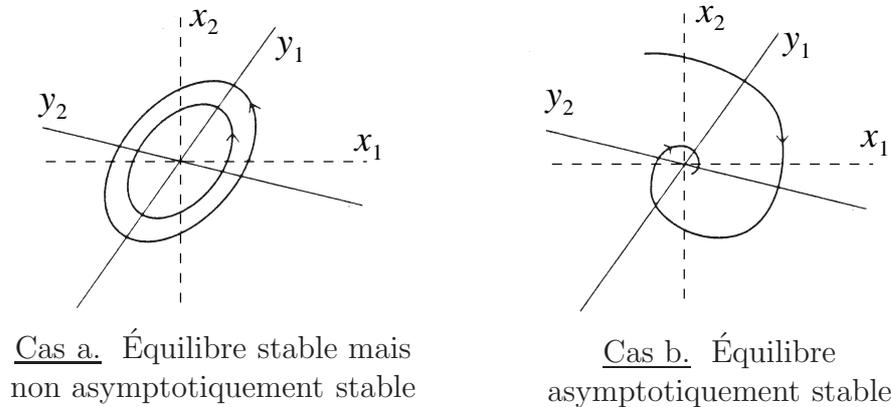


FIGURE 2.2 – Exemples de portraits de phase stables pour $f(x) = Ax$ avec $A \in M_2(\mathbb{R})$.

l'origine est un équilibre stable de $x' = Ax$ si et seulement si toutes les valeurs propres de A sont de partie réelle négative ou nulle et si pour toute valeur propre de partie réelle nulle, les multiplicités algébrique et géométrique coïncident, (c'est-à-dire $\mathbb{R}^n = E^s + E^c$ et $A|_{E^c}$ est diagonalisable dans \mathbb{C}).

Le cas affine

Considérons maintenant un champ de vecteurs affine $f(x) = Ax + b$ sur \mathbb{R}^n , où $A \in M_n(\mathbb{R})$ est une matrice et $b \in \mathbb{R}^n$ un vecteur. Un équilibre de l'équation

$$x'(t) = Ax(t) + b$$

est un point x_0 qui vérifie $Ax_0 + b = 0$ (noter qu'un tel point n'existe que si $b \in \text{Im } A$). En remplaçant b par $-Ax_0$, on réécrit l'équation différentielle sous la forme

$$\frac{d}{dt}(x(t) - x_0) = A(x(t) - x_0).$$

Ainsi la stabilité et la stabilité asymptotique d'un équilibre de l'équation affine $x'(t) = Ax(t) + b$ sont équivalentes respectivement à celles de l'origine pour l'équation linéaire $y'(t) = Ay(t)$.

2.3.2 La stabilité par la linéarisation

Soit x_0 un équilibre de l'équation différentielle (2.20). Nous allons montrer dans les deux théorèmes suivants que l'étude des valeurs propres de la matrice $Df(x_0)$ permet souvent de caractériser la stabilité de l'équilibre.

Théorème 2.32. *Si toutes les valeurs propres de $Df(x_0)$ sont de partie réelle strictement négative, alors x_0 est un équilibre asymptotiquement stable.*

Remarque. Contrairement au cas des équations linéaires, la condition du théorème est suffisante mais pas nécessaire. Prenons par exemple l'équation $y'(t) = -y^3(t)$ dans \mathbb{R} . L'équilibre 0 ne satisfait pas la condition du théorème puisque $Df(0) = 0$. En revanche c'est un équilibre asymptotiquement stable puisque la solution valant $y_0 \neq 0$ en $t = 0$ est

$$y(t) = \frac{\text{signe}(y_0)}{\sqrt{2t + \frac{1}{y_0^2}}}, \quad t \geq 0,$$

qui est décroissante et converge vers 0 quand $t \rightarrow +\infty$.

*PREUVE.

▷ Pour simplifier, on se ramène par translation à $x_0 = 0$. D'après l'hypothèse, il existe $\alpha > 0$ tel que $-\alpha$ est strictement supérieur à la partie réelle de toute valeur propre de $Df(0)$. D'après un résultat classique d'algèbre linéaire, il existe alors un produit scalaire $\langle \cdot, \cdot \rangle_\alpha$ sur \mathbb{R}^n tel que

$$\langle Df(0)x, x \rangle_\alpha \leq -\alpha \|x\|_\alpha^2, \quad \forall x \in \mathbb{R}^n,$$

où $\|\cdot\|_\alpha$ est la norme associée au produit scalaire $\langle \cdot, \cdot \rangle_\alpha$ (le résultat est clair quand $Df(0)$ est diagonalisable, sinon il se montre en utilisant la décomposition de Jordan).

Or, par définition de la différentielle,

$$\langle f(x), x \rangle_\alpha = \langle Df(0)x, x \rangle_\alpha + o(\|x\|_\alpha^2).$$

Ainsi, pour x suffisamment proche de 0, disons $\|x\| < \delta$, on obtient

$$\langle f(x), x \rangle_\alpha \leq -\frac{\alpha}{2} \|x\|_\alpha^2.$$

▷ Prenons maintenant un point $v \neq 0$ vérifiant $\|v\| < \delta$ et notons $x(t) = \phi_t(v)$ la solution issue de v . On peut choisir un temps $t_0 > 0$ suffisamment petit pour que $\|x(t)\| < \delta$ pour tout $t \in [0, t_0]$. La fonction $t \mapsto \|x(t)\|_\alpha$ est dérivable (car $x(t) \neq 0 \forall t$), et

$$\frac{d}{dt} \|x(t)\|_\alpha = \frac{\langle x'(t), x(t) \rangle_\alpha}{\|x(t)\|_\alpha} = \frac{\langle f(x(t)), x(t) \rangle_\alpha}{\|x(t)\|_\alpha} \leq -\frac{\alpha}{2} \|x(t)\|_\alpha.$$

Ceci implique d'abord que $\|x(t)\|_\alpha$ est décroissante : $x(t)$ reste donc confinée dans le compact $\|x\|_\alpha \leq \|v\|_\alpha$, ce qui entraîne que $x(\cdot)$ est définie pour tout $t > 0$ (proposition 2.4). D'autre part, d'après le lemme de Gronwall,

$$\|x(t)\|_\alpha \leq e^{-\frac{\alpha}{2}t} \|v\|_\alpha.$$

Finalement, on a montré que, si $\|v\| < \delta$, alors $\phi_t(v)$ reste dans $B(0, \delta)$ et tend vers 0, ce qui montre que 0 est un équilibre asymptotiquement stable. □

Théorème 2.33. *Si x_0 est un équilibre stable, alors toutes les valeurs propres de $Df(x_0)$ sont de partie réelle négative ou nulle.*

On utilisera généralement la contraposée de ce théorème : *si $Df(x_0)$ a au moins une valeur propre de partie réelle strictement positive, alors l'équilibre x_0 n'est pas stable.*

Il est important de noter que les réciproques des théorèmes 2.32 et 2.33 sont fausses, comme le montre l'exemple ci-dessous. La stabilité d'un équilibre n'est donc pas forcément déterminée par le linéarisé. Nous allons ensuite introduire une classe d'équilibres pour lesquels les réciproques des théorèmes 2.32 et 2.33 sont vérifiées.

Exemple. Considérons deux équations différentielles dans \mathbb{R}^2 ,

$$x' = f(x) = \begin{pmatrix} x_2 - x_1(x_1^2 + x_2^2) \\ -x_1 - x_2(x_1^2 + x_2^2) \end{pmatrix} \quad \text{et} \quad x' = g(x) = \begin{pmatrix} x_2 + x_1(x_1^2 + x_2^2) \\ -x_1 + x_2(x_1^2 + x_2^2) \end{pmatrix},$$

où $x = (x_1, x_2)$. Ces deux équations ont pour unique équilibre 0. Leurs linéarisés en 0 sont égaux,

$$Df(0) = Dg(0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

et ont pour valeurs propres $\pm i$, dont la partie réelle est évidemment nulle ! Cependant l'équilibre 0 est asymptotiquement stable dans le premier cas alors qu'il n'est pas stable dans le deuxième.

En effet, posons $\rho(x) = x_1^2 + x_2^2$. Si $x(\cdot)$ est une solution de l'équation $x' = f(x)$, alors

$$\frac{d}{dt}\rho(x(t)) = 2(x_1x_1' + x_2x_2') = -2\rho^2(x(t)).$$

Ainsi $\rho(x(t)) = \|x(t)\|^2$ est décroissant et tend vers 0 quand $t \rightarrow +\infty$, ce qui implique que 0 est asymptotiquement stable pour l'équation $x' = f(x)$.

De même, si $x(\cdot)$ est une solution de l'équation $x' = g(x)$, on obtient

$$\frac{d}{dt}\rho(x(t)) = 2\rho^2(x(t)).$$

Dans ce cas, $\rho(x(t)) = \|x(t)\|^2$ tend vers l'infini en temps fini (phénomène d'explosion), ce qui implique que l'équilibre 0 n'est pas stable pour l'équation $x' = g(x)$.

Équilibres hyperboliques

Définition 2.15. Un équilibre x_0 est dit *hyperbolique* si toutes les valeurs propres de $Df(x_0)$ ont une partie réelle non nulle.

Les équilibres hyperboliques jouent un rôle important en pratique puisque, comme nous l'avons vu à la fin de la section 2.2.4, la classe des matrices hyperboliques est ouverte et dense dans $M_n(\mathbb{R})$.

D'après les deux théorèmes précédents, la stabilité d'un équilibre hyperbolique x_0 est totalement caractérisée par le signe des parties réelles des valeurs propres de $Df(x_0)$.

Corollaire 2.34. *Un équilibre hyperbolique est soit asymptotiquement stable (si les valeurs propres de $Df(x_0)$ sont toutes de partie réelle négative), soit non stable.*

Ainsi, un équilibre hyperbolique x_0 est stable (resp. asymptotiquement stable) si et seulement si 0 est un équilibre stable (resp. asymptotiquement stable) pour l'équation linéarisée en x_0 ,

$$y'(t) = Df(x_0) \cdot y(t). \quad (2.21)$$

On a en fait beaucoup plus : les portraits de phase du système et de son linéarisé ont la même allure car ils sont topologiquement équivalents.

Théorème 2.35 (Théorème d'Hartman-Grobmann). *Soit x_0 un équilibre hyperbolique. Notons $\phi_t^L : y \mapsto e^{tDf(x_0)}y$ le flot du linéarisé en x_0 . Alors il existe un homéomorphisme $h : V_{x_0} \rightarrow V_0$, où V_{x_0} et V_0 sont des voisinages respectivement de x_0 et 0 dans \mathbb{R}^n , tel que*

$$\phi_t^L(h(x)) = h(\phi_t(x)),$$

partout où ces expressions ont un sens.

2.3.3 Fonctions de Lyapunov

Il existe une autre approche que la linéarisation pour obtenir des résultats de stabilité **qui ne nécessite pas une connaissance explicite du flot**. Commençons par donner un exemple qui illustre l'idée générale.

Champs de gradient

Un *champ de gradient* est un champ de vecteurs de la forme

$$f(x) = -\nabla V(x),$$

où $V : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction de classe C^2 (de façon à ce que f soit de classe C^1). Rappelons que $\nabla V(x)$ désigne le *gradient de V en x* , c'est-à-dire l'unique vecteur de \mathbb{R}^n vérifiant

$$DV(x) \cdot v = \langle \nabla V(x), v \rangle, \quad \forall v \in \mathbb{R}^n.$$

Dans l'espace euclidien \mathbb{R}^n , en coordonnées, $\nabla V(x) = (\frac{\partial V}{\partial x_1}(x), \dots, \frac{\partial V}{\partial x_n}(x))$.

Un équilibre x_0 de ce champ de vecteur est un point critique de V , *i.e.* $\nabla V(x_0) = 0$. Un équilibre peut donc être un minimum local, un maximum local, ou un point selle, mais nous allons voir que seuls les minima locaux peuvent être des équilibres stables.

La dynamique associée à un champ de gradient possède en effet une propriété qui la rend assez simple. Si $x(\cdot)$ est une solution de l'équation différentielle $x'(t) = -\nabla V(x(t))$, alors

$$\frac{d}{dt} [V(x(t))] = DV(x(t)) \cdot x'(t) = -\|\nabla V(x(t))\|^2, \quad (2.22)$$

pour tout t dans l'intervalle de définition de $x(\cdot)$. Ainsi $V(x(t))$ est soit constante, et dans ce cas $x(t) \equiv x_0$ est un point critique, soit strictement décroissante. Intuitivement,

ceci signifie que toute solution tend à se rapprocher d'un minimum, et donc (nous le montrerons plus loin) que :

- si un équilibre x_0 n'est pas un minimum local (*i.e.* x_0 est un maximum local ou un point selle), alors x_0 n'est pas un équilibre stable ;
- si x_0 est un minimum local strict, alors x_0 est un équilibre stable.

Fonctions de Lyapunov

Considérons maintenant une équation différentielle autonome quelconque

$$x'(t) = f(x(t)), \quad (2.20)$$

associée à un champ de vecteurs $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ de classe C^1 . L'exemple des champs de gradient suggère d'introduire la définition suivante.

Définition 2.16. Soient x_0 un équilibre de (2.20), $U \subset \Omega$ un voisinage de x_0 et $L : U \rightarrow \mathbb{R}$ une fonction continue. On dit que L est une *fonction de Lyapunov locale pour (2.20) en x_0* si :

- (a) $L(x_0) = 0$ et $L(x) > 0$ pour $x \neq x_0$ (*i.e.* x_0 est un minimum strict de L sur U) ;
- (b) la fonction $t \mapsto L(\phi_t(x))$ est décroissante.

Si de plus L satisfait

- (c) pour $x \neq x_0$, la fonction $t \mapsto L(\phi_t(x))$ est strictement décroissante,

on dit que L est une fonction de Lyapunov locale *stricte* pour (2.20) en x_0 .

Enfin, si U est égal à tout l'espace d'état, on peut remplacer dans ce qui précède "locale" par "globale".

Quand L est de classe C^1 , on peut remplacer les conditions (b) et (c) respectivement par les conditions suivantes, qui sont plus fortes mais plus simples à vérifier que (b) et (c) :

- (b)' $\langle \nabla L(x), f(x) \rangle \leq 0$ pour tout $x \in U$;
- (c)' $\langle \nabla L(x), f(x) \rangle < 0$ pour tout $x \in U$, $x \neq x_0$.

Une fonction de Lyapunov est donc une sorte de fonction d'énergie qui décrit le long des trajectoires.

Théorème 2.36. *Supposons que l'équation différentielle (2.20) admette L comme fonction de Lyapunov locale en un équilibre x_0 . Alors x_0 est un équilibre stable.*

- (Loc) *Si de plus L est stricte, alors x_0 est localement asymptotiquement stable.*
- (Glob) *Si de plus L est globale, stricte et L tend vers l'infini lorsque x tend vers l'infini, alors x_0 est globalement asymptotiquement stable.*

*PREUVE.

▷ Soit $L : U \rightarrow \mathbb{R}$ la fonction de Lyapunov. Quitte à remplacer U par une boule fermée centrée en x_0 , on le suppose compact. Pour tout $\varepsilon > 0$, l'ensemble $U_\alpha = \{x \in U : L(x) < \alpha\}$ est inclus dans la boule ouverte $B(x_0, \varepsilon)$ pour $\alpha > 0$ suffisamment petit (en effet, sinon il existe une suite de points x_n en dehors de $B(x_0, \varepsilon)$ vérifiant $\lim L(x_n) = 0$; U étant compact, x_n a alors un point d'accumulation $\bar{x} \neq x_0$, qui doit vérifier $L(\bar{x}) = 0$, ce qui est impossible d'après la propriété (a)). Or, d'après la propriété (b) de L , pour tout $x \in U_\alpha$, la solution $\phi_t(x)$ est confinée dans U_α ; ceci montre la stabilité de l'équilibre x_0 .

▷ Supposons maintenant que L est une fonction de Lyapunov locale et stricte. Considérons un point $x \in U_\alpha$ différent de x_0 . La fonction $t \mapsto L(\phi_t(x))$ étant strictement décroissante et minorée par 0, elle a une limite ℓ quand $t \rightarrow +\infty$. D'autre part, U étant compact, il existe une suite $t_n, t_n \rightarrow +\infty$, telle que $\phi_{t_n}(x)$ est convergente. Notons \bar{x} la limite de cette dernière suite. Par continuité de L , \bar{x} vérifie :

$$L(\bar{x}) = \lim_{t_n \rightarrow \infty} L(\phi_{t_n}(x)) = \lim_{t \rightarrow \infty} L(\phi_t(x)) = \ell.$$

De plus, pour tout $s > 0$, on a :

$$L(\phi_s(\bar{x})) = \lim_{t_n \rightarrow \infty} L(\phi_{s+t_n}(x)) = \ell,$$

ce qui montre que $s \mapsto L(\phi_s(\bar{x})) \equiv L(\bar{x})$ n'est pas décroissante, et donc, d'après la propriété (c) de L , que $\bar{x} = x_0$.

Ainsi, le seul point d'accumulation de $\phi_t(x)$ est x_0 , ce qui montre que $\lim_{t \rightarrow \infty} \phi_t(x) = x_0$. L'équilibre x_0 est donc asymptotiquement stable.

On adapte sans peine cette preuve au cas (Glob) en remarquant que les ensembles $L_c := \{x \in \Omega : L(x) \leq c\}$, $c > 0$, sont compacts.

□

Remarque. Pour un équilibre asymptotiquement stable x_0 , on appelle *bassin d'attraction* l'ensemble des points $x \in \Omega$ tels que $\phi_t(x) \rightarrow x_0$ quand $t \rightarrow +\infty$. Par définition de la stabilité asymptotique, le bassin d'attraction contient un voisinage de x_0 . Une question importante en pratique est de déterminer la taille de ce bassin, voire le bassin lui-même.

Le domaine de définition d'une fonction de Lyapunov stricte, si il en existe une, donne des éléments de réponse à cette question. Supposons par exemple que $\Omega = \mathbb{R}^n$ et que L soit une fonction de Lyapunov vérifiant les hypothèses du cas (Glob). Alors, le bassin d'attraction de x_0 est \mathbb{R}^n tout entier.

Plus généralement, si $L : U \rightarrow \mathbb{R}$ une fonction de Lyapunov stricte en x_0 et $P \subset U$ un sous-ensemble fermé de \mathbb{R}^n positivement invariant par le flot (*i.e.* $\phi_t(P) \subset P$ pour tout $t \geq 0$), alors P est inclus dans le bassin d'attraction de x_0 . Par exemple, les ensembles $\{x \in U : L(x) \leq \alpha\}$ pour α suffisamment petit sont fermés et positivement invariants (cf. Définition 2.6), donc inclus dans le bassin d'attraction.

Dans de nombreuses applications, il n'est pas possible d'avoir une fonction de Lyapunov stricte c'est-à-dire vérifiant la condition (c)'. On a alors le résultat suivant, appelé *principe d'invariance de Lasalle*.

Théorème 2.37. *Supposons que l'équation différentielle (2.20) admette $L : U \rightarrow \mathbb{R}_+$ comme fonction de Lyapunov locale en un équilibre x_0 . Notons D_U , le sous-ensemble de U défini par*

$$D_U := \{x \in U, \quad DL(x).f(x) = 0\}.$$

Alors,

- (Loc) toutes les trajectoires restant dans U convergent asymptotiquement vers le plus grand ensemble invariant (cf. Définition 2.6) contenu dans l'intersection de U et D_U .
- (Glob) Si de plus L est globale (i.e. $U = \Omega$) et tend vers l'infini lorsque x tend vers l'infini, toutes les trajectoires sont définies sur \mathbb{R}_+ et convergent asymptotiquement vers le plus grand ensemble invariant contenu dans D_U .

Le principe d'invariance consiste simplement à écrire le système surdéterminé suivant,

$$x' = f(x), \quad D_x L(x) = 0, \quad x \in U,$$

système caractérisant le plus grand ensemble invariant contenu dans l'intersection de U et D_U .

Exemples d'application

a. Champs de gradient Reprenons le cas d'un champ de gradients $f(x) = -\nabla V(x)$. Supposons que x_0 soit un minimum local strict de V , c'est-à-dire l'unique minimum de V sur un voisinage U de x_0 . La fonction V restreinte à U est bien une fonction de Lyapunov en x_0 , la propriété (b)' étant toujours satisfaite d'après la formule (2.22). Donc x_0 est bien un équilibre stable.

b. Champ de force conservatif Considérons un objet de masse m soumis à une force dérivant d'un potentiel $V(x)$. L'évolution de l'état $x \in \mathbb{R}^n$ de l'objet au cours du temps est régi par le principe fondamental de la dynamique :

$$mx''(t) = -\nabla V(x(t)),$$

que l'on réécrit

$$\begin{pmatrix} x \\ x' \end{pmatrix}'(t) = \begin{pmatrix} x'(t) \\ -\frac{1}{m}\nabla V(x(t)) \end{pmatrix}.$$

Autrement dit, (x, x') est solution de l'équation différentielle du 1er ordre dans \mathbb{R}^{2n}

$$(x, v)'(t) = f((x, v)(t)), \quad \text{où } f(x, v) = \left(v, -\frac{1}{m}\nabla V(x)\right).$$

Un équilibre de cette équation différentielle est un point $(x_0, 0) \in \mathbb{R}^{2n}$ où x_0 est un point critique du potentiel $V(x)$.

Supposons que x_0 soit un minimum local strict de V et cherchons une fonction de Lyapunov. Essayons à partir de l'énergie totale du système :

$$E(x, v) = \frac{1}{2}m\|v\|^2 + V(x).$$

En posant $L(x, v) = E(x, v) - V(x_0)$, on obtient une fonction qui satisfait la propriété (a) d'une fonction de Lyapunov. De plus, comme $\nabla L(x, v) = (\nabla V(x), mv)$, on obtient

$$\langle \nabla L(x, v), f(x, v) \rangle \equiv 0,$$

(c'est la conservation de l'énergie!), c'est-à-dire la propriété (b)'. La fonction L est donc bien une fonction de Lyapunov en $(x_0, 0)$, ce qui montre le résultat bien connu (théorème de Lagrange) :

si l'énergie potentielle $V(x)$ a un minimum local strict en x_0 , l'équilibre $(x_0, 0)$ est stable.

Remarques.

- L'équilibre $(x_0, 0)$ ne peut pas être asymptotiquement stable : la fonction de Lyapunov $L(x, v)$ étant constante le long d'une solution $(x, v)(\cdot) \not\equiv (x_0, 0)$, elle ne peut tendre vers 0, ce qui implique que $(x, v)(t)$ ne peut pas tendre vers $(x_0, 0)$.
- L'approche par linéarisation n'aurait pas permis de conclure ici car $(x_0, 0)$ n'est pas un équilibre hyperbolique (exercice : le montrer).

c. Champ de vecteur linéaire Considérons une équation différentielle linéaire

$$x'(t) = Ax(t), \quad x \in \mathbb{R}^n,$$

et supposons que toutes les valeurs propres de A sont de partie réelle strictement négative. On a vu (proposition 2.31) que dans ce cas l'origine est un équilibre (globalement) asymptotiquement stable. Cherchons une fonction de Lyapunov pour cette équation en 0. Le théorème suivant va nous en fournir.

Théorème 2.38. *Les propositions suivantes sont équivalentes.*

- (i) *A est Hurwitz (i.e. toutes ses valeurs propres sont de partie réelle strictement négative);*
- (ii) *il existe une matrice $P \in M_n(\mathbb{R})$ telle que $P = P^T > 0$ (c'est-à-dire que P est symétrique réelle et définie positive) et*

$$A^T P + PA < 0;$$

- (iii) *Pour toute matrice $Q = Q^T > 0$, il existe une unique matrice $P = P^T > 0$ solution de l'équation de Lyapunov*

$$A^T P + PA = -Q.$$

PREUVE.

▷ Il suffit de montrer les implications (ii) \Rightarrow (i) et (i) \Rightarrow (iii).

Commençons par (ii) \Rightarrow (i). Soit $P = P^T > 0$ telle que $A^T P + PA = -Q$. Notons $\alpha = \frac{\lambda_m(Q)}{\lambda_M(P)} > 0$ avec $\lambda_m(Q)$ et $\lambda_M(P)$ respectivement la plus petite des valeurs propres de Q et la plus grande des valeurs propres de P . Considérons la fonction de Lyapunov $L(x) := x^T P x$. Alors,

$$\dot{L}(x) = 2x^T A P x = x^T (A^T P + PA)x = -x^T Q x \leq -\lambda_m(Q) x^T x \leq -\beta L(x).$$

On en conclut que $L(x(t)) \leq e^{-\beta t} L(x(0))$ (pourquoi?) et donc (i).

Montrons maintenant (i) \Rightarrow (iii). Supposons que A est Hurwitz. Pour $t \geq 0$, on pose

$$P(t) := \int_0^t e^{sA^T} Q e^{sA} ds.$$

Remarquons que $P(t) > 0$ pour $t > 0$ et $t \mapsto P(t)$ est une fonction strictement croissante. Comme A est Hurwitz, l'intégrale généralisée

$$P = \int_0^\infty e^{sA^T} Q e^{sA} ds,$$

est convergente et on en conclue que $P(\cdot)$ admet P comme limite lorsque t tend vers l'infini. De plus, $P(\cdot)$ vérifie l'EDO suivante

$$\dot{P}(t) = -(A^T P(t) + P(t)A + Q).$$

La limite P est nécessairement point d'équilibre de cette EDO (pourquoi?) et donc satisfait l'équation de Lyapunov. Pour l'unicité, on procède comme suit. Soit $M > 0$ une solution de l'équation de Lyapunov. On multiplie alors l'équation de Lyapunov vérifiée par M à gauche par e^{sA^T} et à droite par e^{sA} . On remarque que

$$e^{sA^T} A^T M e^{sA} + e^{sA^T} M A e^{sA} = \frac{d}{ds} (e^{sA^T} M e^{sA}).$$

On intègre entre 0 et l'infini et on obtient $M = P$.

□

d. Critère de Hurwitz Bien qu'on ne puisse pas exprimer en général les racines d'un polynôme de degré $n \geq 5$ en fonction de ses coefficients, il existe des conditions algébriques ne portant **QUE** sur les coefficients d'un polynôme P qui sont nécessaires et suffisantes pour que toute racine de P soit de partie réelle strictement négative (par abus de langage, un tel polynôme est dit Hurwitz). L'une dde ces conditions est *le critère de Hurwitz* : on considère le polynôme complexe :

$$P(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n.$$

On pose $a_{n+1} = a_{n+2} = \dots = a_{2n-1} = 0$. On définit la matrice carrée d'ordre n :

$$H = \begin{pmatrix} a_1 & a_3 & a_5 & \cdots & \cdots & a_{2n-1} \\ a_0 & a_2 & a_4 & \cdots & \cdots & a_{2n-2} \\ 0 & a_1 & a_3 & \cdots & \cdots & a_{2n-3} \\ 0 & a_0 & a_2 & \cdots & \cdots & a_{2n-4} \\ 0 & 0 & a_1 & \cdots & \cdots & a_{2n-5} \\ \vdots & \vdots & \ddots & & & \vdots \\ 0 & 0 & 0 & * & \cdots & a_n \end{pmatrix},$$

où $*$ = a_0 ou a_1 selon la parité de n .

Soient $(H_i)_{i \in \{1, \dots, n\}}$ les mineurs principaux de H , i.e.

$$H_1 = a_1, \quad H_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}, \quad H_3 = \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix}, \quad \dots, \quad H_n = \det H.$$

Proposition 2.39. *Si $a_0 > 0$, toute racine de P est de partie réelle strictement négative si et seulement si $H_i > 0$, pour tout $i \in \{1, \dots, n\}$.*

Remarque. Si $a_0 > 0$, on a :

- Si pour toute racine λ de P , on a $\operatorname{Re} \lambda \leq 0$, alors $a_k \geq 0$ et $H_k \geq 0$, pour tout $k \in \{1, \dots, n\}$.
- Si $n \leq 3$ et si $a_k \geq 0$ et $H_k \geq 0$, pour tout $k \in \{1, 2, 3\}$, alors toute racine λ de P vérifie $\operatorname{Re} \lambda \leq 0$.

Remarque. Une condition nécessaire de stabilité est donc, si $a_0 > 0$:

$$\forall k \in \{1, \dots, n\} \quad a_k \geq 0.$$

Mais cette condition n'est pas suffisante (donner un exemple).

e. Nouvelle preuve du théorème 2.32 Considérons un équilibre x_0 de l'équation différentielle

$$x'(t) = f(x(t)),$$

et supposons que toutes les valeurs propres de $Df(x_0)$ sont de partie réelle strictement négative. Nous allons montrer que l'équation admet une fonction de Lyapunov stricte en x_0 , ce qui donnera une nouvelle preuve du théorème 2.32.

Quitte à faire une translation, on suppose $x_0 = 0$ et on choisit $P = P^T > 0$ solution de l'équation de Lyapunov

$$Df(0)^T P + P Df(0) = -I_n.$$

On considère $L(x) := x^T P x$ qui est une fonction de Lyapunov pour l'équation linéaire $y'(t) = Df(0) \cdot y(t)$ (à vérifier). Remarquons d'autre part que

$$f(x) = Df(0) \cdot x + o(\|x\|).$$

En utilisant le paragraphe précédent, on obtient

$$\begin{aligned} \langle \nabla L(x), f(x) \rangle &= \langle \nabla L(x), Df(0) \cdot x \rangle + \langle \nabla L(x), o(\|x\|) \rangle \\ &= -\|x\|^2 + 2 \int_0^\infty \langle e^{sDf(0)} x, e^{sDf(0)} o(\|x\|) \rangle ds. \end{aligned}$$

Le terme dans la dernière intégrale est un $o(\|x\|^2)$, donc pour $\|x\|$ suffisamment petit, $x \neq 0$, on obtient

$$\langle \nabla L(x), f(x) \rangle \leq -\frac{1}{2}\|x\|^2 < 0,$$

c'est-à-dire la propriété (c)'. La fonction L est donc une fonction de Lyapunov stricte en $x_0 = 0$, ce qui montre que cet équilibre est asymptotiquement stable.

Chapitre 3

Commandabilité et observabilité des systèmes linéaires

3.1 Systèmes de commande

Un large pan de l'automatique est basé sur les équations différentielles : c'est l'approche par représentation d'état des systèmes à temps continu déterministes (les systèmes stochastiques s'appuyant sur les équations différentielles stochastiques). La situation est la suivante : on considère un système physique (par exemple un satellite, une voiture,...), décrit par son état $x(t)$ au temps t (par exemple la position et la vitesse), sur lequel on peut agir à tout moment au moyen d'une *commande ou contrôle* u (par exemple la poussée des moteurs dans le cas du satellite). On représente l'état par un vecteur de \mathbb{R}^n , la commande par un vecteur de \mathbb{R}^m , et on modélise l'évolution du vecteur $x(t)$ au cours du temps par un *système de commande* (ou équation différentielle commandée)

$$(\Sigma) : \quad x'(t) = f(t, x(t), u(t)), \quad t \in [0, \tau],$$

où τ est un temps positif.

Que signifie cette expression ? La fonction $u(t)$, $t \in [0, \tau]$, appelée *loi de commande*, est notre moyen d'action sur le système (Σ) : on va la choisir en fonction des objectifs à réaliser. À une loi de commande $u(\cdot)$ est associée une équation différentielle ordinaire

$$(\Sigma_u) : \quad x'(t) = f_u(t, x(t)), \quad t \in [0, \tau],$$

où on a noté $f_u(t, x) = f(t, x, u(t))$. Ainsi, une fonction $x(\cdot)$ est solution du système (Σ) si il existe une loi de commande $u(\cdot)$ telle que $x(\cdot)$ est solution de l'équation différentielle (Σ_u) .

Les principales questions que l'on est amené à se poser à propos du système (Σ) sont les suivantes.

Commandabilité Étant donné un état de départ $x_0 \in \mathbb{R}^n$, un état cible $v \in \mathbb{R}^n$ et un temps $t = \tau > 0$, est-il possible de trouver une commande $u(\cdot)$ qui amène le

système initialement en $x(0)$ à $t = 0$ en l'état v au temps $t = \tau$? On formule aussi la question précédente comme suit : est-il possible de *commander* le système de x_0 à v en temps τ ?

Planification de trajectoires A la question structurelle précédente, correspond le problème plus concret d'établir une procédure effective qui associe, à une paire d'états $x_0, v \in \mathbb{R}^n$ et un temps τ , une commande $u(\cdot)$ qui amène le système de $x(0)$ à v en temps $t = \tau$.

Stabilisation Est-il possible de construire une commande $u(\cdot)$ qui *stabilise asymptotiquement* le système (Σ) autour d'un équilibre x_0 , c'est-à-dire telle que, pour toutes conditions initiales $x(0)$, on ait

$$\lim_{t \rightarrow +\infty} x(t) = x_0.$$

Observabilité Afin de réaliser un objectif de commande (planification de trajectoire, stabilisation, etc...) et donc de choisir en conséquence une loi de commande appropriée, l'opérateur dispose d'une certaine information sur l'état du système x à l'instant t , celle-ci étant obtenue par le biais de mesures. Cependant, il n'est pas possible en général de mesurer (on dit *observer* en automatique) directement l'état $x(t)$ mais seulement une fonction $y(t)$ de l'état et de la commande :

$$y(t) = g(x(t), u(t), t).$$

Il s'agit alors de "reconstruire" l'état $x(\cdot)$ à partir de la *sortie* $y(\cdot)$. La question d'observabilité est alors la suivante : la connaissance de $y(t)$ et de $u(t)$ pour tout $t \in [0, \tau]$ permet-elle de déterminer l'état $x(\cdot)$ pour tout $t \in [0, \tau]$ (ou, ce qui est équivalent, l'état initial $x(0)$)?

Les techniques que nous avons introduites lors de l'étude des équations différentielles linéaires autonomes vont nous permettre de répondre à ces questions dans le cadre de l'*automatique linéaire autonome* (ou stationnaire, ou encore invariante par rapport au temps).

Dans tout ce chapitre, nous supposons donc que le système (Σ) est linéaire autonome (par rapport à (x, u)), c'est-à-dire de la forme :

$$(\Sigma) : \quad x'(t) = Ax(t) + Bu(t), \quad t \in [0, \tau], \quad (3.1)$$

où A est une matrice (carrée) de $M_n(\mathbb{R})$ et B une matrice (pas nécessairement carrée) de $M_{n,m}(\mathbb{R})$, avec n, m entiers positifs. Si $m = 1$, le système est dit *mono-entrée* et sinon *multi-entrée*.

Par ailleurs, lorsque nous aborderons les questions d'observabilité, nous supposons que la sortie y est un vecteur de \mathbb{R}^p , p entier positif et que y est aussi linéaire par rapport à (x, u) , plus simplement égale à

$$y(t) = Cx(t), \quad t \in [0, \tau], \quad (3.2)$$

avec C une matrice $M_{p,n}(\mathbb{R})$. Dans ce cas, le système de commande linéaire autonomes sera défini par les deux équations (3.1) et (3.2).

Les lois de commande $u(\cdot)$ seront supposées *continues par morceaux*, définies sur l'intervalle $[0, \tau]$ et à valeurs dans \mathbb{R}^m . Nous aurions pu faire d'autres hypothèses sur la fonction $u(\cdot)$, la supposer seulement mesurable par exemple ou encore imposer que ses valeurs soient bornées (ce qui est raisonnable si u représente une force). Ces hypothèses interviennent naturellement dans des problèmes d'automatique, mais nous n'en parlerons pas ici : notre but est de montrer dans un cadre aussi simple que possible comment les techniques d'équations différentielles s'appliquent à ces problèmes.

Le point de départ de l'étude des systèmes linéaires autonomes est la formule dite *de variation de la constante* :

Proposition 3.1. *Soient $u(\cdot)$ une commande et $x_0 \in \mathbb{R}^n$. L'unique solution de $x'(t) = Ax(t) + Bu(t)$ valant x_0 à l'instant $t = 0$ est*

$$x(t) = e^{tA}x_0 + \int_0^t e^{(t-s)A}Bu(s)ds.$$

PREUVE.

▷ Supposons qu'il existe une solution $x(\cdot)$ de $x'(t) = Ax(t) + Bu(t)$ telle que $x(0) = x_0$. Posons $y(t) = e^{-tA}x(t)$ et dérivons $y(\cdot)$ par rapport à t :

$$y'(t) = -Ae^{-tA}x(t) + e^{-tA}(Ax(t) + Bu(t)) = e^{-tA}Bu(t).$$

En intégrant entre 0 et t , on obtient

$$y(t) = y(0) + \int_0^t e^{-sA}Bu(s)ds,$$

et donc la conclusion, puisque $x(t) = e^{tA}y(t)$ et $y(0) = x(0) = x_0$. □

Remarque. Notons en particulier que, si $x(0) = 0$,

$$x(t) = \int_0^t e^{(t-s)A}Bu(s)ds, \tag{3.3}$$

et que cette expression dépend linéairement de la loi de commande $u(\cdot)$.

Remarque. Pour ce qui est de l'observabilité des systèmes linéaires, la formule de la proposition 3.1 montre que la "reconstruction" d'une trajectoire $x(\cdot)$ avec la seule connaissance d'une sortie $y(\cdot)$ (et bien sur de la commande $u(\cdot)$) consiste en fait à retrouver la condition initiale x_0 .

3.2 Commandabilité

Soit (Σ) le système de commande linéaire autonome (3.1). Étant donné $x_0 \in \mathbb{R}^n$, on dit qu'un état $v \in \mathbb{R}^n$ est *atteignable en temps τ* à partir de x_0 (par le système (Σ)) si il existe une loi de commande $u : [0, \tau] \rightarrow \mathbb{R}^m$ telle que $x(\tau) = v$, $x(\cdot)$ étant la solution de (Σ_u) satisfaisant $x(0) = x_0$. On note $\mathcal{A}(\tau, x_0)$ l'ensemble des états atteignables à partir de x_0 en temps τ , c'est-à-dire

$$\mathcal{A}(\tau, x_0) := \left\{ x(\tau) : \begin{array}{l} x(\cdot) \text{ solution de } (\Sigma) \\ \text{t.q. } x(0) = x_0 \end{array} \right\}.$$

Il résulte de la remarque ci-dessus et de la proposition 3.1 que l'ensemble $\mathcal{A}(\tau, 0)$ est un espace vectoriel, et que l'ensemble $\mathcal{A}(\tau, x_0)$ est l'espace affine $e^{\tau A}x_0 + \mathcal{A}(\tau, 0)$. L'ensemble des points atteignables à partir de x_0 est donc complètement caractérisé par l'ensemble $\mathcal{A}_\tau := \mathcal{A}(\tau, 0)$.

Définition 3.1. On dit que le système (Σ) est *commandable en temps τ* si $\mathcal{A}_\tau = \mathbb{R}^n$, ou, de façon équivalente, si tout état de \mathbb{R}^n est atteignable en temps τ à partir de n'importe quel autre.

Nous allons chercher maintenant à caractériser algébriquement la commandabilité. Ceci passe par la détermination de l'ensemble \mathcal{A}_τ .

Théorème 3.2. *L'espace \mathcal{A}_τ est égal à l'image de la matrice $(n \times nm)$*

$$\mathcal{C}(A, B) := [B \quad AB \quad \cdots \quad A^{n-1}B],$$

dite matrice de commandabilité.

Remarque. L'image de $\mathcal{C}(A, B)$ est l'espace vectoriel $\mathcal{R}(A, B) \subset \mathbb{R}^n$ engendré par les $A^i Bz$, $i \in \{0, \dots, n-1\}$, $z \in \mathbb{R}^m$:

$$\mathcal{R}(A, B) = \text{Vect}\{A^i Bz : i = 0, \dots, n-1, z \in \mathbb{R}^m\}.$$

La première conséquence de ce résultat est que \mathcal{A}_τ est indépendant de τ . Notons que ce ne serait évidemment pas le cas si nous avions choisi des commandes bornées. La deuxième conséquence est que la dimension de \mathcal{A}_τ est égale au rang de la matrice de commandabilité. On obtient ainsi un critère de commandabilité algébrique, et donc en général facile à vérifier.

Corollaire 3.3 (Critère de commandabilité de Kalman). *Le système (Σ) est commandable si et seulement si la matrice de commandabilité $\mathcal{C}(A, B)$ est de rang n .*

PREUVE.

▷ du Théorème 3.2 Montrons déjà que $\mathcal{A}_\tau \subset \mathcal{R}(A, B)$. Pour cela observons que, par définition, si v appartient à \mathcal{A}_τ il existe $u : [0, \tau] \rightarrow \mathbb{R}^m$ continue par morceaux telle que

$$v = \int_0^\tau e^{(\tau-s)A} B u(s) ds.$$

Le théorème de Cayley-Hamilton nous enseigne que le polynôme caractéristique de A annule A . Or ce polynôme est un polynôme normalisé (*i.e.* dont le coefficient de plus haut degré égale 1) de degré n . Ainsi A^n est combinaison linéaire de I, \dots, A^{n-1} et par conséquent pour tout entier $i \geq 0$, A^i est combinaison linéaire de I, \dots, A^{n-1} . Par conséquent pour tout $i \geq 0$, A^i laisse invariant l'espace vectoriel

$$\mathcal{R}(A, B) = \text{Vect}\{A^i B z : i = 0, \dots, n-1, z \in \mathbb{R}^m\}.$$

Mais pour tout $s \in [0, \tau]$, l'exponentielle $e^{(\tau-s)A}$ admet le développement

$$e^{(\tau-s)A} = I + (\tau-s)A + \dots + \frac{(\tau-s)^k A^k}{k!} + \dots,$$

et par conséquent $e^{(\tau-s)A}$ laisse également invariant l'espace $\mathcal{R}(A, B)$. Nous avons donc montré que $e^{(\tau-s)A} B u(s) \in \mathcal{R}(A, B)$ pour tout $s \in [0, \tau]$ et par conséquent

$$\int_0^\tau e^{(\tau-s)A} B u(s) ds \in \mathcal{R}(A, B).$$

Ainsi $\mathcal{A}_\tau \subset \mathcal{R}(A, B)$.

▷ Montrons l'inclusion réciproque. Il suffit pour cela de démontrer $\mathcal{A}_\tau^\perp \subset \mathcal{R}(A, B)^\perp$. Soit donc $w \in \mathbb{R}^n$ orthogonal à \mathcal{A}_τ ; le vecteur w est ainsi orthogonal à l'état \tilde{w} que l'on peut atteindre au temps τ par la commande

$$u(t) = B^T (e^{(\tau-s)A})^T w.$$

La formule (3.3) montre que

$$\tilde{w} = \int_0^\tau e^{(\tau-s)A} B B^T (e^{(\tau-s)A})^T w ds,$$

et par conséquent, puisque $\langle \tilde{w}, w \rangle = 0$ on obtient

$$0 = \langle w, \int_0^\tau e^{(\tau-s)A} B B^T (e^{(\tau-s)A})^T w ds \rangle = \int_0^\tau \left((e^{(\tau-s)A} B)^T w \right)^T \left((e^{(\tau-s)A} B)^T w \right) ds,$$

ce qui est équivalent à

$$\forall s \in [0, \tau], (e^{(\tau-s)A} B)^T w = 0.$$

Dérivons une fois, puis deux fois, ... cette égalité par rapport à t : il vient successivement

$$(e^{(\tau-s)A} A B)^T w = 0, \quad (e^{(\tau-s)A} A^2 B)^T w = 0, \quad \dots \quad (e^{(\tau-s)A} A^{n-1} B)^T w = 0,$$

soit, pour $s = \tau$,

$$B^T w = 0, \quad \dots \quad (A^{n-1}B)^T w = 0.$$

Ceci implique que, pour tout $j \in \{0, \dots, n-1\}$ et tout $z \in \mathbb{R}^m$,

$$0 = \langle z, (A^j B)^T w \rangle = \langle A^j B z, w \rangle,$$

c'est-à-dire $w \in \mathcal{R}(A, B)^\perp$. L'inclusion $\mathcal{R}(A, B) \subset \mathcal{A}_\tau$ est donc démontrée. □

Nous allons maintenant décrire ce qui se passe lorsque le rang r de la matrice de commandabilité $\mathcal{C}(A, B)$ est quelconque. Pour cela, nous étudions l'effet d'un changement de coordonnées linéaire sur la commandabilité d'un système.

Définition 3.2. Les systèmes de commande linéaires $\dot{x}_1 = A_1 x_1 + B_1 u_1$ et $\dot{x}_2 = A_2 x_2 + B_2 u_2$ sont dits *linéairement équivalents* s'il existe $P \in GL_n(\mathbb{R})$ tel que $A_2 = P A_1 P^{-1}$ et $B_2 = P B_1$.

Remarque. On a alors $x_2 = P x_1$.

Proposition 3.4. *La propriété de Kalman est intrinsèque, i.e.*

$$(B_2, A_2 B_2, \dots, A_2^{n-1} B_2) = P(B_1, A_1 B_1, \dots, A_1^{n-1} B_1),$$

et donc le rang de la matrice de Kalman est invariant par équivalence linéaire.

Considérons une paire (A, B) où $A \in \mathcal{M}_n(\mathbb{R})$ et $B \in \mathcal{M}_{n,m}(\mathbb{R})$.

Théorème 3.5 (Décomposition de Kalman). *La paire (A, B) est linéairement équivalente à une paire (A', B') de la forme*

$$A' = \begin{pmatrix} A'_1 & A'_2 \\ 0 & A'_3 \end{pmatrix}, \quad B' = \begin{pmatrix} B'_1 \\ 0 \end{pmatrix},$$

où $A'_1 \in \mathcal{M}_r(\mathbb{R})$, $B'_1 \in \mathcal{M}_{r,m}(\mathbb{R})$, r étant le rang de la matrice de Kalman de la paire (A, B) . De plus, la paire (A'_1, B'_1) est commandable.

Remarque. La décomposition précédente est aussi appelée décomposition du système en parties commandable et non commandable. Les valeurs propres de A sont appelées *pôles* du système, commandables pour les valeurs propres de A'_1 et non commandables pour celles de A'_3 .

PREUVE.

▷ Supposons que le rang r de la matrice de Kalman C de la paire (A, B) est strictement plus petit que n (sinon il n'y a rien à montrer). Le sous-espace

$$F = \text{Im } C = \text{Im } B + \text{Im } AB + \cdots + \text{Im } A^{n-1}B$$

est de dimension r , et d'après le théorème d'Hamilton-Cayley il est clairement invariant par A . Soit G un supplémentaire de F dans \mathbb{R}^n , et soient (f_1, \dots, f_r) une base de F , et (f_{r+1}, \dots, f_n) une base de G . Notons P la matrice de passage de la base (f_1, \dots, f_n) à la base canonique de \mathbb{R}^n . Alors, puisque F est invariant par A , on a :

$$A' = PAP^{-1} = \begin{pmatrix} A'_1 & A'_2 \\ 0 & A'_3 \end{pmatrix},$$

et d'autre part, puisque $\text{Im } B \subset F$, on a :

$$B' = PB = \begin{pmatrix} B'_1 \\ 0 \end{pmatrix}.$$

Enfin, on voit facilement que le rang de la matrice de Kalman de la paire (A'_1, B'_1) est égal à celui de la paire (A, B) . □

Dans le cas mono-entrée, il existe un changement de coordonnée très utile.

Théorème 3.6 (Forme compagne). *Si $m = 1$ et si la paire (A, B) est contrôlable, alors elle est linéairement équivalente à la paire (\tilde{A}, \tilde{B}) , où*

$$\tilde{A} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ -a_n & -a_{n-1} & \cdots & -a_1 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

La matrice \tilde{A} et la paire (\tilde{A}, \tilde{B}) sont respectivement appelées forme compagne de la matrice A et forme canonique commandable de la paire (A, B) .

Remarque. Si une matrice carrée est sous forme compagne, alors les coefficients de son polynôme caractéristique se lisent sur la dernière ligne de la matrice.

Remarque. Dans ces nouvelles coordonnées, le système est alors équivalent à l'équation différentielle scalaire d'ordre n :

$$x^{(n)}(t) + a_1 x^{(n-1)}(t) + \cdots + a_n x(t) = u(t).$$

PREUVE.

▷ Considérons $P_A(X)$ le polynôme caractéristique de A

$$P_A(X) := X^n + a_1X^{n-1} + \cdots + a_n.$$

Le changement de coordonnées sera donné par une matrice $n \times n$ inversible F de vecteur colonne (f_1, \dots, f_n) . Ces vecteurs sont définis récursivement comme suit,

$$f_n = b, f_{n-1} = Af_n + a_1f_n, \dots, f_1 = Af_2 + a_{n-1}f_n.$$

La famille (f_1, \dots, f_n) est bien une base de \mathbb{R}^n puisque :

$$\begin{aligned} \text{Vect } \{f_n\} &= \text{Vect } \{b\}, \\ \text{Vect } \{f_n, f_{n-1}\} &= \text{Vect } \{b, Ab\}, \\ &\vdots \\ \text{Vect } \{f_n, \dots, f_1\} &= \text{Vect } \{b, \dots, A^{n-1}b\} = \mathbb{R}^n. \end{aligned}$$

Il reste à vérifier que l'on a bien $Af_1 = -a_nf_n$:

$$\begin{aligned} Af_1 &= A^2f_2 + a_{n-1}Af_n \\ &= A^2(Af_3 + a_{n-2}f_n) + a_{n-1}Af_n \\ &= A^3f_3 + a_{n-2}A^2f_n + a_{n-1}Af_n \\ &\dots \\ &= A^n f_n + a_1A^{n-1}f_n + \cdots + a_{n-1}Af_n \\ &= -a_nf_n \end{aligned}$$

puisque d'après le théorème de Cayley-Hamilton, on a $A^n = -a_1A^{n-1} - \cdots - a_nI$. Dans la base (f_1, \dots, f_n) , la paire (A, b) prend la forme (\tilde{A}, \tilde{b}) . □

Remarque. Ce théorème admet la généralisation suivante, lorsque $m > 1$. Si la paire (A, B) est contrôlable, alors on peut la conjuguer à une paire (\tilde{A}, \tilde{B}) telle que :

$$\tilde{A} = \begin{pmatrix} \tilde{A}_1 & * & \cdots & * \\ 0 & \tilde{A}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \tilde{A}_s \end{pmatrix},$$

les matrices \tilde{A}_i étant des matrices sous forme compagne ; par ailleurs, il existe une matrice $G \in \mathcal{M}_{m,s}(\mathbb{R})$ telle que :

$$\tilde{B}G = \begin{pmatrix} \tilde{B}_1 \\ \vdots \\ \tilde{B}_s \end{pmatrix},$$

où tous les coefficients de chaque matrice \tilde{B}_i sont nuls, sauf celui de la dernière ligne, en i -ème colonne, qui est égal à 1.

3.3 Planification de trajectoires

Nous allons maintenant proposer une stratégie pour résoudre le problème de planification de trajectoires pour un système de commande linéaire autonome commandable. Pour cela, nous allons effectuer des transformations linéaires sur le système afin de l'amener en une forme particulière simple et pour laquelle la planification de trajectoire sera aisée.

3.3.1 Exemple

Effectuons la planification de trajectoires pour le système commandé de dimension un, $\dot{x} = u$, évidemment commandable. Etant donnés deux réels arbitraires, x_0, x_1 , il s'agit de trouver une commande u définie sur $[0, 1]$ telle que

$$x_1 = x_0 + \int_0^1 u(t)dt,$$

et donc plus simplement, de trouver pour tout $x \in \mathbb{R}$, une fonction u telle que $x = \int_0^1 u(t)dt$. Une fonction polynômiale (constante) suffit ici. La généralisation au système commandé $x^{(n)} = u$, avec x, u réels, est immédiate.

3.3.2 Forme de Brunovsky

Dans le cas général du système $\dot{x} = Ax + Bu$, l'idée est de transformer linéairement le système de telle sorte à le faire "ressembler" le plus à $x^{(n)} = u$.

Les transformations linéaires considérées ci-dessous sont plus générales que de simples changements de coordonnées.

Définition 3.3 (bouclage statique régulier). Soit $x_1 = Mx_2$, le changement de coordonnées défini sur \mathbb{R}^n par M matrice carrée inversible $n \times n$. Un bouclage statique régulier est défini par $u_1 = Kx_2 + Nu_2$ avec N matrice carrée inversible $m \times m$ et K une matrice $m \times n$. C'est un changement de variables sur la commande paramétré par l'état. De manière matricielle, il s'écrit

$$\begin{pmatrix} x_1 \\ u_1 \end{pmatrix} := \begin{pmatrix} M & 0 \\ K & N \end{pmatrix} \begin{pmatrix} x_2 \\ u_2 \end{pmatrix}.$$

Les systèmes de commande linéaires $\dot{x}_1 = A_1x_1 + B_1u_1$ et $\dot{x}_2 = A_2x_2 + B_2u_2$ sont dits *équivalents par changement d'état avec bouclage statique régulier* s'il existe $M \in GL_n(\mathbb{R})$, $N \in GL_m(\mathbb{R})$ et $K \in M_{m,n}(\mathbb{R})$, tels que $A_2 = M^{-1}A_1M + M^{-1}BK$ et $B_2 = M^{-1}B_1N$.

Théorème 3.7 (Forme de Brunovsky d'un système commandable). *Soit $\dot{x} = Ax + Bu$ un système linéaire commandable avec $x \in \mathbb{R}^n$ et $u \in \mathbb{R}^m$. Alors il existe un changement d'état avec bouclage statique régulier donné par $x = Mz$ et $u = Kz + Nv$ pour lequel il existe m coordonnées de z , notées y_1, \dots, y_m telles que*

$$(i) \quad z = (y_1, \dots, y_1^{(\alpha_1-1)}, \dots, y_m, \dots, y_m^{(\alpha_m-1)})^T ;$$

(ii) $\dot{x} = Ax + Bu$ devient

$$y_1^{(\alpha_1)} = v_1, \quad \dots \quad y_m^{(\alpha_m)} = v_m,$$

où $y_l^{(j)}$ désigne la dérivée temporelle d'ordre j de la fonction scalaire y_l , $1 \leq l \leq m$.

Les coordonnées y_1, \dots, y_m sont appelées **sorties de Brunovsky** du système.

Ecrivons la forme de Brunovsky d'un système mono-entrée, c-à-d avec une seule commande scalaire ($m = 1$). Dans ce cas, $B = b$ est un vecteur colonne. Il existe donc un changement d'état avec bouclage statique régulier qui transforme $\dot{x} = Ax + bu$ en $\dot{z}_i = z_{i+1}$ pour $1 \leq i \leq n$ et $\dot{z}_n = v$. La sortie de Brunovsky est alors $y = z_1$ et le système commandé s'écrit simplement $y^{(n)} = v$.

3.3.3 Application à la planification de trajectoires

Supposons que le système de commande $\dot{x} = Ax + Bu$ soit commandable et sous forme de Brunovsky. Grâce à la structure bloc-diagonale, le problème de planification en temps $T > 0$ se décompose en m problèmes à une seule commande : pour $1 \leq i \leq m$,

Aller de $(y_i(0), \dots, y_i^{(\alpha_i-1)}(0))^T$ à $(y_i(T), \dots, y_i^{(\alpha_i-1)}(T))^T$ le long de $y_i^{(\alpha_i)} = v_i$.

Le i ème problème de planification défini ci-dessus se résout simplement puisque les conditions sur les points initial et final représentent $2\alpha_i$ contraintes à satisfaire. Par exemple, il suffit de choisir v_i comme un polynôme de degré $2\alpha_i - 1$ et de déterminer ses coefficients.

3.3.4 Preuve du théorème 3.7 pour le cas mono-entrée

Dans le reste de ce paragraphe, on supposera $m = 1$ et donc $B = b$ vecteur colonne. On note $P_A(X) = X^n + a_1X^{n-1} + \dots + a_n$, le polynôme caractéristique de A et on suppose la matrice $n \times n$ inversible F met la paire (A, b) sous forme compagne, c-à-d que $(F^{-1}AF, F^{-1}b) = (\tilde{A}, \tilde{b})$, cf. Théorème 3.6. Si on a $Fy = x$, alors $\dot{y} = F^{-1}AFy + F^{-1}b$ avec

$$\dot{y}_1 = y_2, \quad \dots \quad \dot{y}_{n-1} = y_n,$$

et $\dot{y}_n = -a_n y_1 - \dots - a_1 y_n + u$. Si on définit une nouvelle commande $v := y_n$, on a bien un bouclage statique régulier sur la commande u . Le système est donc sous forme de Brunovsky avec y_1 comme sortie de Brunovsky.

3.4 Stabilisation

Utiliser des lois de commande $u(t)$ dépendant en général du temps s'appelle *commander en boucle ouverte* : la loi de commande est fixée au départ, à $t = 0$, et est appliquée indépendamment du comportement du système pour $t > 0$. Les limitations de ce type de lois de commande sont assez évidentes : la moindre erreur sur les données (la condition initiale par exemple) ne pourra pas être prise en compte. Par exemple une commande en boucle ouverte sur une voiture donnerait ceci : pour suivre une ligne droite, positionnez vos roues dans l'axe, tenez bien votre volant, et fermez les yeux. . .

Pour réguler le système, il faut faire appel à un autre type de loi de commande $u(t) = K(t, x(t))$, dite *commande en boucle fermée* ou *par retour d'état* : à tout instant t , on tient compte de l'état du système à cet instant pour déterminer la commande. Notez que ces lois ont aussi leur inconvénient : elles nécessitent la connaissance (quasi instantanée) de l'état $x(t)$, ce qui peut être impossible, ou très coûteux.

Dans le problème de stabilisation, le but est de construire une loi de commande par retour d'état qui amène le système à l'origine, quel que soit le point de départ. Nous allons chercher le retour d'état sous la forme d'une fonction linéaire indépendante du temps, c'est-à-dire $u(t) = Kx(t)$ avec $K \in M_{m,n}(\mathbb{R})$ (une telle loi est appelée *loi proportionnelle*).

Définition 3.4. Un système de commande (Σ) est *asymptotiquement stabilisable* par retour d'état proportionnel s'il existe une loi de commande $u(t) = Kx(t)$, avec $K \in M_{m,n}(\mathbb{R})$, telle que l'équation (Σ_u) soit asymptotiquement stable, c'est-à-dire que, pour toute condition initiale $x(0)$, la solution $x(t)$ de (Σ_u) tende vers 0 quand $t \rightarrow +\infty$.

Noter que l'intervalle de temps considéré est maintenant infini, *i.e.* $\tau = +\infty$.

Nous nous plaçons comme précédemment dans le cadre des systèmes linéaires autonomes c-à-d que (Σ) est de la forme (3.1). Dans ce cas, pour un retour d'état proportionnel $u(t) = Kx(t)$, l'équation différentielle (Σ_u) s'écrit

$$x'(t) = Ax(t) + BKx(t) = (A + BK)x(t).$$

Or nous savons (corollaire 2.27) qu'une telle équation différentielle est asymptotiquement stable si et seulement si la matrice $A + BK$ a toutes ses valeurs propres de parties réelles strictement négatives. Existe-t-il $K \in M_{m,n}(\mathbb{R})$ telle que la matrice $A + BK$ satisfasse cette condition ?

Le but de cette section est de répondre à cette question. Tout d'abord, remarquons qu'un changement de coordonnées permet de remplacer (A, B) par toute paire qui lui est linéairement équivalente. En effet, considérons $\tilde{A} = P^{-1}AP$ et $\tilde{B} = P^{-1}B$ avec P la

matrice de changement de coordonnées $x = Py$ avec P matrice $n \times n$ inversible. On a alors $\tilde{K} = KP$ et $P^{-1}(A + BK)P = \tilde{A} + \tilde{B}\tilde{K}$.

Sans perte en généralité, nous pouvons donc supposer que (A, B) est décomposée en parties commandable et non commandable, cf. Théorème 3.5 :

$$A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix},$$

avec $r \leq n$ le rang de $\mathcal{C}(A, B)$, $A_1 \in \mathcal{M}_r(\mathbb{R})$, $B_1 \in \mathcal{M}_{r,m}(\mathbb{R})$ et (A_1, B_1) commandable. Soit $K = (K_1 \ K_2)$ avec $K_1 \in \mathcal{M}_{m,r}(\mathbb{R})$ et $K_2 \in \mathcal{M}_{m,(n-r)}(\mathbb{R})$. On a donc

$$A + BK = \begin{pmatrix} A_1 + B_1K_1 & A_2 + B_1K_2 \\ 0 & A_3 \end{pmatrix}.$$

Donc, pour les polynômes caractéristiques, on a

$$P_{A+BK}(X) = P_{A_1+B_1K_1}(X)P_{A_3}(X).$$

Il faut donc nécessairement que les pôles non commandables (cf. Remarque 3.2) soient de partie réelle négative. On va voir que c'est aussi une condition suffisante pour stabiliser (Σ) . Cela suggère la définition suivante.

Définition 3.5. Une paire de matrice (A, B) avec $A \in M_n(\mathbb{R})$ et $B \in M_{n,m}(\mathbb{R})$ est dite *stabilisable* si les pôles de sa partie non commandable sont de partie réelle négative.

Un polynôme P unitaire de $\mathbb{R}_n[X]$ est *assignable* pour (A, B) s'il existe $F \in M_{m,n}(\mathbb{R})$ tel que $P_{A+BF}(X) = P(X)$.

Le résultat principal de cette section est le suivant.

Théorème 3.8 (Théorème du placement de pôles - pole shifting theorem). *Soit (A, B) une paire de matrices avec $A \in M_n(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$ et r le rang de $\mathcal{C}(A, B)$. Les polynômes de $\mathbb{R}_n[X]$ assignables pour (A, B) sont exactement de la forme suivante*

$$P_{A+BF}(X) = Q(X)P_{nc}(X),$$

avec $Q(X)$ polynôme unitaire quelconque de $\mathbb{R}_r[X]$ et $P_{nc}(X)$ polynôme caractéristique de la partie non commandable de A . En particulier, (A, B) est commandable si et seulement si tout polynôme unitaire de $\mathbb{R}_n[X]$ est assignable pour (A, B) .

PREUVE C.

▷ lairement, si deux paires de matrices sont linéairement équivalentes, alors on peut leur assigner les mêmes polynômes. On peut donc supposer (A, B) décomposée selon Kalman et, d'après ce qui précède, il suffit désormais de se limiter au cas (A, B) commandable.

Faisons d'abord la démonstration dans le cas $m = 1$ (on se ramènera ensuite à ce cas). Par le théorème 3.6, le système est linéairement équivalent à

$$A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ -a_n & -a_{n-1} & \cdots & -a_1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Posons alors $K = (k_1 \cdots k_n)$ et $u = Kx$. On a :

$$A + BK = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ k_1 - a_n & k_2 - a_{n-1} & \cdots & k_n - a_1 \end{pmatrix},$$

et donc

$$\chi_{A+BK}(X) = X^n + (a_1 - k_n)X^{n-1} + \cdots + (a_n - k_1).$$

Donc, pour tout polynôme $P(X) = X^n + \alpha_1 X^{n-1} + \cdots + \alpha_n$, il suffit de choisir $k_1 = a_n - \alpha_n, \dots, k_n = a_1 - \alpha_1$.

Dans le cas général où $m \geq 1$, montrons le lemme fondamental suivant dû à Ackerman.

Lemme 3.9. *Si la paire (A, B) est comandable, alors il existe $y \in \mathbb{R}^m$ et $C \in \mathcal{M}_{m,n}(\mathbb{R})$ tels que la paire $(A + BC, By)$ est aussi commandable.*

D'après ce lemme, pour tout polynôme P unitaire de degré n , il existe $K_1 \in \mathcal{M}_{1,n}(\mathbb{R})$ tel que $\chi_{A+BC+ByK_1} = P$, et donc en posant $K = C + yK_1 \in \mathcal{M}_{m,n}(\mathbb{R})$, on a $\chi_{A+BK} = P$, ce qui prouve le théorème.

PREUVE.

▷ [Preuve du lemme] Soit $y \in \mathbb{R}^m$ tel que $By \neq 0$. On pose $x_1 = By$. On a le fait suivant :

Fait 1 : Il existe $x_2 \in Ax_1 + \text{Im } B$ (et donc il existe $y_1 \in \mathbb{R}^m$ tel que $x_2 = Ax_1 + By_1$) tel que $\dim \text{Vect}\{x_1, x_2\} = 2$.

En effet sinon, on a $Ax_1 + \text{Im } B \subset \mathbb{R}x_1$, donc $Ax_1 \in \mathbb{R}x_1$ et $\text{Im } B \subset \mathbb{R}x_1$. D'où

$$\text{Im } AB = A\text{Im } B \subset \mathbb{R}Ax_1 \subset \mathbb{R}x_1,$$

et par récurrence immédiate :

$$\forall k \in \mathbb{N} \quad \text{Im } A^k B \subset \mathbb{R}x_1.$$

On en déduit que

$$\text{Im } (B, AB, \dots, A^{n-1}B) = \text{Im } B + \text{Im } AB + \cdots + \text{Im } A^{n-1}B \subset \mathbb{R}x_1,$$

ce qui contredit la condition de Kalman.

Fait 2 : Pour tout $k \leq n$, il existe $x_k \in Ax_{k-1} + \text{Im } B$ (et donc il existe $y_{k-1} \in \mathbb{R}^m$ tel que $x_k = Ax_{k-1} + By_{k-1}$) tel que $\dim E_k = k$, où $E_k = \text{Vect}\{x_1, \dots, x_k\}$.

En effet sinon, on a $Ax_{k-1} + \text{Im } B \subset E_{k-1}$, d'où $Ax_{k-1} \subset E_{k-1}$ et $\text{Im } B \subset E_{k-1}$. On en déduit que

$$AE_{k-1} \subset E_{k-1}.$$

En effet, on remarque que $Ax_1 = x_2 - By_1 \in E_{k-1} + \text{Im } B \subset E_{k-1}$, de même pour Ax_2 , etc, $Ax_{k-2} = x_{k-1} - By_{k-1} \in E_{k-1} + \text{Im } B \subset E_{k-1}$, et enfin, $Ax_{k-1} \in E_{k-1}$.

Par conséquent :

$$\text{Im } AB = A\text{Im } B \subset AE_{k-1} \subset E_{k-1},$$

et de même :

$$\forall i \in \mathbb{N} \quad \text{Im } A^i B \subset E_{k-1}.$$

D'où :

$$\text{Im } (B, AB, \dots, A^{n-1}B) \subset E_{k-1},$$

ce qui contredit la condition de Kalman.

On a donc ainsi construit une base (x_1, \dots, x_n) de \mathbb{R}^n . On définit alors $C \in \mathcal{M}_{m,n}(\mathbb{R})$ par les relations :

$$Cx_1 = y_1, Cx_2 = y_2, \dots, Cx_{n-1} = y_{n-1}, Cx_n \text{ quelconque.}$$

Alors la paire $(A + BC, x_1)$ vérifie la condition de Kalman, car :

$$(A + BC)x_1 = Ax_1 + By_1 = x_2, \dots, (A + BC)x_{n-1} = Ax_{n-1} + By_{n-1} = x_n. \quad \square$$

Le théorème est prouvé. □

3.5 Observabilité

3.5.1 Définition et critère d'observabilité de Kalman

Considérons à nouveau un système de commande (Σ) linéaire autonome. En général, les mesures dont on dispose ne nous permettent pas d'observer directement l'état $x(t)$ mais seulement un vecteur $y(t) \in \mathbb{R}^p$, fonction de l'état et de la commande. Nous supposons cette fonction linéaire et indépendante du temps, c'est-à-dire que le système est maintenant de la forme :

$$(\tilde{\Sigma}) : \begin{cases} x'(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases}, \quad t \in [0, \tau],$$

où $C \in M_{p,n}(\mathbb{R})$ et $D \in M_{p,m}(\mathbb{R})$. Le problème de l'*observabilité* est le suivant : connaissant $y(t)$ et $u(t)$ pour tout $t \in [0, \tau]$ ($\tau > 0$) est-il possible de déterminer la condition initiale $x(0)$? Remarquons :

- que la connaissance de $x(0)$ est équivalente à celle de $x(t)$ pour tout $t \in [0, \tau]$ puisque d'après la formule de variation de la constante

$$x(t) = e^{tA}x(0) + \int_0^t e^{(t-s)A}Bu(s)ds,$$

- le deuxième terme du membre de droite de l'égalité précédente étant supposé connu ;
- que l'on peut supposer $D = 0$ et $B = 0$ puisque l'on connaît $u(\cdot)$.

Il suffit donc d'étudier le problème de l'observabilité pour le système réduit

$$(\tilde{\Sigma}_0) : \begin{cases} x'(t) = Ax(t) \\ y(t) = Cx(t) \end{cases}, \quad t \in [0, \tau],$$

c'est-à-dire que l'on se ramène à l'étude de $y(t) = Ce^{tA}x_0$. Appelons *espace d'inobservabilité* \mathcal{I}_τ du système $(\tilde{\Sigma}_0)$ l'ensemble des conditions initiales $x(0) \in \mathbb{R}^n$ pour lesquelles la solution $y(t)$ est identiquement nulle sur $[0, \tau]$, *i.e.*

$$\mathcal{I}_\tau = \left\{ x_0 \in \mathbb{R}^n : \begin{array}{l} \text{la solution de } (\tilde{\Sigma}_0) \\ \text{avec } x(0) = x_0 \text{ vérifie } y(t) \equiv 0 \end{array} \right\}.$$

Définition 3.6. On dit que le système $(\tilde{\Sigma}_0)$ est *observable* si son espace d'inobservabilité est réduit à $\{0\}$.

Le résultat élémentaire suivant montre que cette définition de l'observabilité correspond bien à la question que l'on s'était posée initialement

Proposition 3.10. *Si le système $(\tilde{\Sigma}_0)$ est observable, la connaissance de $y(\cdot)$ sur $[0, \tau]$ détermine de façon univoque $x(0)$.*

PREUVE.

- ▷ Si ce n'était pas le cas il existerait deux vecteurs distincts x_0 et \tilde{x}_0 dans \mathbb{R}^n tels que

$$Ce^{tA}x_0 = Ce^{tA}\tilde{x}_0,$$

ce qui entraînerait $Ce^{tA}(x_0 - \tilde{x}_0) = 0$. Mais d'après la définition de l'observabilité ceci implique $x_0 = \tilde{x}_0$. □

Il existe un critère très simple permettant de déterminer si un système est observable.

Théorème 3.11 (Critère d'observabilité de Kalman). *L'espace d'inobservabilité du système $(\tilde{\Sigma}_0)$ est le noyau de la matrice $(np \times n)$*

$$\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}.$$

Autrement dit, le système $(\tilde{\Sigma}_0)$ est observable si et seulement si $\ker \mathcal{O} = \{0\}$.

PREUVE.

▷ D'après le théorème de Cayley-Hamilton (c'est un argument que nous avons déjà utilisé dans la démonstration du théorème 3.2), pour tout $t \in [0, \tau]$

$$Ce^{tA} \in \text{Vect}(C, CA, \dots, CA^{n-1}).$$

Par conséquent, si

$$CA^j v = 0, \quad j = 0, \dots, n-1, \quad (3.4)$$

on a $Ce^{tA}v = 0$, c'est-à-dire que v est dans l'espace d'inobservabilité de $(\tilde{\Sigma}_0)$. Mais la condition (3.4) est équivalente au fait que $v \in \ker \mathcal{O}$. Nous avons donc démontré que $\ker \mathcal{O}$ est inclus dans l'espace d'inobservabilité de $(\tilde{\Sigma}_0)$.

▷ Réciproquement, supposons que pour tout $t \in [0, \tau]$,

$$Ce^{tA}v = 0.$$

Alors, en dérivant j fois l'égalité précédente en $t = 0$ ($0 \leq j \leq n-1$), il vient

$$\forall 0 \leq j \leq n-1, \quad CA^j v = 0,$$

ce qui signifie que $v \in \ker \mathcal{O}$. L'inclusion réciproque est démontrée. □

Remarque. Si l'on compare le théorème précédent avec le théorème 3.2 on s'aperçoit que le système $(\tilde{\Sigma}_0)$ est observable si et seulement si le système dual $(\Sigma) : z'(t) = A^T z(t) + C^T u(t)$ est commandable (prendre la transposée de la matrice \mathcal{O}). C'est la *dualité contrôlabilité/observabilité*. Ce fait, très important, permet de transférer aux systèmes observés tous les résultats établis sur les systèmes contrôlés.

Définition 3.7 (Equivalence linéaire). Les systèmes

$$\begin{cases} \dot{x}_1 = A_1 x_1 + B_1 u_1 \\ y_1 = C_1 x_1 \end{cases} \quad \text{et} \quad \begin{cases} \dot{x}_2 = A_2 x_2 + B_2 u_2 \\ y_2 = C_2 x_2 \end{cases}$$

sont dits *linéairement équivalents* s'il existe une matrice $P \in GL_n(\mathbb{R})$ telle que

$$A_2 = PA_1P^{-1}, \quad B_2 = PB_2, \quad C_2 = C_1P^{-1}$$

(et dans ce cas on a $x_2 = Px_1, u_2 = u_1, y_2 = y_1$).

Proposition 3.12. *Tout système $\dot{x} = Ax + Bu, y = Cx$, est linéairement équivalent à un système $\dot{\bar{x}} = \bar{A}\bar{x} + \bar{B}u, y = \bar{C}\bar{x}$, avec*

$$\bar{A} = \begin{pmatrix} \bar{A}_1 & 0 \\ \bar{A}_2 & \bar{A}_3 \end{pmatrix}, \quad \bar{C} = (\bar{C}_1 \ 0),$$

i.e.

$$\begin{cases} \dot{\bar{x}}_1 = \bar{A}_1\bar{x}_1 + \bar{B}_1u \\ \dot{\bar{x}}_2 = \bar{A}_2\bar{x}_1 + \bar{A}_3\bar{x}_2 + \bar{B}_2u \\ y_1 = \bar{C}_1\bar{x}_1 \end{cases} \quad \text{partie non observable}$$

et la paire (\bar{A}_1, \bar{C}_1) est observable.

PREUVE.

▷ Il suffit d'appliquer le résultat vu en contrôlabilité au système $\dot{x} = A^T x + C^T u$.

□

Définition 3.8. Dans cette décomposition, les valeurs propres de \bar{A}_3 sont appelées *pôles inobservables* de A et les valeurs propres de \bar{A}_1 sont dites *pôles observables* de A .

Proposition 3.13 (Forme canonique d'observabilité, cas $p = 1$). *Dans le cas $p = 1$, le système $\dot{x} = Ax + Bu, y = Cx$, est observable si et seulement si il est linéairement équivalent au système $\dot{x}_1 = A_1x_1 + B_1u, y = C_1x_1$, avec*

$$A_1 = \begin{pmatrix} 0 & \cdots & 0 & -a_n \\ 1 & 0 & & \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -a_1 \end{pmatrix}, \quad C_1 = (0 \ \cdots \ 0 \ 1).$$

3.5.2 Stabilisation par retour d'état statique

On peut se demander si, étant donné un système contrôlable et observable $\dot{x} = Ax + Bu$, $y = Cx$, il existe un feedback $u = Ky$ stabilisant le système, i.e. si la matrice $A + BKC$ est Hurwitz.

La réponse est *NON*. Pour le voir, considérons les matrices

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

Le système $\dot{x} = Ax + Bu$, $y = Cx$, est trivialement contrôlable et observable. Pourtant, pour toute matrice scalaire $K = (k)$, la matrice

$$A + BKC = \begin{pmatrix} 0 & 1 \\ k & 0 \end{pmatrix}$$

n'est pas Hurwitz.

En conclusion, un feedback par retour d'état statique ne suffit pas en général. C'est pourquoi, dans la suite, on va voir comment construire un retour d'état dynamique.

3.5.3 Observateur asymptotique de Luenberger

Motivation : supposons que le système $\dot{x} = Ax + Bu$, $y = Cx$, soit observable. Le but est de construire un *observateur asymptotique* $\hat{x}(\cdot)$ de $x(\cdot)$, i.e. une fonction dynamique $\hat{x}(\cdot)$ de l'observable $y(\cdot)$, telle que $\hat{x}(t) - x(t) \xrightarrow[t \rightarrow +\infty]{} 0$. L'idée est de copier la dynamique du système observé et d'y ajouter un terme correctif qui tient compte de l'écart entre la prédiction et la réalité.

Définition 3.9. Un *observateur asymptotique* (ou *observateur de Luenberger*) $\hat{x}(\cdot)$ de $x(\cdot)$ est une solution d'un système du type

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(C\hat{x}(t) - y(t)),$$

où $L \in \mathcal{M}_{n,p}(\mathbb{R})$ est appelée *matrice de gain*, telle que

$$\forall x(0), \hat{x}(0) \in \mathbb{R}^n \quad \hat{x}(t) - x(t) \xrightarrow[t \rightarrow +\infty]{} 0.$$

Remarque. Introduisons $e(t) = \hat{x}(t) - x(t)$, l'erreur entre la prédiction $\hat{x}(\cdot)$ et l'état réel $x(\cdot)$. On a :

$$\dot{e}(t) = (A + LC)e(t),$$

et donc $e(t) \xrightarrow[t \rightarrow +\infty]{} 0$ pour toute valeur initiale $e(0)$ si et seulement si la matrice $A + LC$ est Hurwitz. Construire un observateur asymptotique revient donc à déterminer une matrice de gain L telle que $A + LC$ soit Hurwitz. Ainsi, de manière duale au théorème de placement de pôles, on a :

Théorème 3.14 (Théorème de placement des pôles de l'observateur). *Si la paire (A, C) est observable, alors le système admet un observateur asymptotique (i.e. on peut construire une matrice de gains L telle que $A + LC$ soit Hurwitz).*

PREUVE.

▷ La paire (A^T, C^T) étant contrôlable, d'après le théorème de placement de pôles il existe une matrice L^T telle que la matrice $A^T + C^T L^T$ soit Hurwitz. □

3.5.4 Stabilisation par retour dynamique de sortie

On a vu comment construire :

- un régulateur (feedback) pour un système contrôlable,
- un observateur asymptotique pour un système observable.

Il semble naturel, pour un système contrôlable et observable, de construire un régulateur en fonction de l'observateur asymptotique de l'état : c'est l'étape de *synthèse régulateur-observateur*.

Définition 3.10. On appelle *feedback dynamique de sortie*, ou *observateur-régulateur*, le feedback $u = K\hat{x}$, où

$$\dot{\hat{x}} = A\hat{x} + Bu + L(C\hat{x} - y).$$

Théorème 3.15 (Théorème de stabilisation par retour dynamique de sortie). *Si le système $\dot{x} = Ax + Bu$, $y = Cx$, est contrôlable et observable, alors il est stabilisable par retour dynamique de sortie, i.e. il existe des matrices de gain $K \in \mathcal{M}_{m,n}(\mathbb{R})$ et $L \in \mathcal{M}_{n,p}(\mathbb{R})$ telles que les matrices $A + BK$ et $A + LC$ soient Hurwitz, et alors le système bouclé*

$$\begin{aligned}\dot{x} &= Ax + BK\hat{x} \\ \dot{\hat{x}} &= (A + BK)\hat{x} + LC(\hat{x} - x)\end{aligned}$$

est asymptotiquement stable.

PREUVE.

▷ Posons $e = \hat{x} - x$. Alors :

$$\frac{d}{dt} \begin{pmatrix} x \\ e \end{pmatrix} = \begin{pmatrix} A + BK & BK \\ 0 & A + LC \end{pmatrix} \begin{pmatrix} x \\ e \end{pmatrix},$$

et donc ce système est asymptotiquement stable si et seulement si les matrices $A + BK$ et $A + LC$ sont Hurwitz, ce qui est possible avec les propriétés de contrôlabilité et d'observabilité.

□

Remarque. Le fait que la tâche de stabilisation se résolve indépendamment de celle de reconstruction porte le nom de *principe de séparation*.

Chapitre 4

Commandabilité des systèmes non linéaires

4.1 Commandabilité locale et globale

Nous commençons ce chapitre en rappelant la notion de commandabilité pour un système de contrôle du type

$$\dot{x} = F(x, u), \quad x \in \Omega \subset \mathbb{R}^n, \quad u \in U \subset \mathbb{R}^m, \quad (4.1)$$

où F est lisse (\mathcal{C}^∞) par rapport à x . On supposera dans toute la suite que Ω est connexe.

Une classe spécialement importante de systèmes est celle dite des *systèmes affines dans le contrôle*, du type

$$\dot{x} = f_0(x) + \sum_{i=1}^m u_i f_i(x), \quad x \in \Omega \subset \mathbb{R}^n, \quad u \in U \subset \mathbb{R}^m.$$

On appelle f_0 la *dérivée* du système et f_1, \dots, f_m les *champs contrôlés*.

Pour simplifier la discussion, nous considérons comme lois de commande admissibles les fonctions du temps qui sont constantes par morceaux et à valeur dans l'ensemble U . Nous ferons toujours l'hypothèse que toutes les trajectoires de (4.1) correspondantes à un contrôle constant sont définies et restent dans Ω pour tout temps (positif et négatif). On dira alors que chaque champs de vecteurs $F(\cdot, u)$ est *complet*.

Étant donné $x_0 \in \mathbb{R}^n$ et une loi de commande $u : [0, \tau] \rightarrow \mathbb{R}^m$, on note $x(t; x_0, u)$ la solution de (4.1) démarrante à x_0 à l'instant 0, associée à la commande $u(\cdot)$ et évaluée au temps t . On dit qu'un état $x_1 \in \mathbb{R}^n$ est *atteignable en temps τ* à partir de x_0 (par le système (4.1) s'il existe une loi de commande $u : [0, \tau] \rightarrow \mathbb{R}^m$ telle que $x(\tau; x_0, u) = x_1$).

On note $\mathcal{A}(\tau, x_0)$ l'ensemble des états atteignables à partir de x_0 en temps τ . On note aussi

$$\mathcal{A}(\leq \tau, x_0) = \cup_{t \in [0, \tau]} \mathcal{A}(t, x_0)$$

et

$$\mathcal{A}(x_0) = \cup_{t \in [0, +\infty[} \mathcal{A}(t, x_0).$$

Définition 4.1. Le système (4.1) est dit *complètement commandable* si $\mathcal{A}(x_0) = \Omega$ pour tout $x_0 \in \Omega$. Le système (4.1) est dit *localement commandable à x_0 en temps petit* si x_0 appartient à l'intérieur de $\mathcal{A}(\leq \tau, x_0)$ pour tout $\tau > 0$.

La linéarisation de (4.1) autour d'un équilibre (x_0, u_0) montre que si le système linéaire

$$\dot{y} = D_x F(x_0, u_0)y + D_u F(x_0, u_0)v, \quad y \in \mathbb{R}^n, \quad v \in \mathbb{R}^m,$$

est commandable, alors (4.1) est localement commandable à x_0 en temps petit.

Ce résultat ne nous suffit pas pour pouvoir nous prononcer sur la commandabilité locale, même dans des situations “non linéaires” relativement simples comme celle du robot E=M6 décrit par les équations

$$(E = M6) \quad \begin{cases} x'(t) &= u_1 \cos(\theta(t)), \\ y'(t) &= u_1 \sin(\theta(t)), \\ \theta'(t) &= u_2, \end{cases}$$

avec $(x, y, \theta) \in \mathbb{R}^2 \times S^1$ et $(u_1, u_2) \in \mathbb{R}^2$. Il a été vu en exercice que le linéarisé de $(E = M6)$ autour d'un équilibre formé d'un point $(\bar{x}, \bar{y}, \bar{\theta})$ quelconque et du contrôle $(0, 0)$ n'est jamais commandable bien que le système non linéaire $(E = M6)$ le soit.

Plutôt que la linéarisation, un outil plus adapté pour l'étude de la commandabilité (locale ou globale) des systèmes de contrôle non linéaires s'avère être le crochet de Lie, présenté dans la section suivante.

4.2 Crochets et algèbres de Lie

Soient f et g deux champs de vecteurs lisses et complets sur $\Omega \subset \mathbb{R}^n$. Le crochet de Lie entre f et g est défini comme le champ de vecteurs $[f, g]$ donné par la formule

$$[f, g](x) = Dg(x)f(x) - Df(x)g(x).$$

Remarquons que, pour tout $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$[f, \lambda_1 g_1 + \lambda_2 g_2] = \lambda_1 [f, g_1] + \lambda_2 [f, g_2]$$

et

$$[g, f] = -[f, g].$$

En particulier $[f, f] = 0$.

Pour comprendre le rôle du crochet de Lie dans l'analyse de commandabilité, associons à f et g , par analogie avec l'exponentielle matricielle, deux familles de transformations

de Ω , notés par e^{tf} et e^{tg} , donnés par les flots correspondants aux équations $\dot{x} = f(x)$ et $\dot{x} = g(x)$. Cela signifie que $e^{tf}(x_0)$, défini pour tout $x_0 \in \Omega$ et $t \in \mathbb{R}$, est l'évaluation au temps t de la solution du problème de Cauchy

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0.$$

Remarquons que, pour tout $t, s \in \mathbb{R}$, $e^{(t+s)f} = e^{tf} \circ e^{sf}$. En particulier e^{tf} est inversible et $(e^{tf})^{-1} = e^{-tf}$.

Le lemme suivant relie la direction du champ de vecteurs $[f, g]$ aux propriétés de commutation des flots associés à f et g .

Lemme 4.1. *Pour tout $x \in \Omega$,*

$$e^{-tg} \circ e^{-tf} \circ e^{tg} \circ e^{tf}(x) = x + t^2[f, g](x) + o(t^2)$$

lorsque t tend vers 0.

PREUVE.

▷ Il suffit de faire, pour chaque flot, un développement limité à l'ordre 3 lorsque t tend vers 0. Ainsi, on a

$$e^{tf}(x) = x + tf(x) + \frac{t^2}{2}Df(x)f(x) + O(t^3),$$

puis

$$e^{tg} \circ e^{tf}(x) = x + t(f(x) + g(x)) + \frac{t^2}{2}Df(x)f(x) + t^2Dg(x)f(x) + \frac{t^2}{2}Dg(x)g(x) + O(t^3),$$

ce qui donne

$$e^{-tf} \circ e^{tg} \circ e^{tf}(x) = x + tg(x) + [f, g](x) + \frac{t^2}{2}Dg(x)g(x) + O(t^3),$$

et enfin le résultat. □

Un corollaire très important du calcul précédent est le suivant.

Corollaire 4.2. *Les flots e^{tf} et e^{tg} correspondant à deux champs de vecteur f, g lisses et complets commutent pour tout temps t si et seulement si leur crochet de Lie $[f, g]$ est nul.*

Le résultat précédent nous fait interpréter une condition du type

$$[f, g](x) \notin \text{Vect}(f(x), g(x)),$$

dans le sens suivant : la liberté de choisir à chaque instant une dynamique entre celle de f et celle de g nous permet d'atteindre, en démarrant de x , des points qui ne sont pas directement atteignables par une combinaison linéaire de $f(x)$ et $g(x)$. Ceci est à la base de la théorie de la commandabilité des systèmes non linéaires et est bien illustré par l'exemple du robot E=M6.

La dynamique ($E = M6$) peut s'écrire comme

$$\dot{z} = u_1 f_1(z) + u_2 f_2(z),$$

avec $z = (x, y, \theta)$ et

$$f_1(z) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix}, \quad f_2(z) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Le crochet de Lie $[f_1, f_2]$ correspond en pratique à effectuer un créneau (infinitésimal) et est égal à

$$[f_1, f_2](z) = \begin{pmatrix} \sin(\theta) \\ -\cos(\theta) \\ 0 \end{pmatrix}.$$

Il est clair que pour tout $z \in \mathbb{R}^2 \times S^1$, les vecteurs $f_1(z)$, $f_2(z)$ et $f_3(z)$ sont linéairement indépendants.

Une notion particulièrement importante qui généralise la situation décrite dans le cas du robot E=M6 est celle de la *génération par crochets de Lie*. Pour l'introduire, nous associons à une famille \mathcal{F} de champs de vecteurs lisses et complets sur Ω , l'espace vectoriel $\text{Lie}(\mathcal{F})$ engendré par \mathcal{F} et tous les champs de vecteurs de la forme

$$[f_1, [f_2, [\dots, [f_{k-2}, [f_{k-1}, f_k] \dots]]],$$

où $k \geq 2$ et $f_1, \dots, f_k \in \mathcal{F}$.

Définition 4.2. On appelle $\text{Lie}(\mathcal{F})$ l'*algèbre de Lie associée à \mathcal{F}* et on dit que \mathcal{F} *satisfait la condition de génération par crochets de Lie* à un point $x \in \Omega$ si la dimension de $\text{Lie}_x(\mathcal{F}) := \{f(x) \mid f \in \text{Lie}(\mathcal{F})\}$ est égale à la dimension de Ω . On dit que \mathcal{F} satisfait la condition de génération par crochets de Lie si elle la satisfait pour tout $x \in \Omega$.

4.3 Accessibilité locale et conditions suffisantes pour la commandabilité globale

Considérons à nouveau le système (4.1) et notons \mathcal{F}_U l'ensemble

$$\mathcal{F}_U = \{F(\cdot, u) \mid u \in U\}.$$

Les éléments de \mathcal{F}_U sont dits *champs admissibles* de (4.1). La condition de génération par crochets de Lie exprime la notion suivante : en commutant entre les dynamiques de la famille \mathcal{F}_U , toutes les directions de l'espace deviennent atteignables par des trajectoires de (4.1).

Cela doit néanmoins tenir compte des contraintes de dérive : nous pouvons penser au cas où $\Omega = \mathbb{R}^2$ et \mathcal{F}_U est donnée par la famille de champs de vecteurs constants $\{(1, 1), (1, -1)\}$. Alors \mathcal{F}_U satisfait la condition de génération par crochets de Lie en tout point de \mathbb{R}^2 , même si le système $(\dot{x}_1, \dot{x}_2) = (1, \pm 1)$ n'est pas complètement commandable ni localement commandable en aucun point.

L'atteignabilité "dans toutes les directions" doit s'interpréter dans le sens du célèbre résultat suivant.

Théorème 4.3 (Krener). *Si \mathcal{F}_U satisfait la condition de génération par crochets de Lie à $x_0 \in \Omega$, alors, pour tout $\tau > 0$, x_0 appartient à la fermeture de l'intérieur de $\mathcal{A}(\leq \tau, x_0)$.*

PREUVE.

▷ Remarquons d'abord que \mathcal{F}_U satisfait la condition de génération par crochets de Lie à tous les points d'un voisinage de x_0 . (Si n champs de vecteurs sont linéairement indépendants à x_0 , ils le sont dans un voisinage.)

Remarquons aussi qu'il existe $f \in \mathcal{F}_U$ tel que $f(x_0) \neq 0$: on aurait sinon $\text{Lie}_{x_0}(\mathcal{F}_U) = \{0\}$.

Si $\dim(\Omega) = 1$ la preuve est finie. Si $\dim(\Omega) > 1$ et si tout champ dans \mathcal{F}_U est tangent à la courbe $t \mapsto e^{tf}(x_0)$, $0 < t < \varepsilon$, alors on déduit du Lemme 4.1 que $\text{Lie}_{e^{tf}(x_0)}(\mathcal{F}_U)$ est aussi tangent à telle courbe et qu'il n'a donc pas la même dimension que Ω . Ceci contredit la première remarque faite dans cette preuve. Ils existent donc $g \in \mathcal{F}_U$ et $0 < \bar{t} < \varepsilon$ tels que f et g sont linéairement indépendants dans un voisinage de $x_1 = e^{\bar{t}f}(x_0)$. Alors $(t, s) \mapsto e^{sg} \circ e^{tf}(x_0)$, $0 < s < \varepsilon'$, $\bar{t} - \varepsilon' < t < \bar{t} + \varepsilon'$, a comme image une surface de dimension deux.

Si $\dim(\Omega) = 2$ la preuve est finie. Sinon on recommence le même argument et on conclue par récurrence par rapport à la dimension de Ω .

□

Le théorème peut être interprété comme suit : sous la condition de génération par crochets de Lie, les trajectoires admissibles démarrant en un point x_0 peuvent accéder à un ensemble d'intérieur non vide en temps arbitrairement petit (on parle aussi, dans la littérature, d'accessibilité locale).

Dans le cas où l'ensemble des directions admissibles de (4.1) est convenablement symétrique, la conclusion du théorème de Krener devient plus explicite.

Théorème 4.4 (Chow). Soient f_1, \dots, f_m des champs de vecteurs lisses et complets sur Ω . Si $U \subset \mathbb{R}^m$ contient 0 dans son intérieur et si $\{f_1, \dots, f_m\}$ satisfait la condition de génération par crochets de Lie (en tout $x \in \Omega$), alors

$$\dot{x} = \sum_{i=1}^m u_i f_i(x)$$

est complètement commandable.

PREUVE.

▷ Sans perte de généralité, nous pouvons supposer que U est une boule de \mathbb{R}^m centrée à l'origine.

Soit $x_0 \in \Omega$. Remarquons que \mathcal{F}_U , avec $F(x, u) = \sum_{i=1}^m u_i f_i(x)$, satisfait la condition de génération par crochets de Lie. Nous déduisons donc du théorème de Krener que $\mathcal{A}(x_0)$ contient un ouvert non vide ω .

Choisissons $u^1, \dots, u^k \in U$ et $t_1, \dots, t_k > 0$ tels que

$$e^{t_k \sum_{i=1}^m u_i^k f_i} \circ \dots \circ e^{t_1 \sum_{i=1}^m u_i^1 f_i}(x_0) \in \omega.$$

Puisque $-u^1, \dots, -u^k \in U$, nous avons que

$$e^{t_1 \sum_{i=1}^m (-u_i^1) f_i} \circ \dots \circ e^{t_k \sum_{i=1}^m (-u_i^k) f_i}(\mathcal{A}(x_0)) \subset \mathcal{A}(x_0).$$

En particulier, $\mathcal{A}(x_0)$ contient

$$e^{t_1 \sum_{i=1}^m (-u_i^1) f_i} \circ \dots \circ e^{t_k \sum_{i=1}^m (-u_i^k) f_i}(\omega)$$

qui est un voisinage de x_0 .

Nous en déduisons que $\mathcal{A}(x_0)$ est ouvert. (En effet, si $x_1 \in \mathcal{A}(x_0)$, alors $x_1 \in \text{int}(\mathcal{A}(x_1)) \subset \mathcal{A}(x_1) \subset \mathcal{A}(x_0)$.) De plus, la symétrie de U implique (comme nous venons de le voir) que $x_1 \in \mathcal{A}(x_0)$ si et seulement si $x_0 \in \mathcal{A}(x_1)$. Donc $\{\mathcal{A}(x_0) \mid x_0 \in \Omega\}$ est une partition de Ω en ouverts disjoints. Ω étant connexe, nous pouvons conclure que $\mathcal{A}(x_0) = \Omega$ pour tout $x_0 \in \Omega$. \square

4.4 Champs compatibles

Une autre conséquence très importante du théorème de Krener est le corollaire suivant, qui sera la clé pour étendre le champ d'applications du théorème de Chow.

Corollaire 4.5. Si \mathcal{F}_U satisfait la condition de génération par crochets de Lie (à tout $x \in \Omega$) et si $\mathcal{A}(x_0)$ est dense dans Ω pour un certain $x_0 \in \Omega$, alors $\mathcal{A}(x_0) = \Omega$.

PREUVE.

▷ Soit $x_1 \in \Omega$ et considérons le système

$$\dot{x} = -F(x, u), \quad x \in \Omega \subset \mathbb{R}^n, \quad u \in U \subset \mathbb{R}^m, \quad (4.2)$$

qui est obtenu à partir de (4.1) par renversement temporel. La famille de champs de vecteurs admissibles pour (4.2) étant $-\mathcal{F}_U$, elle satisfait la condition de génération par crochets de Lie.

L'ensemble atteignable pour (4.2) au départ de x_1 , noté par $\mathcal{A}^-(x_1)$, contient donc un ouvert non vide (Krener). En particulier, $\mathcal{A}^-(x_1)$ a intersection non vide avec $\mathcal{A}(x_0)$. Cela signifie précisément que $x_1 \in \mathcal{A}(x_0)$. Puisque $x_1 \in \Omega$ était arbitraire, nous avons montré que $\mathcal{A}(x_0) = \Omega$. □

Le corollaire 4.5 nous suggère la définition suivante.

Définition 4.3. Un champ de vecteurs lisse et complet g est dit *compatible avec \mathcal{F}_U* si le système de contrôle

$$\dot{x} = \hat{F}(x, u), \quad x \in \Omega, \quad u \in \hat{U}, \quad (4.3)$$

avec $\hat{U} = U \cup \{\hat{u}\}$, $\hat{u} \notin U$, et

$$\hat{F}(x, u) = \begin{cases} F(x, u) & \text{si } u \in U, \\ g(x) & \text{si } u = \hat{u}, \end{cases}$$

satisfait à la condition suivante : pour tout $x_0 \in \Omega$, l'ensemble atteignable pour (4.3) au départ de x_0 est contenu dans la fermeture de $\mathcal{A}(x_0)$.

Nous allons maintenant présenter des critères permettant d'identifier des champs de vecteurs compatibles avec une famille \mathcal{F}_U donnée.

Le premier critère est celui dit de *convexification* : il affirme que la combinaison convexe de champs de vecteurs de \mathcal{F}_U est compatible avec \mathcal{F}_U . Il formalise l'intuition selon laquelle si on commute très rapidement entre la dynamique de deux champs de vecteurs f et g et si on reste à chaque fois pendant la même (très petite) quantité de temps sur chacune des deux dynamiques, alors la trajectoire résultante approchera celle du champ de vecteurs $\frac{f+g}{2}$.

Le lemme suivant formule précisément ce résultat, tout en y ajoutant la remarque que la multiplication d'une dynamique par une constante positive correspond simplement à une reparamétrisation du temps.

Lemme 4.6. *Pour tous $\lambda_1, \dots, \lambda_k \geq 0$ et tous $f_1, \dots, f_k \in \mathcal{F}_U$, le champ de vecteurs $\lambda_1 f_1 + \dots + \lambda_k f_k$ est compatible avec \mathcal{F}_U .*

La preuve du lemme, assez technique, est basée sur l'inégalité de Gronwall.

Nous pouvons déduire du lemme précédent un précieux corollaire du théorème de Chow.

Corollaire 4.7. *Si, pour tout $x \in \Omega$, \mathcal{F}_U satisfait la condition de génération par crochets de Lie et si 0 appartient à l'intérieur de l'enveloppe convexe de $\{F(x, u) \mid u \in U\}$, alors (4.1) est complètement commandable.*

Un deuxième critère garantissant la compatibilité est le suivant : si g est un champ de vecteurs lisse et complet qui est la limite uniforme sur tout compact de Ω d'une suite $\{f_n\}_{n \in \mathbb{N}}$, où chaque f_n est compatible avec \mathcal{F}_U , alors g est compatible avec \mathcal{F}_U . Ce critère peut facilement se déduire de la théorie générale des équations différentielles ordinaires, en n'affirmant rien d'autre qu'un principe de continuité des solutions d'une EDO par rapport au champ de vecteurs la régissant.

Nous déduisons de ce critère le résultat suivant, qui s'applique à la classe des systèmes affines dans le contrôle.

Corollaire 4.8. *Soient f_0, f_1, \dots, f_m des champs de vecteurs lisses et complets sur Ω . Si la famille $\mathcal{G} = \{f_1, \dots, f_m\}$ satisfait la condition de génération par crochets de Lie, alors*

$$\dot{x} = f_0(x) + \sum_{i=1}^m u_i f_i(x), \quad (u_1, \dots, u_m) \in \mathbb{R}^m$$

est complètement commandable.

PREUVE.

▷ Le théorème de Chow nous garantit que le système

$$\dot{x} = \sum_{i=1}^m u_i f_i(x), \quad (u_1, \dots, u_m) \in \mathbb{R}^m$$

est complètement commandable.

Si nous montrons que chaque $\sum_{i=1}^m u_i f_i$ est compatible avec \mathcal{F}_U , alors la thèse suit du corollaire 4.5.

Il suffit donc de remarquer que

$$\sum_{i=1}^m u_i f_i = \lim_{n \rightarrow \infty} \frac{1}{n} (f_0 + \sum_{i=1}^m (n u_i) f_i)$$

et que chaque champ de vecteurs $\frac{1}{n} (f_0 + \sum_{i=1}^m (n u_i) f_i)$ est compatible avec \mathcal{F}_U .

□

Un dernier critère de compatibilité très utile peut être formulé en termes de récurrence d'un champ de vecteurs.

Définition 4.4 (Champ de vecteurs récurrent). Un champ de vecteurs f lisse et complet sur Ω est dit *récurrent* si, pour tout point $x_0 \in \Omega$, tout voisinage V de x_0 et tout temps $t > 0$, il existe $t^* > t$ tel que $e^{t^*f}(x_0) \in V$.

Remarquons que si les trajectoires de f sont périodiques (éventuellement avec des périodes différentes selon les trajectoires) alors f est récurrent.

Lemme 4.9. *Si f est compatible avec \mathcal{F}_U et est récurrent, alors $-f$ est aussi compatible avec \mathcal{F}_U .*

PREUVE.

▷ Soient $x_0 \in \Omega$ et $t > 0$. Il suffit de montrer que $e^{-tf}(x_0)$ est la limite de points atteignables au départ de x_0 .

La définition de champ récurrent implique l'existence d'une suite croissante non bornée $\{t_k\}_{k \in \mathbb{N}}$ de temps positifs telle que $e^{t_k f}(x_0) \rightarrow x_0$ pour $k \rightarrow \infty$.

Nous avons donc que $e^{(t_k-t)f}(x_0) \rightarrow e^{-tf}(x_0)$ pour $k \rightarrow \infty$ et $t_k > t$ pour k suffisamment grand.

Puisque f est compatible avec \mathcal{F}_U , alors $e^{(t_k-t)f}(x_0)$ appartient à la fermeture de $\mathcal{A}(x_0)$ pour tout k tel que $t_k > t$. Nous en déduisons que $e^{-tf}(x_0)$ appartient à la fermeture de $\mathcal{A}(x_0)$. □

En conséquence du lemme précédent, nous pouvons parfois déterminer la commandabilité d'un système affine dans le contrôle même dans le cas où l'ensemble des contrôles est borné.

Corollaire 4.10. *Soient f_0, f_1, \dots, f_m des champs de vecteurs lisses et complets sur Ω et considérons le système*

$$\dot{x} = f_0(x) + \sum_{i=1}^m u_i f_i(x), \quad (u_1, \dots, u_m) \in U. \quad (4.5)$$

Supposons que 0 appartient à l'intérieur de l'enveloppe convexe de U , que $\{f_0, \dots, f_m\}$ satisfait la condition de génération par crochets de Lie et que f_0 soit récurrent. Alors (4.5) est complètement commandable.

PREUVE.

▷ Remarquons que f_0 est compatible avec \mathcal{F}_U à cause du lemme 4.6. Lemme 4.9 affirme l'équivalence entre la commandabilité de (4.5) et celle de

$$\dot{x} = \sum_{i=0}^m u_i f_i(x), \quad (u_0, \dots, u_m) \in \{-1, 1\} \times U.$$

Cette dernière suit du corollaire 4.7. □

Exemple. Soit $\Omega = S^2$, la sphère unité de \mathbb{R}^3 , c'est-à-dire $\{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 = 1\}$. Considérons les deux champs de vecteurs

$$f_0(x) = \begin{pmatrix} -x_2 \\ x_1 \\ 0 \end{pmatrix}, \quad f_1(x) = \begin{pmatrix} 0 \\ -x_3 \\ x_2 \end{pmatrix},$$

dont les flots au temps t sont les rotations d'angle t par rapport, respectivement, à $(0, 0, 1)^T$ et $(1, 0, 0)^T$.

Le système commandé

$$\dot{x} = f_0(x) + u f_1(x), \quad u \in (-1, 1), \quad x \in S^2,$$

est complètement commandable puisque les trajectoires de f_0 sont périodiques (et donc f_0 est récurrent) et le crochet entre f_0 et f_1 est donné par

$$[f_0, f_1](x) = \begin{pmatrix} -x_3 \\ 0 \\ x_1 \end{pmatrix},$$

de telle sorte que $\text{Lie}_x(\{f_0, f_1\})$ est de dimension deux pour tout $x \in S^2$.

4.5 Orbites et conditions nécessaires pour la commandabilité

Nous avons vu dans les sections précédentes plusieurs conditions suffisantes pour la commandabilité d'un système de contrôle non linéaire.

Cette section présente plutôt des conditions nécessaires, déduites d'un résultat profond de nature géométrique, le théorème de l'orbite. Ce théorème assure que, dans les cas non "pathologiques", l'algèbre de Lie associée à la famille \mathcal{F}_U mesure de façon précise la taille de l'ensemble des directions auxquelles on peut accéder à partir d'un point.

Nous définissons l'orbite à partir d'un point $x_0 \in \Omega$ pour le système (4.1) comme l'ensemble

$$\mathcal{O}(x_0) = \{e^{t_k f_k} \circ \dots \circ e^{t_1 f_1}(x_0) \mid k \in \mathbb{N}, t_1, \dots, t_k \in \mathbb{R}, f_1, \dots, f_k \in \mathcal{F}_U\}.$$

Rappelons, par comparaison, que

$$\mathcal{A}(x_0) = \{e^{t_k f_k} \circ \dots \circ e^{t_1 f_1}(x_0) \mid k \in \mathbb{N}, t_1, \dots, t_k \geq 0, f_1, \dots, f_k \in \mathcal{F}_U\}.$$

Nous avons alors le théorème suivant.

Théorème 4.11 (Orbite). *Pour chaque $x_0 \in \Omega$, l'ensemble $\mathcal{O}(x_0)$ a la structure d'une variété immergée. En particulier, celle-ci a même dimension en tout point. De plus, l'espace des directions tangentes à $\mathcal{O}(x_0)$ à un point $x \in \mathcal{O}(x_0)$ contient $\text{Lie}_x(\mathcal{F}_U)$ et les deux espaces sont égaux si une des deux conditions suivantes est vérifiée : (i) chaque composante de chaque champ de vecteurs de \mathcal{F}_U est une fonction analytique, ou (ii) la dimension de $\text{Lie}_x(\mathcal{F}_U)$ est constante par rapport à $x \in \mathcal{O}(x_0)$.*

Le corollaire suivant retient du théorème de l'orbite les conséquences les plus directement exploitables pour l'analyse de commandabilité d'un système non linéaire.

Corollaire 4.12. *Si \mathcal{F}_U ne satisfait pas la condition de génération par crochets de Lie et si (i) chaque composante de chaque champ de vecteurs dans \mathcal{F}_U est une fonction analytique, ou (ii) la dimension de $\text{Lie}_x(\mathcal{F}_U)$ est constante par rapport à $x \in \Omega$, alors (4.1) n'est pas complètement commandable.*

Chapitre 5

Théorie linéaire-quadratique

Dans ce chapitre, on s'intéresse aux systèmes de contrôle linéaires avec un coût quadratique. Ces systèmes sont d'une grande importance dans la pratique, comme on le verra en section 5.4. En effet un coût quadratique est souvent très naturel dans un problème, par exemple lorsqu'on veut minimiser l'écart au carré par rapport à une trajectoire nominale (problème de poursuite). Par ailleurs même si les systèmes de contrôle sont en général non linéaires, on est très souvent amené à linéariser le système le long d'une trajectoire, par exemple dans des problèmes de stabilisation.

Nous allons donc considérer un système de contrôle linéaire dans \mathbb{R}^n :

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(0) = x_0, \quad (5.1)$$

muni d'un coût quadratique du type :

$$C(u) = {}^t x(T)Qx(T) + \int_0^T ({}^t x(t)W(t)x(t) + {}^t u(t)U(t)u(t))dt, \quad (5.2)$$

où $T > 0$ est fixé, pour tout t , $U(t) \in \mathcal{M}_m(\mathbb{R})$ est symétrique définie positive, $W(t) \in \mathcal{M}_n(\mathbb{R})$ est symétrique positive, et $Q \in \mathcal{M}_n(\mathbb{R})$ est une matrice symétrique positive. On suppose que la dépendance en t de A , B , W et U est L^∞ sur $[0, T]$. Par ailleurs le coût étant quadratique, l'espace naturel des contrôles est $L^2([0, T], \mathbb{R}^m)$.

Le problème de contrôle optimal est alors le suivant, nous l'appellerons *problème LQ* (linéaire-quadratique) :

Problème LQ : Un point initial $x_0 \in \mathbb{R}^n$ étant fixé, l'objectif est de déterminer les trajectoires partant de x_0 qui minimisent le coût $C(u)$.

Notons que l'on n'impose aucune contrainte sur le point final $x(T)$. Pour toute la suite, on pose :

$$\|x(t)\|_W^2 = {}^t x(t)W(t)x(t), \quad \|u(t)\|_U^2 = {}^t u(t)U(t)u(t), \quad \text{et } g(x) = {}^t xQx,$$

de sorte que

$$C(u) = g(x(T)) + \int_0^T (\|x(t)\|_W^2 + \|u(t)\|_U^2)dt.$$

Les matrices Q, W, U sont des matrices de *pondération*.

Remarque. Par hypothèse, les matrices Q et $W(t)$ sont symétriques positives, mais pas nécessairement définies. Par exemple si $Q = 0$ et $W = 0$ alors le coût est toujours minimal pour le contrôle $u = 0$.

Remarque. Comme dans le chapitre précédent, on suppose pour alléger les notations que le temps initial est égal à 0. Cependant tous les résultats qui suivent sont toujours valables si on considère le problème LQ sur un intervalle $[t_0, T]$, avec des contrôles dans l'espace $L^2([t_0, T], \mathbb{R}^m)$.

Remarque. Les résultats des sections 5.1 et 5.2 seront en fait valables pour des systèmes linéaires perturbés $\dot{x} = Ax + Bu + r$, et aussi avec une fonction g de \mathbb{R}^n dans \mathbb{R} continue ou C^1 pour laquelle on puisse montrer que (g1) la fonction coût $C(u)$ soit bornée inférieurement; (g2) toute suite (u_n) convergente vers la borne inférieure du coût est bornée en norme L^2 . Ceci a lieu lorsque que par exemple g vérifie l'hypothèse suivante : il existe un réel a tel que $\frac{g(x)}{\|x\|} + a$ est borné inférieurement pour $\|x\|$ assez grand. Nous précisons pour chaque résultat les extensions possibles.

De même nous envisagerons le cas où $T = +\infty$.

5.1 Existence de trajectoires optimales

Introduisons l'hypothèse de coercivité suivante sur U :

$$\exists \alpha > 0 \mid \forall u \in L^2([0, T], \mathbb{R}^m) \quad \int_0^T \|u(t)\|_U^2 dt \geq \alpha \int_0^T {}^t u(t) u(t) dt. \quad (5.3)$$

Par exemple cette hypothèse est satisfaite si l'application $t \mapsto U(t)$ est continue sur $[0, T]$ et $T < +\infty$, ou encore s'il existe une constante $c > 0$ telle que pour tout $t \in [0, T]$ et pour tout vecteur $v \in \mathbb{R}^m$ on ait ${}^t U(t)v \geq c {}^t v v$.

On a le théorème d'existence suivant :

Théorème 5.1. *Sous l'hypothèse (5.3), il existe une unique trajectoire minimisante pour le problème LQ.*

PREUVE.

▷ Montrons tout d'abord l'existence d'une telle trajectoire. Considérons une suite minimisante $(u_n)_{n \in \mathbb{N}}$ de contrôles sur $[0, T]$, i.e. la suite $C(u_n)$ converge vers la borne inférieure des coûts. En particulier cette suite est bornée. Par hypothèse, il existe une constante $\alpha > 0$ telle que pour tout $u \in L^2([0, T], \mathbb{R}^m)$ on ait $C(u) \geq \alpha \|u\|_{L^2}$. On en déduit que la suite $(u_n)_{n \in \mathbb{N}}$ est bornée dans $L^2([0, T], \mathbb{R}^m)$. Par conséquent à sous-suite près elle converge faiblement vers

un contrôle u de L^2 . Notons x_n (resp. x) la trajectoire associée au contrôle u_n (resp. u) sur $[0, T]$. D'après la formule de variation de la constante, on a, pour tout $t \in [0, T]$:

$$x_n(t) = M(t)x_0 + M(t) \int_0^t M(s)^{-1} B(s) u_n(s) ds \quad (5.4)$$

(et la formule analogue pour $x(t)$). On montre alors aisément que, à sous-suite près, la suite (x_n) converge simplement vers l'application x sur $[0, T]$ (en fait on peut même montrer que la convergence est uniforme).

Passant maintenant à la limite dans (5.4), on obtient, pour tout $t \in [0, T]$:

$$x(t) = M(t)x_0 + M(t) \int_0^t M(s)^{-1} B(s) u(s) ds,$$

et donc x est une solution du système associée au contrôle u . Montrons qu'elle est minimisante. Pour cela on utilise le fait que puisque $u_n \rightharpoonup u$ dans L^2 , on a l'inégalité :

$$\int_0^T \|u(t)\|_U^2 dt \leq \liminf \int_0^T \|u_n(t)\|_U^2 dt,$$

et donc $C(u) \leq \liminf C(u_n)$. Mais comme (u_n) est une suite minimisante, $C(u)$ est donc égal à la borne inférieure des coûts, i.e. le contrôle u est minimisant, ce qui montre l'existence d'une trajectoire optimale.

Pour l'unicité on a besoin du lemme suivant.

Lemme 5.2. *La fonction C est strictement convexe.*

PREUVE.

▷ [Preuve du lemme] Tout d'abord, remarquons que pour tout $t \in [0, T]$, la fonction $f(u) = {}^t U(t)u$ définie sur \mathbb{R}^m est strictement convexe puisque par hypothèse la matrice $U(t)$ est symétrique définie positive. Ensuite, notons $x_u(\cdot)$ la trajectoire associée à un contrôle u . On a pour tout $t \in [0, T]$:

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

Par conséquent, l'application qui à un contrôle u associe $x_u(t)$ est convexe (pourquoi?), ceci pour tout $t \in [0, T]$. Or la matrice $W(t)$ étant symétrique positive, ceci implique que l'application qui à un contrôle u associe ${}^t x(t)W(t)u(t)$ est convexe. On raisonne de même pour le terme ${}^t x(T)Qx(T)$. Enfin, l'intégration respectant la convexité, on a bien que le coût est strictement convexe en u . □

L'unicité de la trajectoire optimale en résulte trivialement. □

Remarque (Extension du théorème 5.1). Si la fonction g apparaissant dans le coût est une fonction continue quelconque de \mathbb{R}^n dans \mathbb{R} vérifiant les conditions (g1) et (g2) (cf. remarque 5), et/ou si le système de contrôle est perturbé par une fonction $r(t)$, alors le théorème précédent reste vrai.

Remarque (Cas d'un intervalle infini). Le théorème est encore valable si $T = +\infty$, avec $g = 0$, pourvu que le système (5.1) soit contrôlable (en temps quelconque).

En effet il suffit juste de montrer qu'il existe des trajectoires solutions du système (5.1) sur $[0, +\infty[$ et de coût fini. Or si le système est contrôlable, alors il existe un contrôle u et un temps $T > 0$ tel que la trajectoire associée à u relie x_0 à 0 sur $[0, T]$. On étend alors le contrôle u par 0 sur $]T, +\infty[$, de sorte que la trajectoire reste en 0. On a ainsi construit une trajectoire solution du système sur $[0, +\infty[$ et de coût fini. Ceci permet d'affirmer l'existence d'une suite de contrôles minimisants. Les autres arguments de la preuve sont inchangés. On obtient donc le résultat suivant.

Proposition 5.3. *Considérons le problème de déterminer une trajectoire solution de*

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

sur $[0, +\infty[$ et minimisant le coût

$$C(u) = \int_0^{+\infty} (\|x(t)\|_W^2 + \|u(t)\|_U^2) dt.$$

Si le système est contrôlable en un temps $T > 0$, et si l'hypothèse (5.3) est satisfaite sur $[0, +\infty[$, alors il existe une unique trajectoire minimisante.

Remarque. – Si on suppose de plus que les applications $A(\cdot)$ et $B(\cdot)$ sont L^2 sur $[0, +\infty[$, et si $W(\cdot)$ vérifie comme U une hypothèse de coercivité (5.3), alors la trajectoire minimisante tend vers 0 lorsque t tend vers l'infini.

En effet on montre facilement en utilisant l'inégalité de Cauchy-Schwarz que l'application $\dot{x}(\cdot)$ est dans L^1 , et par conséquent $x(t)$ converge. Sa limite est alors forcément nulle.

– Dans le cas autonome (A et B sont constantes), si $W(\cdot)$ vérifie comme U une hypothèse de coercivité (5.3), alors la trajectoire minimisante tend vers 0 lorsque t tend vers l'infini.

En effet il suffit d'écrire l'inégalité :

$$\|\dot{x}(t)\| \leq \|A\|\|x(t)\| + \|B\|\|u(t)\| \leq Cste(\|x(t)\|^2 + \|u(t)\|^2),$$

puis en intégrant on montre de même que l'application $\dot{x}(\cdot)$ est dans L^1 .

5.2 Condition nécessaire et suffisante d'optimalité : principe du maximum dans le cas LQ

Théorème 5.4. *La trajectoire x , associée au contrôle u , est optimale pour le problème LQ si et seulement s'il existe un vecteur adjoint $p(t)$ satisfaisant pour presque tout $t \in [0, T]$:*

$$\dot{p}(t) = -p(t)A(t) + {}^t x(t)W(t) \quad (5.8)$$

et la condition finale

$$p(T) = -{}^t x(T)Q. \quad (5.9)$$

De plus le contrôle optimal u s'écrit, pour presque tout $t \in [0, T]$:

$$u(t) = U(t)^{-1} {}^t B(t) {}^t p(t). \quad (5.10)$$

PREUVE.

▷ Soit u un contrôle optimal et x la trajectoire associée sur $[0, T]$. Le coût est donc minimal parmi toutes les trajectoires solutions du système, partant de x_0 , le point final étant non fixé. Considérons alors des perturbations du contrôle u dans $L^2([0, T], \mathbb{R}^m)$:

$$u_{pert}(t) = u(t) + \delta u(t),$$

engendrant les trajectoires :

$$x_{pert}(t) = x(t) + \delta x(t) + o(\|\delta u\|_{L^2}),$$

avec $\delta x(0) = 0$. La trajectoire x_{pert} devant être solution du système $\dot{x}_{pert} = Ax_{pert} + Bu_{pert}$, on en déduit que :

$$\delta \dot{x} = A\delta x + B\delta u,$$

et par conséquent, pour tout $t \in [0, T]$:

$$\delta x(t) = M(t) \int_0^t M(s)^{-1} B(s) \delta u(s) ds. \quad (5.11)$$

Par ailleurs il est bien clair que le coût $C(\cdot)$ est une fonction lisse sur $L^2([0, T], \mathbb{R}^m)$ (elle est même analytique) au sens de Fréchet. Le contrôle u étant minimisant on doit avoir :

$$dC(u) = 0.$$

Or

$$C(u_{pert}) = g(x_{pert}(T)) + \int_0^T (\|x_{pert}(t)\|_W^2 + \|u_{pert}(t)\|_U^2) dt,$$

et comme Q , $W(t)$ et $U(t)$ sont symétriques, on en déduit que :

$$\frac{1}{2}dC(u).\delta u = {}^t x(T)Q\delta x(T) + \int_0^T ({}^t x(t)W(t)\delta x(t) + {}^t u(t)U(t)\delta u(t))dt = 0, \quad (5.12)$$

ceci étant valable pour toute perturbation δu . Cette équation va nous conduire à l'expression du contrôle optimal u . Mais introduisons tout d'abord le vecteur adjoint $p(t)$ comme solution du problème de Cauchy suivant :

$$\dot{p}(t) = -p(t)A(t) + {}^t x(t)W(t), \quad p(T) = -{}^t x(T)Q.$$

La formule de variation de la constante nous conduit à :

$$p(t) = \Lambda M(t)^{-1} + \int_0^t {}^t x(s)W(s)M(s)ds M(t)^{-1}$$

pour tout $t \in [0, T]$, où :

$$\Lambda = -{}^t x(T)QM(T) - \int_0^T {}^t x(s)W(s)M(s)ds.$$

Revenons alors à l'équation (5.12). Tout d'abord, en tenant compte de (5.11) puis en intégrant par parties, il vient :

$$\begin{aligned} \int_0^T {}^t x(t)W(t)\delta x(t)dt &= \int_0^T {}^t x(t)W(t)M(t) \int_0^t M(s)^{-1}B(s)\delta u(s)ds dt \\ &= \int_0^T {}^t x(s)W(s)M(s)ds \int_0^T M(s)^{-1}B(s)\delta u(s)ds \\ &\quad - \int_0^T \int_0^t {}^t x(s)W(s)M(s)ds M(t)^{-1}B(t)\delta u(t) dt. \end{aligned}$$

Or

$$p(t) - \Lambda M(t)^{-1} = \int_0^t {}^t x(s)W(s)M(s)ds M(t)^{-1},$$

et d'après l'expression de Λ on arrive à :

$$\int_0^T {}^t x(t)W(t)\delta x(t)dt = -{}^t x(T)QM(T) \int_0^T M(t)^{-1}B(t)\delta u(t)dt - \int_0^T p(t)B(t)\delta u(t)dt.$$

Injectons cette égalité dans (5.12), en tenant compte du fait que :

$${}^t x(T)Q\delta x(T) = {}^t x(T)QM(T) \int_0^T M(t)^{-1}B(t)\delta u(t)dt.$$

On trouve alors que :

$$\frac{1}{2}dC(u).\delta u = \int_0^T ({}^t u(t)U(t) - p(t)B(t))\delta u(t) dt = 0,$$

ceci pour toute application $\delta u \in L^2([0, T], \mathbb{R}^m)$. Ceci implique donc l'égalité pour presque tout $t \in [0, T]$:

$${}^t u(t)U(t) - p(t)B(t) = 0,$$

ce qui est la conclusion souhaitée. Réciproquement s'il existe un vecteur adjoint $p(t)$ vérifiant (5.8) et (5.9) et si le contrôle u est donné par (5.10), alors il est bien clair d'après le raisonnement précédent que :

$$dC(u) = 0.$$

Or C étant strictement convexe ceci implique que u est un minimum global de C .

□

Remarque. Si le système de contrôle est perturbé par une fonction $r(t)$, alors le théorème précédent reste vrai. Il le reste, de même, si la fonction g apparaissant dans le coût est une fonction C^1 de \mathbb{R}^n dans \mathbb{R} vérifiant les conditions (g1) et (g2) (cf. remarque5), sauf que la condition finale sur le vecteur adjoint (5.9) devient :

$$p(T) = -\frac{1}{2}\nabla g(x(T)), \tag{5.13}$$

comme on le voit facilement dans la démonstration. Cette condition s'appelle *condition de transversalité*.

Remarque. Dans le cas d'un intervalle infini ($T = +\infty$) la condition devient :

$$\lim_{t \rightarrow +\infty} p(t) = 0. \tag{5.14}$$

Remarque. Définissons la fonction $H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ par :

$$H(x, p, u) = p(Ax + Bu) - \frac{1}{2}({}^t x W x + {}^t u U u),$$

en utilisant toujours la convention que p est un vecteur ligne de \mathbb{R}^n . Alors les équations données par le principe du maximum LQ s'écrivent :

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p} = Ax + Bu, \\ \dot{p} &= -\frac{\partial H}{\partial x} = -pA + {}^t x W, \end{aligned}$$

et

$$\frac{\partial H}{\partial u} = 0,$$

puisque $pB - {}^t u U = 0$. Ceci annonce le principe du maximum général. Mais en fait ici dans le cas LQ on peut dire mieux : d'une part le principe du maximum LQ est une condition nécessaire et suffisante de minimalité (alors que dans le cas général c'est une condition nécessaire seulement), d'autre part il est possible d'exprimer le contrôle sous forme de boucle fermée, grâce à la théorie de Riccati, voir section suivante.

Remarque. L'approche développée dans la démonstration du théorème 5.4 est variationnelle. Nous avons ici préféré l'approche du calcul des variations classique, car elle permet une preuve plus rapide et élégante. L'autre approche est en fait plus générale et sera privilégiée dans le cas général (non linéaire) où elle conduit au principe du maximum de Pontryagin général.

5.3 Fonction valeur et équation de Riccati

5.3.1 Définition de la fonction valeur

Soit $T > 0$ fixé, et soit $x \in \mathbb{R}^n$. Considérons le problème LQ de trouver une trajectoire solution de

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(0) = x, \quad (5.15)$$

minimisant le coût quadratique

$$C_T(u) = {}^t x(T)Qx(T) + \int_0^T (\|x(t)\|_W^2 + \|u(t)\|_U^2) dt. \quad (5.16)$$

On définit alors :

Définition 5.1. La fonction valeur S_T au point x est la borne inférieure des coûts pour le problème LQ. Autrement dit :

$$S_T(x) = \inf\{C_T(u) \mid x_u(0) = x\}.$$

Remarque. Sous l'hypothèse (5.3) on a existence d'une unique trajectoire optimale d'après le théorème 5.1, et dans ce cas cette borne inférieure est un minimum.

5.3.2 Equation de Riccati

Théorème 5.5. Sous l'hypothèse (5.3), pour tout $x \in \mathbb{R}^n$ il existe une unique trajectoire optimale x associée au contrôle u pour le problème (5.15), (5.16). Le contrôle optimal se met sous forme de boucle fermée :

$$u(t) = U(t)^{-1} {}^t B(t) E(t) x(t), \quad (5.20)$$

où $E(t) \in \mathcal{M}_n(\mathbb{R})$ est solution sur $[0, T]$ de l'équation matricielle de Riccati :

$$\dot{E}(t) = W(t) - {}^t A(t)E(t) - E(t)A(t) - E(t)B(t)U(t)^{-1} {}^t B(t)E(t), \quad E(T) = -Q. \quad (5.21)$$

De plus, pour tout $t \in [0, T]$, la matrice $E(t)$ est symétrique, et :

$$S_T(x) = -{}^t x E(0) x. \quad (5.22)$$

Remarque. En particulier le théorème affirme que le contrôle optimal u se met sous forme de *boucle fermée* :

$$u(t) = K(t)x(t),$$

où $K(t) = U(t)^{-1} {}^tB(t)E(t)$. Cette forme se prête bien aux problèmes de stabilisation, comme nous le verrons plus loin.

PREUVE.

▷ D'après le théorème 5.1, il existe une unique trajectoire optimale qui, d'après le théorème 5.4, est caractérisée par le système d'équations :

$$\begin{aligned}\dot{x} &= Ax + BU^{-1} {}^tB {}^t p, \\ \dot{p} &= -pA + {}^t x W,\end{aligned}$$

avec $x(0) = x$ et $p(T) = -{}^t x(T)Q$. De plus, le contrôle s'écrit :

$$u = U^{-1} {}^tB {}^t p.$$

Il faut donc montrer que l'on peut écrire $p(t) = {}^t x(t)E(t)$, où $E(t)$ est solution de (5.21). Notons que si p s'écrit ainsi alors d'après l'équation satisfaite par le couple (x, p) , on trouve facilement que $E(t)$ doit satisfaire l'équation (5.21). En utilisant l'unicité de la trajectoire optimale, on va maintenant montrer que p s'écrit effectivement ainsi. Soit $E(t)$ solution de l'équation :

$$\dot{E} = W - {}^tAE - EA - EBU^{-1} {}^tBE, \quad E(T) = -Q.$$

Tout d'abord $E(t)$ est symétrique car le second membre de l'équation différentielle l'est, et la matrice Q est symétrique. A priori on ne sait pas cependant que la solution est bien définie sur $[0, T]$ tout entier. On montrera cela plus loin (lemme 5.6).

Posons maintenant $p_1(t) = {}^t x_1(t)E(t)$, où x_1 est solution de

$$\dot{x}_1 = Ax_1 + Bu_1,$$

et $u_1 = U^{-1} {}^tBE x_1$. On a alors :

$$\begin{aligned}\dot{p}_1 &= \dot{{}^t x}_1 E + {}^t x_1 \dot{E} \\ &= {}^t(Ax_1 + BU^{-1} {}^tBE x_1)E + {}^t x_1 (W - {}^tAE - EA - EBU^{-1} {}^tBE) \\ &= -p_1 A + {}^t x_1 W.\end{aligned}$$

Autrement dit le triplet (x_1, p_1, u_1) vérifie exactement les équations du théorème 5.4. Par conséquent la trajectoire x_1 est optimale, et par unicité il vient $x_1 = x$, $u_1 = u$, puis $p_1 = p$. En particulier on a donc $p = {}^t x E$, et $u = U^{-1} {}^tBE x$. Déduisons-en la formule (5.22). Pour cela calculons d'abord, le long de la trajectoire $x(t)$:

$$\begin{aligned}\frac{d}{dt} {}^t x(t)E(t)x(t) &= \frac{d}{dt} p(t)x(t) = \dot{p}(t)x(t) + p(t)\dot{x}(t) \\ &= (-p(t)A(t) + {}^t x(t)W(t))x(t) + p(t)(A(t)x(t) + B(t)u(t)) \\ &= {}^t x(t)W(t)x(t) + p(t)B(t)u(t).\end{aligned}$$

Par ailleurs de l'expression de u on déduit :

$${}^t_u U u = {}^t(U^{-1}{}^t B E x) U U^{-1}{}^t B E x = {}^t_x E B U^{-1}{}^t B E x = p B u.$$

Finalement on a l'égalité :

$$\frac{d}{dt} {}^t_x(t) E(t) x(t) = {}^t_x(t) W(t) x(t) + {}^t_u(t) U(t) u(t),$$

et par conséquent :

$$S_T(x) = {}^t_x(T) Q x(T) + \int_0^T \frac{d}{dt} {}^t_x(t) E(t) x(t) dt.$$

Or puisque $E(T) = -Q$ et $x(0) = x$, il vient $S_T(x) = -{}^t_x E(0) x$.

Lemme 5.6. *L'application $t \mapsto E(t)$ est bien définie sur $[0, T]$ tout entier.*

PREUVE.

▷ [Preuve du lemme] Si l'application $E(t)$ n'est pas définie sur $[0, T]$ entier, alors il existe $0 < t_* < T$ tel que $\|E(t)\|$ tend vers $+\infty$ lorsque t tend vers t_* par valeurs supérieures. En particulier pour tout $\alpha > 0$ il existe $t_0 \in]t_*, T]$ et $x_0 \in \mathbb{R}^n$, avec $\|x_0\| = 1$, tels que

$$|{}^t_{x_0} E(t_0) x_0| \geq \alpha. \quad (5.23)$$

D'après le théorème 5.1, il existe une unique trajectoire optimale $x(\cdot)$ pour le problème LQ sur $[t_0, T]$, telle que $x(t_0) = x_0$ (voir remarque 5). Cette trajectoire est caractérisée par le système d'équations :

$$\begin{aligned} \dot{x} &= Ax + BU^{-1}{}^t B {}^t p, & x(t_0) &= x_0, \\ \dot{p} &= -pA + {}^t_x W, & p(T) &= -{}^t_x(T) Q. \end{aligned}$$

Il résulte du théorème de dépendance continue des solutions d'une équation différentielle par rapport à la condition initiale que les extrémités $x(T)$ au temps T des trajectoires issues au temps t_0 de x_0 , sont uniformément bornées lorsque $0 \leq t_0 < T$ et $\|x_0\| = 1$, et donc les solutions correspondantes $x(t), p(t)$ du système différentiel précédent sont uniformément bornées sur $[0, T]$. En particulier la quantité $p(t_0)x(t_0)$ doit être bornée indépendamment de t_0 . Or on sait que $p(t) = {}^t_x(t) E(t)$, donc :

$$p(t_0)x(t_0) = {}^t_{x_0} E(t_0) x_0,$$

et on obtient une contradiction avec (5.23). □

Ceci achève la preuve du théorème. □

Remarque. Il est clair d'après l'expression (5.22) du coût minimal que la matrice $E(0)$ est symétrique négative. On peut améliorer ce résultat si la matrice Q est de plus définie :

Lemme 5.7. *Si la matrice Q est symétrique définie positive, ou bien si pour tout $t \in [0, T]$ la matrice $W(t)$ est symétrique définie positive, alors la matrice $E(0)$ est symétrique définie négative.*

PREUVE.

▷ [Preuve du lemme 5.7.] Soit x_0 tel que ${}^t x_0 E(0) x_0 = 0$, et montrons que $x_0 = 0$. Pour cela on considère le problème LQ :

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = x_0, \\ \min \quad & {}^t x(T) Q x(T) + \int_0^T (\|x(t)\|_W^2 + \|u(t)\|_U^2) dt, \end{aligned}$$

pour lequel, d'après le théorème 5.5, le coût minimal vaut $-{}^t x_0 E(0) x_0 = 0$. Par conséquent, puisque pour tout t la matrice $U(t)$ est définie positive, on a $u(t) = 0$ sur $[0, T]$. Si par ailleurs Q est définie positive on a aussi $x(T) = 0$. Donc la trajectoire $x(\cdot)$ est solution du problème de Cauchy $\dot{x} = Ax, x(T) = 0$, et par unicité $x(\cdot)$ est identiquement nulle. En particulier $x(0) = x_0 = 0$, ce qui achève la preuve. Dans le deuxième cas où $W(t)$ est définie positive, la conclusion est immédiate. □

Variante du problème précédent. Soit $T > 0$ fixé. Pour tout $t \in [0, T]$ et tout $x \in \mathbb{R}^n$, considérons le problème LQ qui consiste à trouver une trajectoire solution de

$$\dot{x} = Ax + Bu, \quad x(t) = x, \tag{5.24}$$

minimisant le coût quadratique

$$C_T(t, u) = g(x(T)) + \int_t^T \lambda (\|x(t)\|_W^2 + \|u(t)\|_U^2 \rho) dt. \tag{5.25}$$

On définit alors :

Définition 5.2. La fonction valeur S au point (t, x) est la borne inférieure des coûts pour ce problème LQ. Autrement dit :

$$S_T(t, x) = \inf \{ C_T(t, u) \mid x_u(t) = x \}.$$

Théorème 5.8. *Sous l'hypothèse (5.3), pour tout $x \in \mathbb{R}^n$ et tout $t \in [0, T]$ il existe une unique trajectoire optimale x associée au contrôle u pour le problème (5.24), (5.25). Le contrôle optimal se met sous forme de boucle fermée :*

$$u(s) = U(s)^{-1} {}^t B(s) E(s) x(s), \quad (5.29)$$

pour tout $s \in [t, T]$, et où $E(s) \in \mathcal{M}_n(\mathbb{R})$ est solution sur $[t, T]$ de l'équation matricielle de Riccati :

$$\dot{E} = W - {}^t A E - E A - E B U^{-1} {}^t B E, \quad E(T) = -Q. \quad (5.30)$$

De plus, pour tout $s \in [t, T]$, la matrice $E(s)$ est symétrique, et pour tout $t \in [0, T]$ on a :

$$S_T(t, x) = -{}^t x E(t) x. \quad (5.31)$$

PREUVE.

▷ La différence par rapport au cas précédent est que l'on paramétrise le temps initial. Le seul changement est donc la formule (5.31). Comme dans la démonstration précédente, on a :

$$S_T(t, x) = {}^t x(T) Q x(T) + \int_t^T \frac{d}{ds} {}^t x(s) E(s) x(s) ds.$$

Or puisque $E(T) = -Q$ et $x(t) = x$, il vient $S_T(t, x) = -{}^t x E(t) x$.

□

Remarque. L'équation de Riccati étant fondamentale, notamment dans les problèmes de régulateur (voir section suivante), la question de son implémentation numérique se pose naturellement. On peut procéder de manière directe : il s'agit alors, en tenant compte du fait que $E(t)$ est symétrique, d'intégrer un système différentiel non linéaire de $n(n+1)/2$ équations.

5.3.3 Représentation linéaire de l'équation de Riccati

On a la propriété suivante.

Proposition 5.9. *Plaçons-nous dans le cadre du théorème 5.5. Soit*

$$R(t) = \begin{pmatrix} R_1(t) & R_2(t) \\ R_3(t) & R_4(t) \end{pmatrix}$$

la résolvante du système linéaire

$$\begin{aligned} \dot{x} &= Ax + BU^{-1}{}^tB{}^t p, \\ {}^t\dot{p} &= -{}^tA{}^t p + Wx, \end{aligned}$$

telle que $R(T) = Id$. Alors pour tout $t \in [0, T]$ on a :

$$E(t) = (R_3(t) - R_4(t)Q)(R_1(t) - R_2(t)Q)^{-1}.$$

PREUVE.

▷ Par définition de la résolvante on a :

$$\begin{aligned} x(t) &= R_1(t)x(T) + R_2(t){}^t p(T), \\ {}^t p(t) &= R_3(t)x(T) + R_4(t){}^t p(T). \end{aligned}$$

Or on sait que ${}^t p(T) = -Qx(T)$, donc :

$$x(t) = (R_1(t) - R_2(t)Q)x(T) \quad \text{et} \quad {}^t p(t) = (R_3(t) - R_4(t)Q)x(T).$$

On conclut en remarquant que ${}^t p(t) = E(t)x(t)$. Notons que la matrice $R_1(t) - R_2(t)Q$ est inversible sur $[0, T]$ car le problème LQ est bien posé, comme nous l'avons vu précédemment. \square

Par conséquent pour résoudre l'équation de Riccati (5.21), il suffit d'intégrer un système linéaire (il faut calculer une résolvante), ce qui est très facile à programmer. Cette méthode (due à Kalman-Englar) est notamment préférable à la méthode directe dans le cas stationnaire.

5.4 Applications de la théorie LQ

5.4.1 Problèmes de régulation

Le problème du régulateur d'état (ou "problème d'asservissement", ou "problème de poursuite", en anglais "tracking problem")

Considérons le système de contrôle linéaire perturbé :

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + r(t), \quad x(0) = x_0, \quad (5.32)$$

et soit $\xi(t)$ une certaine trajectoire de \mathbb{R}^n sur $[0, T]$, partant d'un point ξ_0 (et qui n'est pas forcément solution du système (5.32)). Le but est de déterminer un contrôle tel que la trajectoire associée, solution de (5.32), suive le mieux possible la trajectoire de référence $\xi(t)$.

On introduit alors l'erreur sur $[0, T]$:

$$z(t) = x(t) - \xi(t),$$

qui est solution du système de contrôle :

$$\dot{z}(t) = A(t)z(t) + B(t)u(t) + r_1(t), \quad z(0) = z_0, \quad (5.33)$$

où $z_0 = x_0 - \xi_0$ et $r_1(t) = A(t)\xi(t) - \dot{\xi}(t) + r(t)$. Il est alors raisonnable de vouloir minimiser le coût :

$$C(u) = {}^t z(T)Qz(T) + \int_0^T (\|z(t)\|_W^2 + \|u(t)\|_U^2) dt,$$

où Q, W, U sont des matrices de pondération. Pour absorber la perturbation r_1 , on augmente le système d'une dimension, en posant :

$$z_1 = \begin{pmatrix} z \\ 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} A & r_1 \\ 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} B \\ 0 \end{pmatrix}, \quad Q_1 = \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix}, \quad W_1 = \begin{pmatrix} W & 0 \\ 0 & 0 \end{pmatrix},$$

de sorte que l'on se ramène à minimiser le coût

$$C(u) = {}^t z_1(T)Q_1 z_1(T) + \int_0^T (\|z_1(t)\|_{W_1}^2 + \|u(t)\|_U^2) dt,$$

pour le système de contrôle

$$\dot{z}_1 = A_1 z_1 + B_1 u,$$

partant du point $z_1(0)$.

La théorie LQ faite précédemment prévoit alors que le contrôle optimal existe, est unique, et s'écrit

$$u(t) = U(t)^{-1} {}^t B_1(t) E_1(t) z_1(t),$$

où $E_1(t)$ est solution de l'équation de Riccati :

$$\dot{E}_1 = W_1 - {}^t A_1 E_1 - E_1 A_1 - E_1 B_1 U^{-1} {}^t B_1 E_1, \quad E_1(T) = -Q_1.$$

Posons :

$$E_1(t) = \begin{pmatrix} E(t) & h(t) \\ {}^t h(t) & \alpha(t) \end{pmatrix}.$$

En remplaçant dans l'équation précédente, on établit facilement les équations différentielles de E, h, α :

$$\begin{aligned} \dot{E} &= W - {}^t A E - E A - E B U^{-1} {}^t B E, & E(T) &= -Q, \\ \dot{h} &= -{}^t A h - E r_1 - E B U^{-1} {}^t B h, & h(T) &= 0, \\ \dot{\alpha} &= -2 {}^t r_1 h - {}^t h B U^{-1} {}^t B h, & \alpha(T) &= 0. \end{aligned} \quad (5.34)$$

Résumons tout ceci dans la proposition suivante.

Proposition 5.10. *Soit ξ une trajectoire de \mathbb{R}^n sur $[0, T]$, et considérons le problème de poursuite pour le système de contrôle :*

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

où l'on veut minimiser le coût :

$$C(u) = {}^t(x(T) - \xi(T))Q(x(T) - \xi(T)) + \int_0^T (\|x(t) - \xi(t)\|_W^2 + \|u(t)\|_U^2) dt.$$

Alors il existe un unique contrôle optimal, qui s'écrit :

$$u(t) = U(t)^{-1}{}^tB(t)E(t)(x(t) - \xi(t)) + U(t)^{-1}{}^tB(t)h(t),$$

où $E(t) \in \mathcal{M}_n(\mathbb{R})$ et $h(t) \in \mathbb{R}^n$ sont solutions sur $[0, T]$ de

$$\begin{aligned} \dot{E} &= W - {}^tAE - EA - EBU^{-1}{}^tBE, & E(T) &= -Q, \\ \dot{h} &= -{}^tAh - E(A\xi - \dot{\xi} + r) - EBU^{-1}{}^tBh, & h(T) &= 0, \end{aligned}$$

et de plus $E(t)$ est symétrique. Par ailleurs le coût minimal est alors égal à

$$\begin{aligned} &- {}^t(x(0) - \xi(0))E(0)(x(0) - \xi(0)) - 2{}^th(0)(x(0) - \xi(0)) \\ &- \int_0^T \left(2{}^t(A(t)\xi(t) - \dot{\xi}(t) + r(t))h(t) + {}^th(t)B(t)U(t)^{-1}{}^tB(t)h(t) \right) dt. \end{aligned}$$

Remarque. Notons que le contrôle optimal s'écrit bien sous forme de boucle fermée

$$u(t) = K(t)(x(t) - \xi(t)) + H(t).$$

Remarque. Si $\dot{\xi} = A\xi + r$, i.e. la trajectoire de référence est solution du système sans contrôle, alors dans les notations précédentes on a $r_1 = 0$, et d'après les équations (5.34) on en déduit que $h(t)$ et $\alpha(t)$ sont identiquement nuls. On retrouve alors le cadre LQ de la section précédente. En fait :

- Si $\xi = 0$ et $r = 0$, le problème est un problème LQ standard.
- Si $r = 0$, il s'agit d'un problème de poursuite de la trajectoire ξ .
- Si $\xi = 0$, c'est un problème de régulation avec la perturbation r .

Variante : le problème de poursuite d'une sortie (ou "output tracking")

On ajoute au problème précédent une variable de sortie :

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) + r(t), \quad x(0) = x_0, \\ y(t) &= C(t)x(t), \end{aligned}$$

et étant donné un signal de référence $\xi(t)$ on cherche un contrôle tel que, le long de la trajectoire associée, l'observable $z(\cdot)$ soit proche de $\xi(\cdot)$. Notons qu'on retrouve le cas précédent si $y(t) = x(t)$.

Posant $z(t) = y(t) - \xi(t)$, on cherche à minimiser le coût :

$$C(u) = {}^t z(T)Qz(T) + \int_0^T (\|z(t)\|_W^2 + \|u(t)\|_U^2) dt.$$

Posons alors :

$$x_1 = \begin{pmatrix} x \\ 1 \end{pmatrix}, \quad Q_1 = \begin{pmatrix} {}^t C(T)QC(T) & -{}^t C(T)Q\xi(T) \\ -{}^t \xi(T)QC(T) & {}^t \xi(T)Q\xi(T) \end{pmatrix}, \quad W_1 = \begin{pmatrix} {}^t CWC & -{}^t CW\xi \\ -{}^t \xi WC & {}^t \xi W\xi \end{pmatrix},$$

et A_1, B_1 comme précédemment (avec $r_1 = r$). Alors on cherche un contrôle u , associé à la trajectoire x_1 solution de $\dot{x}_1 = A_1 x_1 + B_1 u$, minimisant le coût

$$C(u) = {}^t x_1(T)Q_1 x_1(T) + \int_0^T (\|x_1(t)\|_{W_1}^2 + \|u(t)\|_U^2) dt.$$

En raisonnant comme précédemment, on arrive au résultat suivant.

Proposition 5.11. Soit ξ une trajectoire de \mathbb{R}^p sur $[0, T]$, et considérons le problème de poursuite de la sortie r pour le système de contrôle avec sortie :

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) + r(t), \quad x(0) = x_0, \\ y(t) &= C(t)x(t), \end{aligned}$$

où l'on veut minimiser le coût :

$$C(u) = {}^t(y(T) - \xi(T))Q(y(T) - \xi(T)) + \int_0^T (\|y(t) - \xi(t)\|_W^2 + \|u(t)\|_U^2) dt.$$

Alors il existe un unique contrôle optimal, qui s'écrit :

$$u(t) = U(t)^{-1}{}^tB(t)E(t)x(t) + U(t)^{-1}{}^tB(t)h(t),$$

où $E(t) \in \mathcal{M}_n(\mathbb{R})$ et $h(t) \in \mathbb{R}^p$ sont solutions sur $[0, T]$ de

$$\begin{aligned} \dot{E} &= {}^tCWC - {}^tAE - EA - EBU^{-1}{}^tBE, \quad E(T) = -{}^tC(T)QC(T), \\ \dot{h} &= -{}^tCW\xi - {}^tAh - Er - EBU^{-1}{}^tBh, \quad h(T) = -{}^tC(T)Q\xi(T), \end{aligned}$$

et de plus $E(t)$ est symétrique. Par ailleurs le coût minimal est alors égal à

$$-{}^tx(0)E(0)x(0) - 2{}^th(0)x(0) - \alpha(0),$$

où $\alpha(t)$ est solution de

$$\dot{\alpha} = {}^t\xi W \xi - 2{}^trh - {}^thBU^{-1}{}^tBh, \quad \alpha(T) = {}^t\xi(T)Q\xi(T).$$

Remarque. Il existe d'autres variantes de ce problème, notamment le même problème que ci-dessus, sauf que le coût s'écrit :

$$C(u) = {}^tx(T)Qx(T) + \int_0^T (\|y(t) - \xi(t)\|_W^2 + \|u(t)\|_U^2) dt.$$

Le seul changement est dans la matrice augmentée Q_1 , et donc dans les conditions aux limites de E et h , qui deviennent dans ce cas : $E(T) = -Q$ et $h(T) = 0$.

Enfin, il y a aussi une autre variante du problème LQ, celle où la fonction g apparaissant dans le coût est linéaire en x . Nous laissons l'écriture de toutes ces variantes au lecteur, la méthode étant de toute façon la même que précédemment.

5.4.2 Filtre de Kalman déterministe

Ce problème célèbre est le suivant. Connaissant un signal de référence $\xi(t)$ sur $[0, T]$, on cherche une trajectoire solution sur $[0, T]$ de

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

minimisant le coût

$$C(u) = {}^t x(0)Qx(0) + \int_0^T (\|(C(t)x(t) - \xi(t))\|_W^2 + \|u(t)\|_U^2) dt.$$

Il s'agit d'une variante des problèmes de poursuite précédents, sauf que l'on n'impose aucune condition sur $x(0)$ et $x(T)$, et de plus le coût pénalise le point initial $x(0)$. En revanche dans ce problème on suppose que la matrice Q est symétrique *définie* positive.

Pour se ramener aux cas précédents, il convient donc tout d'abord d'inverser le temps, de façon à ce que le coût pénalise, comme avant, le point final. On pose donc, pour tout $t \in [0, T]$:

$$\begin{aligned} \tilde{x}(t) &= x(T-t), \quad \tilde{u}(t) = u(T-t), \quad \tilde{A}(t) = -A(T-t), \quad \tilde{B}(t) = -B(T-t), \\ \tilde{\xi}(t) &= \xi(T-t), \quad \tilde{W}(t) = W(T-t), \quad \tilde{U}(t) = U(T-t), \quad \tilde{C}(t) = C(T-t), \end{aligned}$$

de sorte que l'on se ramène au problème de déterminer une trajectoire solution de $\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\tilde{u}$, minimisant le coût :

$$\tilde{C}(\tilde{u}) = {}^t \tilde{x}(T)Q\tilde{x}(T) + \int_0^T (\|(\tilde{C}(t)\tilde{x}(t) - \tilde{\xi}(t))\|_{\tilde{W}}^2 + \|\tilde{u}(t)\|_{\tilde{U}}^2) dt.$$

Notons que, par construction, on a $\tilde{C}(\tilde{u}) = C(u)$.

Fixons une donnée initiale $\tilde{x}(0)$, et appliquons, pour cette donnée initiale, le même raisonnement que dans les cas précédents. On obtient alors :

$$\tilde{u}(t) = \tilde{U}^{-1} {}^t \tilde{B} \tilde{E} \tilde{x} + \tilde{U}^{-1} {}^t \tilde{B} \tilde{h},$$

où :

$$\begin{aligned} \dot{\tilde{E}} &= {}^t \tilde{C} \tilde{W} \tilde{C} - {}^t \tilde{A} \tilde{E} - \tilde{E} \tilde{A} - \tilde{E} \tilde{B} \tilde{U}^{-1} {}^t \tilde{B} \tilde{E}, & \tilde{E}(T) &= -Q, \\ \dot{\tilde{h}} &= -{}^t \tilde{C} \tilde{W} \tilde{\xi} - {}^t \tilde{A} \tilde{h} - \tilde{E} \tilde{B} \tilde{U}^{-1} {}^t \tilde{B} \tilde{h}, & \tilde{h}(T) &= 0, \\ \dot{\tilde{\alpha}} &= {}^t \tilde{\xi} \tilde{W} \tilde{\xi} - {}^t \tilde{h} \tilde{B} \tilde{U}^{-1} {}^t \tilde{B} \tilde{h}, & \tilde{\alpha}(T) &= 0, \end{aligned}$$

et le coût minimal pour cette donnée initiale fixée $\tilde{x}(0)$ vaut :

$$-{}^t \tilde{x}(0) \tilde{E}(0) \tilde{x}(0) - 2 {}^t \tilde{x}(0) \tilde{h}(0) - \tilde{\alpha}(0).$$

Il faut maintenant trouver $\tilde{x}(0)$ tel que ce coût soit minimal. Posons donc :

$$f(x) = -{}^t x \tilde{E}(0) x - 2 {}^t x \tilde{h}(0) - \alpha(0).$$

Il faut donc déterminer un minimum de f . Notons tout d'abord que, la matrice Q étant par hypothèse définie positive, la matrice $\tilde{E}(0)$ est d'après le lemme 5.7 symétrique définie négative. En particulier la fonction f est strictement convexe et de ce fait admet un unique minimum. En un tel point on doit avoir $f'(x) = 0$, d'où $x = -\tilde{E}(0)^{-1}\tilde{h}(0)$.

Finalement, en reprenant le cours positif du temps, et en posant pour tout $t \in [0, T]$:

$$E(t) = -\tilde{E}(T-t), \quad h(t) = -\tilde{h}(T-t),$$

on arrive au résultat suivant.

Proposition 5.12. *Soit $\xi(\cdot)$ une trajectoire définie sur $[0, T]$ à valeurs dans \mathbb{R}^p . On considère le problème de déterminer une trajectoire solution sur $[0, T]$ de*

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

minimisant le coût

$$C(u) = {}^t x(0)Qx(0) + \int_0^T (\|C(t)x(t) - \xi(t)\|_W^2 + \|u(t)\|_U^2) dt,$$

où la matrice Q est de plus supposée définie positive. Alors il existe une unique trajectoire minimisante, associée au contrôle

$$u(t) = U(t)^{-1}{}^t B(t)E(t)x(t) + U(t)^{-1}{}^t B(t)h(t),$$

et à la condition finale

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

où

$$\begin{aligned} \dot{E} &= {}^t CWC - {}^t AE - EA - EBU^{-1}{}^t BE, & E(0) &= Q, \\ \dot{h} &= -{}^t CW\xi - {}^t Ah - EBU^{-1}{}^t Bh, & h(0) &= 0, \end{aligned}$$

et le coût minimal vaut alors :

$$-{}^t h(T)E(T)^{-1}h(T) + \int_0^T ({}^t \xi(t)W(t)\xi(t) - {}^t h(t)B(t)U(t)^{-1}{}^t B(t)h(t)) dt.$$

L'état final $x(T) = -E(T)^{-1}h(T)$ est la donnée qui nous intéresse principalement dans le problème du filtre de Kalman, qui est un problème d'estimation. L'estimation de cet état final peut être simplifiée de la manière suivante.

Posons $F(t) = E(t)^{-1}$. On trouve facilement, puisque $\dot{F} = -F\dot{E}F$:

$$\dot{F} = BU^{-1}{}^t B + AF + F{}^t A - F{}^t CWCF, \quad F(0) = Q^{-1}.$$

Par ailleurs si on pose $z(t) = -F(t)h(t)$, on trouve que :

$$\dot{z} = (A - F{}^t CWCF)z + F{}^t CW\xi, \quad z(0) = 0.$$

Finalement on arrive au résultat suivant.

Proposition 5.13. *Sous les hypothèses de la proposition 5.12, l'état final $x(T)$ de la solution optimale est égal à $z(T)$, où*

$$\begin{aligned} \dot{z} &= (A - F^t C W C)z + F^t C W \xi, & z(0) &= 0, \\ \dot{F} &= B U^{-1} {}^t B + A F + F^t A - F^t C W C F, & F(0) &= Q^{-1}. \end{aligned}$$

Application au filtrage. Le problème est d'estimer, d'après une observation, un signal bruité. Le modèle est le suivant :

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t), & x(0) &= x_0, \\ \xi(t) &= y(t) + v(t), \end{aligned}$$

où $y(t) = C(t)x(t)$ et les fonctions u et v sont des *bruits*, i.e. des perturbations affectant le système. La donnée initiale x_0 est inconnue. Le signal $\xi(t)$ représente une observation de la variable $y(t)$, et à partir de cette observation on veut construire une estimation de l'état final $x(T)$. On cherche une estimation optimale dans le sens que les perturbations u et v , ainsi que la donnée initiale x_0 , doivent être aussi petites que possible. On cherche donc à minimiser un coût de la forme :

$${}^t x(0) Q x(0) + \int_0^T (\|v(t)\|_W^2 + \|u(t)\|_U^2) dt.$$

Il s'agit donc exactement du problème LQ suivant :

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t), \\ y(t) &= C(t)x(t), \\ C(u) &= {}^t x(0) Q x(0) + \int_0^T (\|y(t) - \xi(t)\|_W^2 + \|u(t)\|_U^2) dt, \end{aligned}$$

i.e. le problème que l'on vient d'étudier ($x(0)$ non fixé).

L'estimation optimale de l'état est donc égale à $z(T)$, voir proposition 5.13.

Remarque. La bonne manière d'interpréter le filtre de Kalman est statistique, ce qui dépasse le cadre de cet ouvrage. En fait il faut interpréter les perturbations u et b comme des bruits blancs gaussiens, et x_0 comme une variable aléatoire gaussienne, tous supposés centrés en 0 (pour simplifier). Les matrices $Q, W(t), U(t)$ sont alors les matrices de variance de $x_0, v(t), u(t)$, et le problème de minimisation s'interprète comme le problème d'estimer l'état final de variance minimale, connaissant l'observation $\xi(t)$.

Par ailleurs les pondérations doivent être choisies en fonction de l'importance des bruits. Par exemple si le bruit v est très important comparé au bruit u et à l'incertitude sur la condition initiale alors on choisit une matrice $W(t)$ petite.

5.4.3 Régulation sur un intervalle infini et rapport avec la stabilisation

Considérons le problème LQ sur l'intervalle $[0, +\infty[$. Il s'agit d'un problème de régulation où l'on cherche à rendre l'erreur petite pour tout temps. Nous nous restreignons au cas de systèmes stationnaires. Le cadre est le suivant.

On cherche à déterminer une trajectoire solution de

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0,$$

minimisant le coût

$$C(u) = \int_0^\infty (\|x(t)\|_W^2 + \|u(t)\|_U^2) dt,$$

où de même les matrices W et U sont constantes.

On a le résultat suivant.

Théorème 5.14. *On suppose que les matrices W et U sont symétriques définies positives, et que le système est contrôlable. Alors il existe une unique trajectoire minimisante pour ce problème, associée sur $[0, +\infty[$ au contrôle optimal :*

$$u(t) = U^{-1} {}^t B E x(t), \quad (5.37)$$

où $E \in \mathcal{M}_n(\mathbb{R})$ est l'unique matrice symétrique définie négative solution de l'équation de Riccati stationnaire :

$${}^t A E + E A + E B U^{-1} {}^t B E = W. \quad (5.38)$$

De plus le coût minimal vaut $-{}^t x_0 E x_0$. Par ailleurs le système bouclé

$$\dot{x} = (A + B U^{-1} {}^t B E)x$$

est globalement asymptotiquement stable, et la fonction $V(x) = -{}^t x E x$ est une fonction de Lyapunov stricte pour ce système.

Remarque. En particulier, la trajectoire minimisante associée à ce problème en horizon infini tend vers 0 lorsque t tend vers l'infini.

PREUVE.

▷ On sait déjà (voir proposition 5.3 et remarque 5.2) qu'il existe une unique trajectoire optimale, vérifiant les équations :

$$\dot{x} = Ax + Bu, \quad \dot{p} = -pA + {}^t x W, \quad \lim_{t \rightarrow +\infty} p(t) = 0,$$

avec $u = U^{-1} {}^t B p$. De manière tout à fait similaire à la preuve du théorème 5.4 on montre, par un argument d'unicité, que $p(t) = {}^t x(t)E$, où E est solution, pourvu qu'elle existe, de l'équation (5.38). Il faut donc montrer l'existence d'une telle solution. C'est l'objet du lemme :

Lemme 5.15. *Il existe une unique matrice E symétrique définie négative solution de l'équation (5.38).*

PREUVE.

▷ [Preuve du lemme.] Il est bien clair que si $x(\cdot)$ est minimisante pour le problème LQ sur $[0, +\infty[$, alors elle l'est aussi sur chaque intervalle $[0, T]$, $T > 0$. Considérons donc le problème LQ sur $[0, T]$:

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = x_0, \\ C(T, u) &= \int_0^T (\|x(t)\|_W^2 + \|u(t)\|_U^2) dt, \end{aligned}$$

et appelons $E(T, t)$ la solution de l'équation de Riccati associée :

$$\dot{E} = W - {}^t A E - E A - E B U^{-1} {}^t B E, \quad E(T, T) = 0.$$

On sait que de plus le coût minimal est $C(T, u) = -{}^t x_0 E(T, 0) {}^t x_0$. Posons alors $D(T, t) = -E(T, T - t)$. Il est bien clair que :

$$\dot{D} = W + {}^t A D + D A - D B U^{-1} {}^t B D, \quad D(T, 0) = 0.$$

Cette équation étant en fait indépendante de T , on peut poser $D(t) = D(T, t)$, et $D(t)$ est solution de l'équation de Riccati ci-dessus sur \mathbb{R}^+ . De plus pour tout $T > 0$ on a $D(T) = -E(T, 0)$, et comme la matrice W est symétrique définie positive on déduit du lemme 5.7 que $D(T)$ est symétrique définie positive.

Par ailleurs on a, pour tout $T > 0$: $C(T, u) = {}^t x_0 D(T) x_0$. Il est clair que si $0 < t_1 \leq t_2$ alors $C(t_1, u) \leq C(t_2, u)$, et donc ${}^t x_0 D(t_1) x_0 \leq {}^t x_0 D(t_2) x_0$. Ceci est en fait indépendant de x_0 , car l'équation de Riccati ne dépend nullement de la donnée initiale. Ainsi pour tout $x \in \mathbb{R}^n$ la fonction $t \mapsto {}^t x D(t) x$ est croissante.

Montrons qu'elle est également majorée. Le système étant contrôlable, l'argument de la remarque 5.1 montre qu'il existe au moins un contrôle v sur $[0, +\infty[$ de coût fini. Comme le contrôle u est optimal, on en déduit que la fonction $t \mapsto C(t, u)$ est majorée (par $C(v)$). Pour tout $x \in \mathbb{R}^n$, la fonction $t \mapsto {}^t x D(t) x$ étant croissante et majorée, on en déduit qu'elle converge. En appliquant cette conclusion aux éléments d'une base (e_i) de \mathbb{R}^n , on en déduit que chaque élément $d_{ij}(t)$ de la matrice $D(t)$ converge, car en effet :

$$d_{ij}(t) = {}^t e_i D(t) e_j = \frac{1}{2} {}^t e_i + e_j D(t) (e_i + e_j) - {}^t e_i D(t) e_i - {}^t e_j D(t) e_j.$$

Ainsi la matrice $D(t)$ converge vers une matrice $-E$, qui est nécessairement symétrique définie négative d'après la croissance de la fonction $t \mapsto {}^t x D(t) x$.

Par ailleurs de l'équation différentielle vérifiée par D on déduit que $\dot{D}(t)$ converge, et cette limite est alors nécessairement nulle. En passant à la limite dans cette équation différentielle on obtient finalement l'équation de Riccati stationnaire (5.38).

Enfin, en passant à la limite on a $C(u) = -{}^t x_0 E x_0$, d'où on déduit aisément l'unicité de la solution. □

Pour montrer la deuxième partie du théorème, il suffit de voir que la fonction $V(x) = -{}^t x E x$ est une fonction de Lyapunov pour le système bouclé $\dot{x} = (A + B U^{-1} {}^t B E)x$. La forme quadratique V est bien définie positive puisque E est symétrique définie négative. Par ailleurs on calcule facilement le long d'une trajectoire $x(t)$ solution du système bouclé :

$$\frac{d}{dt} V(x(t)) = -{}^t x(t) (W + E B U^{-1} {}^t B E) x(t).$$

Or la matrice W est par hypothèse définie positive, et la matrice $E B U^{-1} {}^t B E$ est positive, donc cette quantité est strictement négative si $x(t) \neq 0$. On a donc bien une fonction de Lyapunov stricte, ce qui prouve que le système bouclé est asymptotiquement stable. □

Remarque. Le contrôle optimal s'écrit sous forme de boucle fermée $u = Kx$, avec $K = U^{-1} {}^t B E$. On retrouve le fait que si le système est contrôlable alors il est stabilisable par feedback linéaire (voir le théorème de placement de pôles). Cependant, alors que la méthode de stabilisation décrite par le théorème de placement de pôles consiste à réaliser un placement de pôles, ici la matrice K est choisie de manière à minimiser un certain critère. On parle de stabilisation par retour d'état optimal. C'est donc une méthode (parmi beaucoup d'autres) de stabilisation.

Remarque. En général l'équation (5.38) admet plusieurs solutions, mais elle n'admet qu'une seule solution symétrique définie négative.

Chapitre 6

Temps-optimalité pour les systèmes linéaires

6.1 Existence de trajectoires temps-optimales

Avant de traiter à proprement parler du sujet de ce chapitre, prouvons tout d'abord un résultat à propos de l'ensemble atteignable en temps fini pour le système de contrôle dans \mathbb{R}^n :

$$\dot{x}(t) = A(t)x(t) + b(t)u(t) + r(t),$$

où les contrôles u sont à valeurs dans un compact d'intérieur non vide $\Omega \subset \mathbb{R}^m$.

Théorème 6.1. *Considérons le système de contrôle linéaire dans \mathbb{R}^n :*

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

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

Remarque. La convexité de $\mathcal{A}(x_0, t)$ est facile à établir si Ω est convexe (petit exercice laissé au lecteur). Pourtant, et ce résultat est surprenant, la conclusion de ce théorème est encore vraie si Ω n'est pas convexe et nous admettrons ce résultat. Mentionnons juste le fait que la preuve de la convexité nécessite un lemme de Lyapunov en théorie de la mesure. Ce résultat implique en particulier le corollaire suivant très utile.

Corollaire 6.2. Si on note $\mathcal{A}_\Omega(x_0, t)$ l'ensemble accessible depuis x_0 en temps t pour des contrôles à valeurs dans Ω , alors on a :

$$\mathcal{A}_\Omega(x_0, t) = \mathcal{A}_{\text{Conv}(\Omega)}(x_0, t),$$

où $\text{Conv}(\Omega)$ est l'enveloppe convexe de Ω . En particulier, on a $\mathcal{A}_{\partial\Omega}(x_0, t) = \mathcal{A}_\Omega(x_0, t)$.

Démonstration du théorème 6.1. Il reste à montrer que $\mathcal{A}(x_0, t)$ est compact. Grâce au corollaire précédent, on peut donc supposer sans perdre en généralité que, dans la suite de la preuve, Ω est convexe. Cela simplifie grandement la preuve de la compacité.

Il s'agit de montrer que toute suite (x_n) de points de $\mathcal{A}(x_0, t)$ admet une sous-suite convergente. Pour tout entier n soit u_n un contrôle reliant x_0 à x_n en temps t , et soit $x_n(\cdot)$ la trajectoire correspondante. On a donc :

$$x_n = x_n(t) = M(t)x_0 + \int_0^t M(t)M(s)^{-1}(B(s)u_n(s) + r(s))ds. \quad (6.1)$$

Par définition les contrôles u_n sont à valeurs dans le compact convexe Ω , et par conséquent la suite (u_n) est bornée dans $L^\infty([0, t], \Omega)$ qui est compact pour la topologie faible-*. En conséquence (u_n) converge faiblement, à une sous-suite près, vers une fonction $u \in L^\infty([0, t], \Omega)$. Il est alors immédiat de montrer que (x_n) converge, à une sous-suite près, vers

$$x = M(t)x_0 + \int_0^t M(t)M(s)^{-1}(B(s)u(s) + r(s))ds,$$

ce qui prouve la compacité de $\mathcal{A}(x_0, t)$.

Montrons enfin la continuité par rapport à t de $\mathcal{A}(x_0, t)$. Soit $\varepsilon > 0$. On va chercher $\delta > 0$ tel que :

$$\forall t_1, t_2 \in [0, T] \quad |t_1 - t_2| \leq \delta \Rightarrow d(\mathcal{A}(t_1), \mathcal{A}(t_2)) \leq \varepsilon,$$

où on note pour simplifier $\mathcal{A}(t) = \mathcal{A}(x_0, t)$, et où :

$$d(\mathcal{A}(t_1), \mathcal{A}(t_2)) = \sup \lambda \left(\sup_{y \in \mathcal{A}(t_2)} d(y, \mathcal{A}(t_1)), \sup_{y \in \mathcal{A}(t_1)} d(y, \mathcal{A}(t_2)) \right).$$

Par la suite, on suppose $0 \leq t_1 < t_2 \leq T$. Il suffit de montrer que :

1. $\forall y \in \mathcal{A}(t_2) \quad d(y, \mathcal{A}(t_1)) \leq \varepsilon,$
2. $\forall y \in \mathcal{A}(t_1) \quad d(y, \mathcal{A}(t_2)) \leq \varepsilon.$

Montrons juste le premier point (2. étant similaire). Soit $y \in \mathcal{A}(t_2)$. Il suffit de montrer que :

$$\exists z \in \mathcal{A}(t_1) \quad / \quad d(y, z) \leq \varepsilon.$$

Par définition de $\mathcal{A}(t_2)$, il existe un contrôle $u \in L^\infty([0, T], \Omega)$ tel que la trajectoire associée à u , partant de x_0 , vérifie : $x(t_2) = y$, voir figure 6.1. On va voir que $z = x(t_1)$

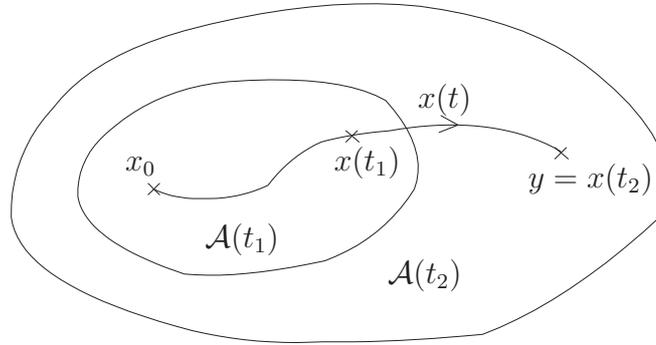


FIGURE 6.1 –

convient. En effet on a :

$$\begin{aligned}
 x(t_2) - x(t_1) &= M(t_2)x_0 + \int_0^{t_2} M(t_2)M(s)^{-1}(B(s)u(s) + r(s))ds \\
 &\quad - \lambda(M(t_1)x_0 + \int_0^{t_1} M(t_1)M(s)^{-1}(B(s)u(s) + r(s))ds) \\
 &= M(t_2) \int_{t_1}^{t_2} M(s)^{-1}(B(s)u(s) + r(s))ds \\
 &\quad + \lambda(M(t_2) - M(t_1)) \lambda(x_0 + \int_0^{t_1} M(s)^{-1}(B(s)u(s) + r(s))ds)
 \end{aligned}$$

Si $|t_1 - t_2|$ est petit, le premier terme de cette somme est petit par continuité de l'intégrale ; le deuxième terme est petit par continuité de $t \mapsto M(t)$. D'où le résultat. \square

Revenons maintenant à l'existence de trajectoires en temps minimal. Il faut tout d'abord formaliser, à l'aide de $\mathcal{A}(x_0, t)$, la notion de temps minimal. Considérons comme précédemment le système de contrôle dans \mathbb{R}^n :

$$\dot{x}(t) = A(t)x(t) + b(t)u(t) + r(t),$$

où les contrôles u sont à valeurs dans un compact d'intérieur non vide $\Omega \subset \mathbb{R}^m$. Soient $x_0, x_1 \in \mathbb{R}^n$. On suppose que x_1 est accessible depuis x_0 , i.e. il existe au moins une trajectoire reliant x_0 à x_1 . Parmi toutes les trajectoires reliant x_0 à x_1 on aimerait caractériser celles qui le font en temps minimal t^* , voir figure 6.2.

Si t^* est le temps minimal, alors pour tout $t < t^*$, $x_1 \notin \mathcal{A}(x_0, t)$ (en effet sinon x_1 serait accessible à partir de x_0 en un temps inférieur à t^*). Par conséquent :

$$t^* = \inf\{t > 0 / x_1 \in \mathcal{A}(x_0, t)\}.$$

Ce temps t^* est bien défini car, d'après le théorème 6.1, $\mathcal{A}(x_0, t)$ varie continûment avec t , donc $\{t > 0 / x_1 \in \mathcal{A}(x_0, t)\}$ est fermé dans \mathbb{R} . En particulier cette borne inférieure est un minimum.

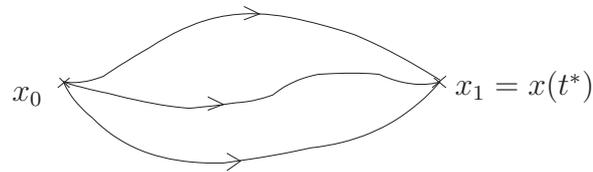


FIGURE 6.2 –

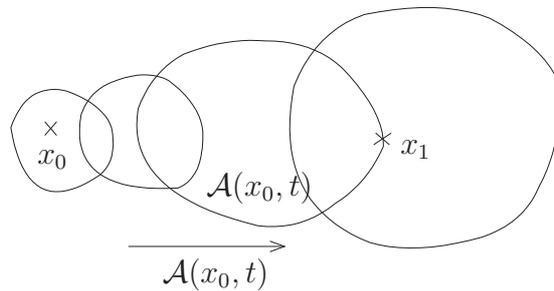


FIGURE 6.3 – Temps minimal

Le temps $t = t^*$ est le premier temps pour lequel $\mathcal{A}(x_0, t)$ contient x_1 (voir figure 6.3). D'autre part, on a nécessairement :

$$x_1 \in \partial\mathcal{A}(x_0, t^*) = \mathcal{A}(x_0, t^*) \setminus \overset{\circ}{\mathcal{A}}(x_0, t^*).$$

En effet, si x_1 appartenait à l'intérieur de $\mathcal{A}(x_0, t^*)$, alors pour $t < t^*$ proche de t^* , x_1 appartiendrait encore à $\mathcal{A}(x_0, t)$ car $\mathcal{A}(x_0, t)$ varie continûment avec t . Mais ceci contredit le fait que t^* est le temps minimal.

En particulier on a prouvé le théorème d'existence suivant.

Théorème 6.3. *Si le point x_1 est accessible depuis x_0 alors il existe une trajectoire temps-minimal reliant x_0 à x_1 .*

Remarque. On peut aussi se poser le problème d'atteindre une cible non réduite à un point. Ainsi, soit $(M_1(t))_{0 \leq t \leq T}$ une famille de sous-ensembles compacts de \mathbb{R}^n variant continûment en t . Tout comme précédemment, on voit que s'il existe un contrôle u à valeurs dans Ω joignant x_0 à $M_1(T)$, alors il existe un contrôle temps-minimal défini sur $[0, t^*]$ joignant x_0 à $M(t^*)$.

Ces remarques donnent une vision géométrique de la notion de temps minimal, et conduisent à la définition :

Définition 6.1. Le contrôle u est dit *extrémal* sur $[0, t]$ si la trajectoire du système $\dot{x}(t) = A(t)x(t) + b(t)u(t) + r(t)$ associée à u vérifie : $x(t) \in \partial\mathcal{A}(x_0, t)$.

En particulier, tout contrôle temps-minimal est extrémal. La réciproque est évidemment fautive car l'extrémalité ne fait pas la différence entre la minimalité et la maximalité.

Dans le paragraphe suivant on donne une caractérisation de cette propriété.

6.2 Condition nécessaire d'optimalité : principe du maximum dans le cas linéaire

Le théorème suivant donne une condition nécessaire et suffisante pour qu'un contrôle soit extrémal.

Théorème 6.4. *Considérons le système de contrôle linéaire :*

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

où le domaine de contraintes $\Omega \subset \mathbb{R}^m$ sur le contrôle est compact. Soit $T > 0$. Le contrôle u est extrémal sur $[0, T]$ si et seulement si il existe une solution non triviale $p(t)$ de l'équation $\dot{p}(t) = -p(t)A(t)$ telle que

$$p(t)B(t)u(t) = \max_{v \in \Omega} p(t)B(t)v \tag{6.3}$$

pour presque tout $t \in [0, T]$. Le vecteur ligne $p(t) \in \mathbb{R}^n$ est appelé *vecteur adjoint*.

Remarque. La condition initiale $p(0)$ dépend en fait du point final x_1 , comme on le voit dans la démonstration. Comme elle n'est pas directement connue, l'usage de ce théorème sera plutôt indirect, comme on le verra dans les exemples.

Remarque. Dans le cas mono-entrée (contrôle scalaire), et si de plus $\Omega = [-a, a]$ où $a > 0$, la condition de maximisation implique immédiatement que $u(t) = a \operatorname{sign}(p(t)B(t))$. La fonction $\varphi(t) = p(t)B(t)$ est appelée *fonction de commutation*, et un temps t_c auquel le contrôle extrémal $u(t)$ change de signe est appelé un *temps de commutation*. C'est en particulier un zéro de la fonction φ .

PREUVE.

▷ On a vu que $Acc_{\Omega}(x_0, T) = Acc_{\operatorname{Conv}(\Omega)}(x_0, T)$, et par conséquent on peut supposer que Ω est convexe. Si u est extrémal sur $[0, T]$, soit x la trajectoire associée à u . On a $x(T) \in \partial\mathcal{A}(x_0, T)$. Par convexité de $\mathcal{A}(x_0, T)$, il existe d'après le théorème du convexe un hyperplan séparant au sens large $x(T)$ et $\mathcal{A}(x_0, T)$. Soit p_T un vecteur normal à cet hyperplan, voir figure 6.4.

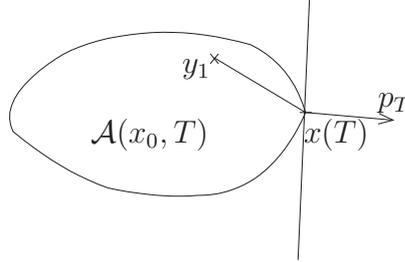


FIGURE 6.4 –

D'après le théorème du convexe :

$$\forall y_1 \in \mathcal{A}(x_0, T) \quad p_T(y_1 - x(T)) \leq 0. \quad (6.4)$$

Par définition de $\mathcal{A}(x_0, T)$, il existe un contrôle u_1 tel que la trajectoire associée $y(t)$ vérifie : $y_1 = y(T)$. L'inégalité (6.4) se réécrit :

$$p_T x(T) \geq p_T y(T).$$

D'où :

$$\int_0^T p_T M(T) M(s)^{-1} (B(s)u(s) + r(s)) ds \geq \int_0^T p_T M(T) M(s)^{-1} (B(s)u_1(s) + r(s)) ds.$$

Appelons $p(t)$ la solution sur $[0, T]$ de $\dot{p} = -pA$, telle que $p(T) = p_T$. Alors il est clair que : $p(t) = p(0)M(t)^{-1}$ et $p_T = p(T) = p(0)M(T)^{-1}$. D'où :

$$\forall s \in [0, T] \quad p_T M(T) M(s)^{-1} = p(0) M(s)^{-1} = p(s),$$

et donc :

$$\int_0^T p(s) B(s) u_1(s) ds \leq \int_0^T p(s) B(s) u(s) ds \quad (6.5)$$

Si (6.3) n'est pas vraie alors

$$p(t) B(t) u(t) < \max_{v \in \Omega} p(t) B(t) v.$$

sur un sous-ensemble de $[0, T]$ de mesure positive. Soit alors $u_1(\cdot)$ sur $[0, T]$ à valeurs dans Ω tel que :

$$p(t) B(t) u_1(t) = \max_{v \in \Omega} p(t) B(t) v.$$

En appliquant un lemme de sélection mesurable de théorie de la mesure, on peut montrer que l'application $u_1(\cdot)$ peut être choisie mesurable sur $[0, T]$.

Comme u_1 est à valeurs dans Ω , l'inégalité (6.5) est vraie, alors que par ailleurs la définition de u_1 conduit immédiatement à l'inégalité stricte inverse, d'où la contradiction. Par conséquent (6.3) est vraie.

Réciproquement, supposons qu'il existe un vecteur adjoint tel que le contrôle u satisfait (6.3). Notons $x(\cdot)$ la trajectoire associée à u . On voit facilement en remontant le raisonnement précédent que :

$$\forall y_1 \in \mathcal{A}(x_0, T) \quad p(T)(y_1 - x(T)) \leq 0. \quad (6.6)$$

Raisonnons alors par l'absurde, et supposons que $x(T) \in \text{Int}\mathcal{A}(x_0, T)$. Alors il existe un point y_1 de $\mathcal{A}(x_0, T)$ qui est sur la demi-droite d'origine $x(T)$ et de direction $p(T)$, voir figure 6.5. Mais alors : $p(T)(y_1 - x(T)) > 0$, ce qui contredit (6.6). Donc $x(T) \in \partial\mathcal{A}(x_0, T)$, et donc u

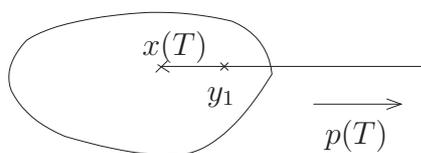


FIGURE 6.5 –

est extrémal.

□

Remarque. Si u est extrémal sur $[0, T]$ alors u est aussi extrémal sur $[0, t]$ pour tout $t \in [0, T]$, et de plus $p(t)$ est un vecteur normal extérieur à $\mathcal{A}(x_0, t)$. Cela découle facilement de la preuve et de la propriété (6.3).

Remarque. Puisque tout contrôle temps-minimal est extrémal, le théorème précédent, qui est le principe du maximum dans le cas linéaire, donne une *condition nécessaire* d'optimalité.

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

Comme exemple théorique d'application, montrons le résultat suivant.

Proposition 6.5. *Considérons dans \mathbb{R}^n le système linéaire autonome $\dot{x}(t) = Ax(t) + Bu(t)$, où $B \in \mathbb{R}^n$ et $|u(t)| \leq 1$. Supposons que la paire (A, B) satisfait la condition de Kalman.*

1. *Si toute valeur propre de A est réelle, alors tout contrôle extrémal a au plus $n - 1$ commutations sur \mathbb{R}^+ .*
2. *Si toute valeur propre de A a une partie imaginaire non nulle, alors tout contrôle extrémal a un nombre infini de commutations sur \mathbb{R}^+ .*

PREUVE.

▷ Comme le système est commandable, le système peut s'écrire sous forme équivalente de Brunovski, et il est alors équivalent à une équation différentielle scalaire d'ordre n de la forme :

$$x^{(n)} + a_1 x^{(n-1)} + \dots + a_n x = u, \quad |u| \leq 1.$$

De plus, tout contrôle extrémal est de la forme $u(t) = \text{signe } \lambda(t)$, où $\lambda(t)$ est la dernière coordonnée du vecteur adjoint, qui vérifie l'équation différentielle :

$$\lambda^{(n)} - a_1 \lambda^{(n-1)} + \dots + (-1)^n a_n \lambda = 0.$$

En effet le vecteur adjoint vérifie $p'(t) = -p(t)A(t)$.

1. Si toute valeur propre de A est réelle, alors $\lambda(t)$ s'écrit sous la forme :

$$\lambda(t) = \sum_{j=1}^r P_j(t) e^{\lambda_j t},$$

où P_j est un polynôme de degré inférieur ou égal à $n_j - 1$, et où $\lambda_1, \dots, \lambda_r$, sont les r valeurs propres distinctes de $-A$, de multiplicités respectives n_1, \dots, n_r . Notons que $n = n_1 + \dots + n_r$. On montre alors facilement, par récurrence, que $\lambda(t)$ admet au plus $n - 1$ zéros.

2. Si toute valeur propre de A a une partie imaginaire non nulle, alors, comme précédemment, on peut écrire :

$$\lambda(t) = \sum_{j=1}^r (P_j(t) \cos \beta_j t + Q_j(t) \sin \beta_j t) e^{\alpha_j t},$$

où $\lambda_j = \alpha_j + i\beta_j$, et P_j, Q_j sont des polynômes réels non nuls. En mettant en facteur un terme $t^k e^{\alpha_j t}$ de plus haut degré (i.e. dominant), on voit facilement que $\lambda(t)$ a un nombre infini de zéros. □

6.3 Exemple : Synthèse optimale pour le problème de l'oscillateur harmonique

Appliquons la théorie précédente à l'exemple de l'oscillateur harmonique présenté en introduction, pour $k_2 = 0$, et répondons aux deux questions suivantes :

1. Pour toute condition initiale $x(0) = x_0, \dot{x}(0) = y_0$, existe-t-il une force extérieure horizontale (un contrôle) qui permette d'amener la masse ponctuelle à sa position d'équilibre $x(T) = 0, \dot{x}(T) = 0$ en un temps fini T ?
2. Si la première condition est satisfaite, peut-on de plus déterminer cette force de manière à minimiser le temps ?

Enfin, ces deux problèmes résolus, nous représenterons dans le plan de phase la trajectoire optimale obtenue.

6.3.1 Contrôlabilité du système

Le système s'écrit :

$$\begin{cases} \dot{X} &= AX + Bu \\ X(0) &= X_0 \end{cases} \quad \text{avec } A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

On a facilement $\text{rg}(B, AB) = 2$, et par ailleurs les valeurs propres de A sont de partie réelle nulle, donc d'après le critère de Kalman, le système est contrôlable à 0, i.e. il existe des contrôles tels que les trajectoires associées relient X_0 à 0, ce qui répond à la première question.

6.3.2 Interprétation physique

- Si on n'applique aucune force extérieure : $u = 0$. L'équation du mouvement est : $\ddot{x} + x = 0$. La masse ponctuelle oscille, et ne s'arrête jamais, donc ne parvient pas à sa position d'équilibre en un temps fini.
- Si on applique certaines forces extérieures, on a tendance à amortir les oscillations. La théorie prévoit qu'on parvient à stopper l'objet en un temps fini.

6.3.3 Calcul du contrôle optimal

D'après le paragraphe précédent, il existe des contrôles permettant de relier X_0 à 0. On cherche maintenant à le faire en temps minimal. Pour cela, on applique le théorème 6.4 :

$$u(t) = \text{signe } p(t)B,$$

où $p(t) \in \mathbb{R}^2$ est solution de : $\dot{p} = -pA$. Posons $p = (p_1, p_2)$. Alors : $u(t) = \text{signe } p_2(t)$, et : $\dot{p}_1 = p_2, \dot{p}_2 = -p_1$, d'où $\ddot{p}_2 + p_2 = 0$. Donc $p_2(t) = \lambda \cos t + \mu \sin t$. En particulier, la durée entre deux zéros consécutifs de $p_2(t)$ est exactement π . Par conséquent le contrôle optimal est constant par morceaux sur des intervalles de longueur π , et prend alternativement les valeurs ± 1 .

- Si $u = -1$, on obtient le système différentiel :

$$\begin{cases} \dot{x} = y \\ \dot{y} = -x - 1 \end{cases} \quad (6.7)$$

- Si $u = +1$:

$$\begin{cases} \dot{x} = y \\ \dot{y} = -x + 1 \end{cases} \quad (6.8)$$

La trajectoire optimale finale, reliant X_0 à 0, sera constituée d'arcs successifs, solutions de (6.7) et (6.8).

Solutions de (6.7). On obtient facilement $(x+1)^2 + y^2 = \text{cste} = R^2$, donc les courbes solutions de (6.7) sont des cercles centrés en $x = -1, y = 0$, et de période 2π (en fait : $x(t) = -1 + R \cos t, y(t) = R \sin t$).

Solutions de (6.8). On obtient : $x(t) = 1 + R \cos t, y(t) = R \sin t$. Les solutions sont des cercles centrés en $x = 1, y = 0$, de période 2π .

La trajectoire optimale de X_0 à 0 doit donc suivre alternativement un arc de cercle centré en $x = -1, y = 0$, et un arc de cercle centré en $x = 1, y = 0$.

Quitte à changer t en $-t$, nous allons raisonner à l'envers, et construire la trajectoire optimale menant de 0 à X_0 . Pour cela, nous allons considérer toutes les trajectoires optimales partant de 0, et nous sélectionnerons celle qui passe par X_0 .

C'est en faisant varier $p(0)$ que l'on change de trajectoire optimale : en effet, $p(0)$ détermine $p(t)$ pour tout t d'après le théorème de Cauchy-Lipschitz, ce qui détermine un contrôle optimal $u(t)$, et donc une trajectoire optimale.

Prenons des exemples pour commencer à représenter l'allure des trajectoires optimales possibles :

- si $p_1(0) = 1, p_2(0) = 0$: alors $p_2(t) = -\sin t$, donc sur $]0, \pi[$ on a $u(t) = \text{sign} p_2(t) = -1$. La trajectoire optimale correspondante, partant de 0, suit donc pendant un temps π l'arc de cercle Γ_- solution de (6.7), passant par 0, voir figure 6.6.

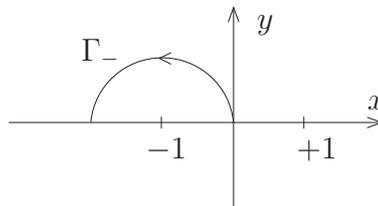


FIGURE 6.6 –

- si $p_1(0) = -1, p_2(0) = 0$: alors $p_2(t) = \sin t$, donc sur $]0, \pi[$ on a $u(t) = \text{sgn} p_2(t) = +1$. La trajectoire optimale correspondante, partant de 0, suit donc pendant un temps π l'arc de cercle Γ_+ solution de (6.8), passant par 0, voir figure 6.7.

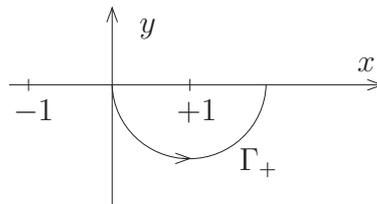


FIGURE 6.7 –

- Pour tout autre choix de $p(0)$ tel que $p_2(0) > 0$, la trajectoire optimale correspondante part de l'origine en suivant Γ_+ jusqu'à ce que $p_2(t) = 0$. Au-delà de ce point, $p_2(t)$ change de signe, donc le contrôle commute et prend la valeur -1 , pendant une durée π (i.e. jusqu'à ce que $p_2(t)$ change à nouveau de signe). La trajectoire optimale doit alors être solution de (6.7), en partant de ce point de commutation M , et doit donc suivre un arc de cercle centré en $x = -1, y = 0$, pendant un temps π (c'est donc un demi-cercle, vu la paramétrisation des courbes de (6.7)), voir figure 6.8.

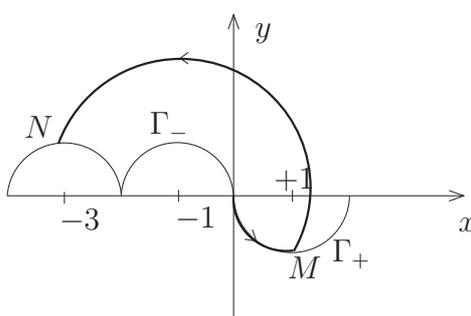


FIGURE 6.8 –

La trajectoire optimale rencontre un deuxième point de commutation N lorsque à nouveau $p_2(t)$ change de signe. On remarque que M et N sont symétriques par rapport au point $x = -1, y = 0$ (en effet ce sont les extrémités d'un demi-cercle centré en ce point). Le point M appartenant au demi-cercle Γ_+ , le point N appartient au demi-cercle image de Γ_+ par la symétrie par rapport au point $x = -1, y = 0$ qui est aussi, comme on le voit facilement, le translaté à gauche de Γ_- par la translation de vecteur $(-2, 0)$.

Poursuivons alors notre raisonnement : on se rend compte que les points de commutation de cette trajectoire optimale partant de 0 sont situés sur la courbe W construite de la manière suivante : W est l'union de tous les translatés à gauche de Γ_- par la translation précédente, et aussi de tous les translatés à droite de Γ_+ , voir figure 6.9.

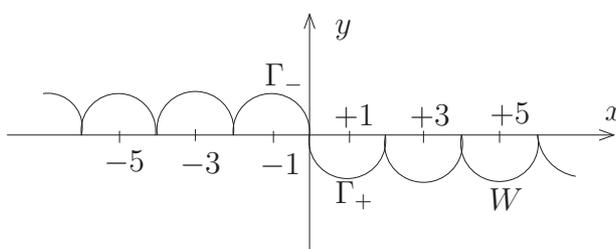


FIGURE 6.9 – Ensemble W

Les trajectoires optimales sont alors construites de la manière suivante : on part de 0 et on suit un morceau de Γ_+ ou Γ_- , jusqu'à un premier point de commutation. Si par exemple on était sur Γ_+ , alors partant de ce point on suit un arc de cercle centré en $x = -1, y = 0$, au-dessus de W , jusqu'à ce qu'on rencontre W . De ce deuxième point de commutation, on suit un arc de cercle centré en $x = +1, y = 0$ jusqu'à rencontrer W en un troisième point de commutation, etc, voir figure 6.10.

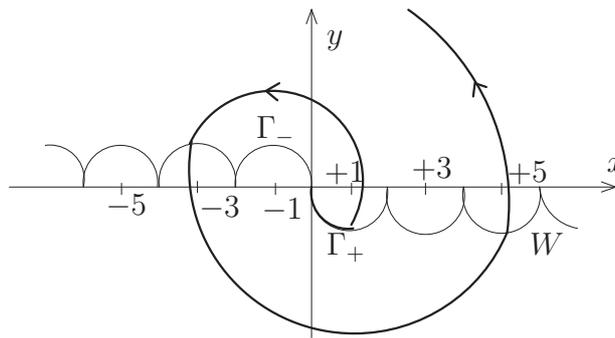


FIGURE 6.10 –

On est maintenant en mesure de répondre à la deuxième question, du moins graphiquement : le but est de relier 0 et X_0 par une trajectoire optimale. La théorie prévoit qu'on peut effectivement le faire. Une trajectoire partant de 0 est, comme on vient de le voir ci-dessus, déterminée par deux choix :

1. partant de 0, on peut suivre un morceau de Γ_+ ou de Γ_- .
2. il faut choisir le premier point de commutation.

Si maintenant on se donne un point $X_0 = (x_0, y_0)$ du plan de phase, on peut déterminer graphiquement ces deux choix, et obtenir un tracé de la trajectoire optimale, voir figure 6.11. Dans la pratique il suffit d'inverser le temps, i.e. de partir du point final et d'atteindre le point initial.

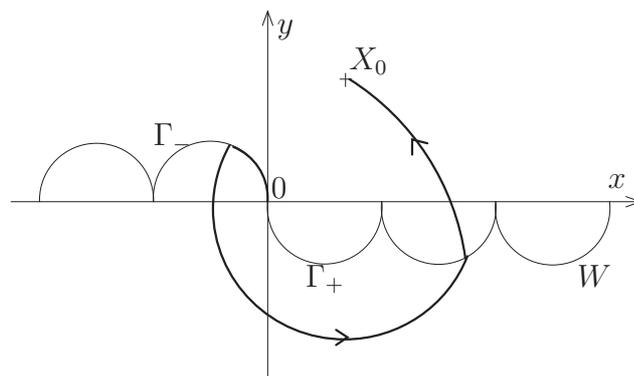


FIGURE 6.11 – Synthèse optimale

Chapitre 7

Contrôle optimal non-linéaire

Dans le chapitre suivant, nous allons énoncer une forme générale du Principe du Maximum de Pontryagin (en bref PMP) dans le cas non linéaire. Ce théorème constitue l'instrument fondamental pour calculer les trajectoires optimales. Nous l'appliquerons ensuite à deux classes de problèmes : le problème sous-riemaniennien et le problème de temps minimal pour des systèmes commandés affines avec contrôles bornés.

7.1 Énoncé général du Principe du maximum de Pontryagin

On considère le système commandé défini sur \mathbb{R}^n comme suit :

$$\dot{x}(t) = f(x(t), u(t)), \quad (7.1)$$

où $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ est de classe C^∞ et où les contrôles sont des applications mesurables, bornées, définies sur un intervalle de \mathbb{R}^+ et à valeurs dans $U \subset \mathbb{R}^m$. On note \mathcal{U} l'ensemble des contrôles admissibles u dont les trajectoires associées relient un point initial de x_0 à un point final de x_1 .

Par ailleurs on définit $C(t, u)$, le coût d'un contrôle u sur $[0, t]$, par :

$$C(t, u) = \int_0^t f^0(x(s), u(s)) ds.$$

où $f^0 : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ est C^∞ , et $x(\cdot)$ est la trajectoire solution de (7.1) associée au contrôle u .

On considère le problème de contrôle optimal suivant : déterminer une trajectoire reliant x_0 à x_1 et minimisant le coût C . Le temps final peut être fixé ou non. Si une telle trajectoire existe, on dira qu'elle est optimale et que le contrôle associé l'est aussi.

Théorème 7.1 (Principe du maximum de Pontryagin (PMP)). *Si le contrôle $u \in \mathcal{U}$ associé à la trajectoire $x(\cdot)$ est optimal sur $[0, T]$, alors il existe une application $p(\cdot)$:*

$[0, T] \rightarrow \mathbb{R}^n$ lipschitzienne appelée vecteur adjoint (ou covecteur), et un réel $p^0 \leq 0$, tels que le couple $(p(\cdot), p^0)$ est non trivial, et tels que les équations suivantes sont vérifiées pour presque tout $t \in [0, T]$:

$$\dot{x}(t) = \frac{\partial H}{\partial p}(x(t), p(t), p^0, u(t)), \quad (7.2)$$

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(x(t), p(t), p^0, u(t)), \quad (7.3)$$

où $H(x, p, p^0, u) = p \cdot f(t, x, u) + p^0 f^0(t, x, u)$ est l'Hamiltonien du système.

De plus, on a la condition de maximisation suivante, vraie presque partout sur $[0, T]$:

$$H(x(t), p(t), p^0, u(t)) = \max_{v \in U} H(x(t), p(t), p^0, v). \quad (7.4)$$

Enfin,

$$H(x(t), p(t), p^0, u(t)) = \text{const}, \quad (7.5)$$

et cette constante est nulle si le temps final pour joindre la cible x_1 n'est pas fixé.

Il faut remarquer que le covecteur p est un vecteur ligne.

Remarque. Dans le cas où $U = \mathbb{R}^m$, i.e. lorsqu'il n'y a pas de contrainte sur le contrôle, la condition de maximum (7.4) peut être remplacée par la condition plus faible $\frac{\partial H}{\partial u} = 0$. On parle alors de principe du maximum faible. Voir le paragraphe 7.4. Par ailleurs, le lecteur attentif de ce cours aura noté que, dans le cas de la commande LQ, les deux versions du PMP sont équivalentes.

Remarque. Il est important de remarquer que le PMP n'est qu'une condition *nécessaire* d'optimalité. Cela signifie pour obtenir en pratique les trajectoires optimales on doit d'abord montrer l'existence de ces trajectoires puis appliquer le PMP.

Définition 7.1. Une trajectoire *extrême* du problème de contrôle optimal est une solution des équations (7.2) et (7.4). Si $p_0 = 0$, on dit que l'extrême est *anormale*, et si $p^0 \neq 0$ l'extrême est dite *normale*.

Remarque. Dans le cas général et contrairement au cas linéaire, l'équation différentielle pour le covecteur dépend de la trajectoire. Appliquer le le PMP s'avère donc plus compliqué.

7.2 Le problème sous-riemannien

Une classe importante de problèmes de contrôle optimal est celle des problèmes sous-riemanniens, c'est-à-dire que la dynamique et le coût sont respectivement linéaire et quadratique par rapport à la commande. De plus, on suppose que le coût ne dépend que de la commande,

$$\dot{x} = \sum_{i=1}^m u_i F_i(x), \quad \int_0^T \sum u_i(t)^2 dt, \quad x(0) = x_0, \quad , x(T) = x_1. \quad (7.6)$$

Ici, $x \in \mathbb{R}^n$, F_i ($i = 1, \dots, m$) sont des champs de vecteurs réguliers, $m \geq 2$ et le temps final T est fixé.

On a alors,

$$H(x, p, p^0, u) = \sum_{i=1}^m u_i p \cdot F_i(x) + p^0 \sum u_i^2.$$

Dans de nombreuses applications intéressantes, il est possible de prouver qu'il n'y a pas d'extrémales anormales et dans la suite de ce paragraphe, nous le supposons. On peut alors normaliser l'Hamiltonien en choisissant $p^0 = 1/2$. D'après la condition de maximalité, on a

$$\frac{\partial H}{\partial u}(x(t), p(t), 1/2, u(t)) = 0,$$

et on en déduit $u_i(t) = p(t) \cdot F_i(x(t))$. En définissant $\mathcal{H}(x, p) := H(x, p, 1/2, p \cdot F_i(x)) = \frac{1}{2} \sum_{i=1}^m (p \cdot F_i(x))^2$, on obtient les équations hamiltoniennes suivantes,

$$\dot{x}(t) = \frac{\partial H}{\partial p}(x(t), p(t), p^0, u(t)) = \frac{\partial \mathcal{H}}{\partial p}(x(t), p(t)) \quad (7.7)$$

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(x(t), p(t), p^0, u(t)) = -\frac{\partial \mathcal{H}}{\partial x}(x(t), p(t)). \quad (7.8)$$

Une conséquence très importante des équations précédentes est que les trajectoires extrémales (et donc les trajectoires optimales lorsqu'elles existent) sont de classe C^∞ .

Remarquons aussi que $\mathcal{H}(x(t), p(t)) = \sum_{i=1}^m (u_i(t))^2$ est une constante du mouvement.

Remarque. Soit $(x(\cdot), p(\cdot))$ une solution de (7.7), (7.8), correspondant aux contrôles $(u_i(\cdot))_{1 \leq i \leq m}$, telle que $x(0) = x_0$, $x(T) = x_1$ et $\mathcal{H}(x(t), p(t)) = \beta > 0$. Alors, pour tout $\alpha > 0$, la paire $(\bar{x}(t), \bar{p}(t)) := (x(\alpha t), \alpha p(\alpha t))$ est encore solution de (7.7), (7.8), correspondant à $(\alpha u_i(\cdot))_{1 \leq i \leq m}$ avec $x(0) = x_0$, $x(T/\alpha) = x_1$ et $\mathcal{H}(\bar{x}(t), \bar{p}(t)) = \alpha^2 \beta > 0$. En d'autres termes, la ligne de niveau de l'Hamiltonien est déterminée par le temps final T pour lequel la cible est atteinte.

7.3 Temps minimum pour un système affine bidimensionnel

Dans ce paragraphe, nous allons présenter la théorie du temps minimum pour les systèmes commandés affines en la commande et définis sur \mathbb{R}^2 :

$$\dot{x} = F(x) + uG(x), \quad x \in \mathbb{R}^2, \quad |u| \leq 1. \quad (7.9)$$

Il s'agit d'atteindre chaque point du plan en temps minimum en partant de l'origine.

Dans ce cas, l'Hamiltonien prend une forme particulièrement simple $H(x, p, p^0, u) = p \cdot F(x) + u p \cdot G(x) + p^0$. En appliquant le PMP, on obtient donc :

(i) $\dot{p}(t) = -p(t) \cdot (\nabla F + u(t) \nabla G)(x(t))$,

(ii) $p(t) \cdot F(x(t)) + u(t) p(t) \cdot G(x(t)) + p^0 = 0$,

(iii) $u(t) p(t) \cdot G(x(t)) = \max_{v \in [-1,1]} v p(t) \cdot G(x(t))$.

D'une manière analogue au cas des systèmes linéaires, nous allons définir la fonction de commutation.

Définition 7.2. (fonction de commutation) Soit $(x(\cdot), p(\cdot)) : [0, \tau] \rightarrow \mathbb{R}^2 \times \mathbb{R}^2$ une trajectoire extrême avec le covecteur correspondant. La fonction de commutation associée est définie par $\phi(\cdot) := p(\cdot) \cdot G(x(\cdot))$. Notons que $\phi(\cdot)$ est absolument continue.

Le lemme suivant est une conséquence facile des équations du PMP.

Lemme 7.2. Soit $(x(\cdot), p(\cdot)) : [0, \tau] \rightarrow \mathbb{R}^2 \times \mathbb{R}^2$ une trajectoire extrême avec le covecteur correspondant et $\phi(\cdot)$ la fonction de commutation associée. Si ϕ n'a que des zéros isolés, alors $u(t) = \text{sgn}(\phi(t))$ dans $[0, \tau]$.

Définition 7.3. (Trajectoire bang-bang et trajectoire singulière) Une trajectoire qui satisfait les hypothèses du lemme précédent est appelée une trajectoire *bang-bang*.

À l'opposé, on dira qu'une trajectoire extrême est *singulière* sur $[t_1, t_2]$ si la fonction de commutation associée est identiquement nulle sur $[t_1, t_2]$.

Comme ϕ est absolument continue, on peut toujours la dériver. On a alors :

Lemme 7.3. Soit $(x(\cdot), p(\cdot)) : [0, \tau] \rightarrow \mathbb{R}^2 \times \mathbb{R}^2$ une trajectoire extrême avec son covecteur correspondant et $\phi(\cdot)$ la fonction de commutation associée. Alors $\phi(\cdot)$ est de classe \mathcal{C}^1 et

$$\dot{\phi}(t) = p(t) \cdot [F, G](x(t)).$$

PREUVE.

▷ En utilisant le PMP, on a pour presque tout t :

$$\begin{aligned} \dot{\phi}(t) &= \frac{d}{dt}(p(t) \cdot G(x(t))) = \dot{p}(t) \cdot G(x(t)) + p \cdot \dot{G}(x(t)) \\ &= -p(t)(\nabla F + u(t)\nabla G)(x(t)) \cdot G(x(t)) + p \cdot \nabla G(x(t))(F + u(t)G)(x(t)) \\ &= p(t) \cdot [F, G](x(t)). \end{aligned} \tag{7.10}$$

□

L'étape suivante consiste à déterminer quand le contrôle change de signe ou peut prendre des valeurs en $] - 1, +1[$.

7.3.1 Trajectoires singulières et détermination des commutations

Afin de déterminer les trajectoires singulières ainsi que les commutations, on définit $\Delta_A(\cdot)$ et $\Delta_B(\cdot)$, les fonctions suivantes :

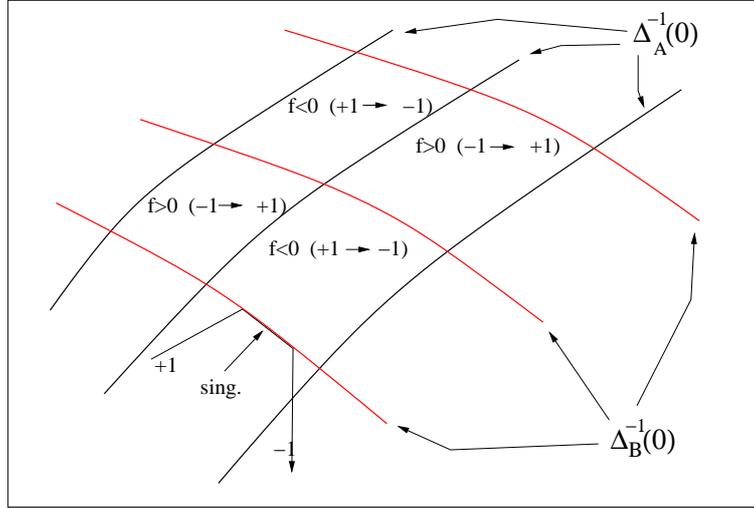


FIGURE : Possible commutations dans les composantes connexes de $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$ en relation avec le signe de f_S , et un exemple de trajectoire extrémale contenant un arc singulier.

$$\Delta_A(x) := \det(F(x), G(x)) = F_1(x)G_2(x) - F_2(x)G_1(x), \quad (7.11)$$

$$\Delta_B(x) := \det(G(x), [F, G](x)) = G_1(x)[F, G]_2(x) - G_2(x)[F, G]_1(x). \quad (7.12)$$

L'ensemble des zéros de ces fonctions $\Delta_A^{-1}(0), \Delta_B^{-1}(0)$, sont respectivement l'ensemble des points où F et G sont parallèles et l'ensemble des points où G est parallèle à $[F, G]$. Comme on va le voir, ces lieux sont cruciaux pour ce qui est de la construction des trajectoires optimales. En effet, en supposant que ces lieux sont inclus dans des courbes lisses, on montre que

- dans chaque composante connexe de $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$, toute trajectoire extrémale est bang-bang avec au plus une commutation. De plus, le contrôle commute de -1 to $+1$ si $f_S := -\Delta_B \setminus \Delta_A > 0$ et de $+1$ to -1 si $f_S < 0$;
- le support des trajectoires singulières est contenu dans l'ensemble $\Delta_B^{-1}(0)$. (Remarquons que l'on ne sait rien encore sur les valeurs prises par le contrôle le long d'une trajectoire singulière.)

Afin de se convaincre des résultats précédents, on commence par étudier ce qui se passe sur $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$. Pour cela, on considère la définition suivante.

Définition 7.4. Les points de $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$ sont appelés *points ordinaires*. En un tel point, $F(x)$ et $G(x)$ sont linéairement indépendants et on définit les fonctions scalaires f_S et g_S comme les coefficients de la combinaison linéaire $[F, G](x) = f_S(x)F(x) + g_S(x)G(x)$.

On a alors :

Lemme 7.4. *En un point ordinaire x ,*

$$f_S(x) = -\frac{\Delta_B(x)}{\Delta_A(x)}. \quad (7.13)$$

PREUVE.

▷ Par un simple calcul,

$$\begin{aligned} \Delta_B(x) &= \det(G(x), [F, G](x)) = \det(G(x), f_S(x)F(x) + g_S(x)G(x)) \\ &= f_S(x) \det(G(x), F(x)) = -f_S(x) \Delta_A(x). \end{aligned}$$

□

Sur $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$, la structure des trajectoires optimales est simple :

Théorème 7.5. *Soit Ω un ouvert de $\mathbb{R}^2 \setminus (\Delta_A^{-1}(0) \cup \Delta_B^{-1}(0))$. Alors, toute trajectoire extrême $x(\cdot)$ du système commandé (7.9) contenue dans Ω est bang-bang avec au plus une commutation. De plus, si $f_S > 0$ (resp. $f_S < 0$) sur Ω , alors la commande associée à $x(\cdot)$ est ou bien constante égale à 1, ou -1 , ou bien commute de -1 à 1 (resp. de 1 à -1).*

PREUVE.

▷ On ne traitera que le cas où $f_S > 0$ sur Ω , l'autre cas étant similaire. Soit $(x(\cdot), p(\cdot))$ une trajectoire extrême avec le covecteur correspondant telle que $x(\cdot)$ est contenue dans Ω . Soit un temps \bar{t} tel que $p(\bar{t}) \cdot G(x(\bar{t})) = 0$. Alors, on a

$$\dot{\phi}(\bar{t}) = p(\bar{t}) \cdot [F, G](x(\bar{t})) = p(\bar{t}) \cdot (fF + gG)(x(\bar{t})) = f(x(\bar{t})) p(\bar{t}) \cdot F(x(\bar{t})).$$

D'après le PMP, on obtient que $p(\bar{t}) \cdot F(x(\bar{t})) \geq 0$. Cela implique que $\dot{\phi} > 0$, puisque $F(x(\bar{t}))$ et $G(x(\bar{t}))$ sont linéairement indépendants. Cela prouve que ϕ a au plus un seul zéro et on conclue.

□

Enfin, le résultat annoncé sur les trajectoires singulières est une conséquence du lemme suivant.

Lemme 7.6. *Soit $(x(\cdot), p(\cdot))$ une trajectoire extrême avec le covecteur correspondant et $\phi(\cdot)$ la fonction de commutation associée. Supposons que $\phi(t) \equiv 0$ sur $[a, b]$. Alors, $x([a, b]) \subset \Delta_B^{-1}(0)$.*

PREUVE.

▷ Comme $\phi(t) = p(t) \cdot G(x(t)) \equiv 0$ et $\dot{\phi}(t) = p(t) \cdot [F, G](x(t)) \equiv 0$, il résulte que G est parallèle à $[F, G]$ le long de $x(\cdot)$ sur $[a, b]$. On conclue.

□

7.4 Principe du maximum de Pontryagin faible

Dans cette section, nous prouvons le Principe du maximum de Pontryagin faible, c'est-à-dire :

Théorème 7.7 (Principe du maximum de Pontryagin faible). *On considère le système de contrôle optimal :*

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \quad x(T) = x_1, \quad T \text{ fixé} \quad (7.14)$$

$$\int_0^T f^0(x(s), u(s)) ds \rightarrow \min \quad (7.15)$$

où $f, f_0 : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ sont de classe C^∞ et où les contrôles sont des applications mesurables et bornés définies sur un intervalle de \mathbb{R}^+ et à valeurs dans \mathbb{R}^m .

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

$$\dot{x}(t) = \frac{\partial H}{\partial p}(x(t), p(t), p^0, u(t)), \quad (7.16)$$

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(x(t), p(t), p^0, u(t)), \quad (7.17)$$

$$\frac{\partial H}{\partial u}(x(t), p(t), p^0, u(t)) = 0. \quad (7.18)$$

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

Dans ce but, il faut introduire le concept d'application entrée-sortie et de contrôles singuliers.

Considérons de nouveau le système de contrôle général :

$$\dot{x}(t) = f(x(t), u(t)), \quad (7.19)$$

où f est une application de classe C^∞ de \mathbb{R}^{n+m} dans \mathbb{R}^n , et où les contrôles sont des applications mesurables et bornés définies sur un intervalle de \mathbb{R}^+ et à valeurs dans \mathbb{R}^m .

Définition 7.5. Soit $T > 0$. L'application entrée-sortie en temps T du système contrôlé (7.19) initialisé à 0 est l'application :

$$E_T : \begin{array}{l} \mathcal{U} \longrightarrow \mathbb{R}^n \\ u \longmapsto x(T) \end{array}$$

où \mathcal{U} est l'ensemble des contrôles admissibles, i.e. l'ensemble de contrôles u tels que la trajectoire associée est bien définie sur $[0, T]$.

Autrement dit, l'application entrée-sortie en temps T associée à un contrôle u le point final de la trajectoire associée à u . Une question importante en théorie du contrôle est d'étudier cette application en décrivant son image, ses singularités, etc.

7.4.1 Régularité de l'application entrée-sortie

En toute généralité on a le résultat suivant :

Proposition 7.8. *Considérons le système (7.19) où f est C^p , $p \geq 1$, et soit $\mathcal{U} \subset L^\infty([0, T], \mathbb{R}^m)$ le domaine de définition de E_T , c'est-à-dire l'ensemble des contrôles dont la trajectoire associée est bien définie sur $[0, T]$. Alors \mathcal{U} est un ouvert de $L^\infty([0, T], \mathbb{R}^m)$, et E_T est C^p au sens L^∞ .*

De plus la différentielle (au sens de Fréchet) de E_T en un point $u \in \mathcal{U}$ est donnée par le système linéarisé en u de la manière suivante. Posons, pour tout $t \in [0, T]$:

$$A(t) = \frac{\partial f}{\partial x}(x_u(t), u(t)) \quad , \quad B(t) = \frac{\partial f}{\partial u}(x_u(t), u(t)).$$

Le système de contrôle linéaire :

$$\begin{aligned} \dot{y}_v(t) &= A(t)y_v(t) + B(t)v(t) \\ y_v(0) &= 0 \end{aligned}$$

est appelé système linéarisé le long de la trajectoire $x(\cdot)$. La différentielle de Fréchet de E_T en u est alors l'application $dE_T(u)$ telle que, pour tout $v \in L^\infty([0, T], \mathbb{R}^m)$:

$$dE_T(u).v = y_v(T) = M(T) \int_0^T M^{-1}(s)B(s)v(s)ds \quad (7.20)$$

où M est la résolvante du système linéarisé, i.e. la solution matricielle de : $\dot{M} = AM$, $M(0) = \text{Id}$.

Démonstration. Pour la démonstration du fait que \mathcal{U} est ouvert, voir par exemple le livre de Sontag. Par hypothèse $u(\cdot)$ et sa trajectoire associée $x(\cdot, x_0, u)$ sont définis sur $[0, T]$. L'ensemble des contrôles étant les applications mesurables et bornées muni de la norme L^∞ , l'application E_T est de classe C^p sur un voisinage de $u(\cdot)$ en vertu des théorèmes de dépendance par rapport à un paramètre. Exprimons sa différentielle au sens de Fréchet. On note $x(\cdot) + \delta x(\cdot)$ la trajectoire associée à $u(\cdot) + \delta u(\cdot)$, issue en $t = 0$ de x_0 . Par un développement de Taylor, on obtient :

$$\begin{aligned} \frac{d}{dt}(x + \delta x)(t) &= f(x(t) + \delta x(t), u(t) + \delta u(t)) \\ &= f(x(t), u(t)) + \frac{\partial f}{\partial x}(x(t), u(t))\delta x(t) + \frac{\partial f}{\partial u}(x(t), u(t))\delta u(t) \\ &\quad + o(\delta x(t), \delta u(t)). \end{aligned}$$

Comme $\dot{x}(t) = f(x(t), u(t))$, on a donc :

$$\frac{d}{dt}(\delta x)(t) = \frac{\partial f}{\partial x}(x(t), u(t))\delta x(t) + \frac{\partial f}{\partial u}(x(t), u(t))\delta u(t) + o(\delta x(t), \delta u(t)).$$

Par ailleurs, on a aussi $x(0) + \delta x(0) = x_0 = x(0)$, donc $\delta x(0) = 0$.

En ne retenant que les termes du premier ordre (c-à-d en négligeant le "o"), on obtient le résultat. \square

Définition 7.6. Soit $u(\cdot)$ un contrôle défini sur $[0, T]$ tel que sa trajectoire associée $x(\cdot)$ issue de $x(0) = x_0$ est définie sur $[0, T]$. On dit que le contrôle $u(\cdot)$ (ou la trajectoire $x(\cdot)$) est singulier sur $[0, T]$ si la différentielle de Fréchet $dE_T(u)$ de l'application entrée-sortie au point u n'est pas surjective. Sinon on dit qu'il est régulier.

Proposition 7.9. Soient x_0 et T fixés. Si u est un contrôle régulier, alors E_T est ouverte dans un voisinage de u .

PREUVE.

▷ Par hypothèse, il existe n contrôles v_i tels que $dE_T(u).v_i = e_i$ où (e_1, \dots, e_n) est la base canonique de \mathbb{R}^n . On considère l'application :

$$(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \longmapsto E_T(u + \sum_{i=1}^n \lambda_i v_i).$$

Par construction, c'est un difféomorphisme local, et le résultat s'ensuit. □

Autrement dit, en un point x_1 atteignable en temps T depuis x_0 par une trajectoire régulière $x(\cdot)$, l'ensemble accessible $\mathcal{A}(T, x_0)$ est *localement ouvert*, i.e. est un voisinage du point x_1 . En particulier cela implique que le système est *localement contrôlable* autour du point x_1 . On parle aussi de *contrôlabilité le long de la trajectoire* $x(\cdot)$. Ainsi :

Proposition 7.10. Si u est un contrôle régulier sur $[0, T]$, alors le système est localement contrôlable le long de la trajectoire associée à ce contrôle.

Le corollaire suivant est immédiat.

Corollaire 7.11. Soit $u(\cdot)$ un contrôle défini sur $[0, T]$ tel que sa trajectoire associée $x(\cdot)$ issue de $x(0) = x_0$ est définie sur $[0, T]$ et vérifie au temps T :

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

Alors le contrôle u est singulier sur $[0, T]$.

7.4.2 Caractérisation hamiltonienne des contrôles singuliers

Définition 7.7. Le *Hamiltonien* du système (7.19) est la fonction :

$$\begin{aligned} \mathbf{H} : \mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}^m &\longrightarrow \mathbb{R} \\ (x, p, u) &\longmapsto \mathbf{H}(x, p, u) = p \cdot f(x, u). \end{aligned}$$

Proposition 7.12. Soit u un contrôle singulier sur $[0, T]$, et soit $x(\cdot)$ la trajectoire singulière associée. Alors il existe une application absolument continue $p : [0, T] \longrightarrow \mathbb{R}^n \setminus \{0\}$,

appelée *vecteur adjoint*, telle que les équations suivantes sont vérifiées pour presque tout $t \in [0, T]$:

$$\dot{x}(t) = \frac{\partial \mathbf{H}}{\partial p}(x(t), p(t), u(t)), \quad (7.21)$$

$$\dot{p}(t) = -\frac{\partial \mathbf{H}}{\partial x}(x(t), p(t), u(t)), \quad (7.22)$$

$$\frac{\partial \mathbf{H}}{\partial u}(x(t), p(t), u(t)) = 0, \quad (7.23)$$

où \mathbf{H} est le hamiltonien du système.

L'équation (7.23) est appelée *équation de contrainte*. Il faut remarquer que $p(t) \neq 0$ pour tout $t \in [0, T]$.

PREUVE.

▷ Par définition, le couple (x, u) est singulier sur $[0, T]$ si $dE_T(u)$ n'est pas surjective. Donc il existe un vecteur ligne $\psi \in \mathbb{R}^n \setminus \{0\}$ tel que pour tout contrôle v dans L^∞ on ait :

$$\psi \cdot dE_T(u) \cdot v = \psi \int_0^T M(T)M^{-1}(s)B(s)v(s)ds = 0$$

Par conséquent :

$$\psi M(T)M^{-1}(s)B(s) = 0 \quad \text{p.p. sur } [0, T].$$

On pose $p(t) = \psi M(T)M^{-1}(t)$ pour tout $t \in [0, T]$. C'est un vecteur ligne de $\mathbb{R}^n \setminus \{0\}$, et $p(T) = \psi$. On a par dérivation :

$$\dot{p}(t) = -p(t) \frac{\partial f}{\partial x}(x(t), u(t)).$$

En introduisant le Hamiltonien $\mathbf{H}(x, p, u) = p \cdot f(x, u)$, on obtient :

$$f(x(t), u(t)) = \frac{\partial \mathbf{H}}{\partial p}(x(t), p(t), u(t)),$$

et

$$-p(t) \frac{\partial f}{\partial x}(x(t), u(t)) = -\frac{\partial \mathbf{H}}{\partial x}(x(t), p(t), u(t)).$$

La dernière relation vient de $p(t)B(t) = 0$ car $B(t) = \frac{\partial f}{\partial u}(x(t), u(t))$.

□

7.4.3 Démonstration du Théorème 7.7

Associons au système (7.14), (7.15) le *système augmenté* suivant :

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)), \\ \dot{x}^0(t) &= f^0(x(t), u(t)), \end{aligned} \quad (7.24)$$

et notons $\tilde{x} = (x, x^0)$, $\tilde{f} = (f, f^0)$. Le problème revient donc à chercher une trajectoire solution de (7.24) joignant les points $\tilde{x}_0 = (x_0, 0)$ et $\tilde{x}_1 = (x_1, x^0(T))$, et minimisant la dernière coordonnée $x^0(T)$.

L'ensemble des états accessibles à partir de \tilde{x}_0 pour le système (7.24) est $\tilde{\mathcal{A}}(\tilde{x}_0, T) = \bigcup_{u(\cdot)} \tilde{x}(T, \tilde{x}_0, u)$.

Le lemme crucial est alors le suivant.

Lemme 7.13. *Si le contrôle u associé au système de contrôle (7.14) est optimal pour le coût (7.15), alors il est singulier sur $[0, T]$ pour le système augmenté (7.24).*

PREUVE.

▷ Notons \tilde{x} la trajectoire associée, solution du système augmenté (7.24), issue de $\tilde{x}_0 = (x_0, 0)$. Le contrôle u étant optimal pour le coût (7.15), il en résulte que le point $\tilde{x}(T)$ appartient à la frontière de l'ensemble $\tilde{\mathcal{A}}(T, \tilde{x}_0)$, voir figure 7.1. En effet sinon, il existerait un voisinage du point $\tilde{x}(T) = (x_1, x^0(T))$ dans $\tilde{\mathcal{A}}(T, \tilde{x}_0)$ contenant un point $\tilde{y}(T)$ solution du système (7.24) et tel que l'on ait $y^0(T) < x^0(T)$, ce qui contredirait l'optimalité du contrôle u . Par conséquent, d'après la proposition 7.9, le contrôle u est un contrôle singulier pour le système augmenté (7.24) sur $[0, T]$. □

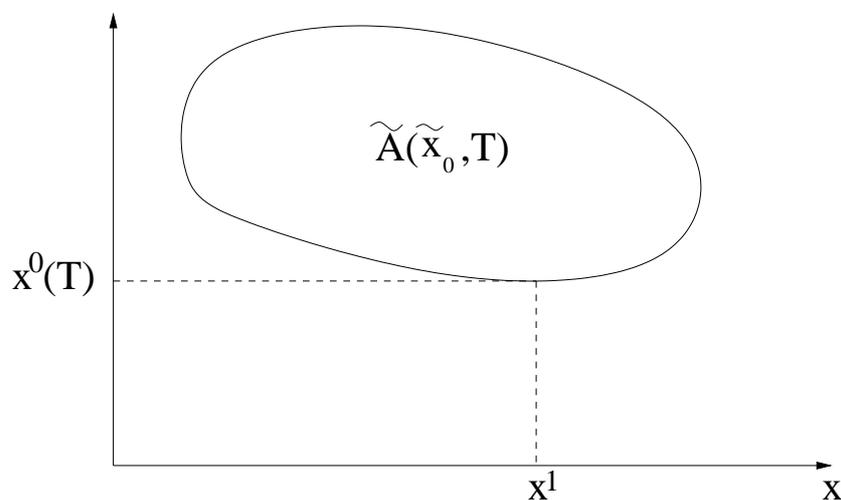


FIGURE 7.1 – Ensemble accessible augmenté.

Dans la situation du lemme, d'après la proposition 7.12, il existe une application $\tilde{p} : [0, T] \rightarrow \mathbb{R}^{n+1} \setminus \{0\}$ telle que $(\tilde{x}, \tilde{p}, \tilde{u})$ soit solution du système hamiltonien :

$$\dot{\tilde{x}}(t) = \frac{\partial \tilde{H}}{\partial \tilde{p}}(\tilde{x}(t), \tilde{p}(t), u(t)), \quad \dot{\tilde{p}}(t) = -\frac{\partial \tilde{H}}{\partial \tilde{x}}(\tilde{x}(t), \tilde{p}(t), u(t)), \quad (7.25)$$

$$\frac{\partial \tilde{H}}{\partial u}(\tilde{x}(t), \tilde{p}(t), u(t)) = 0 \quad (7.26)$$

où $\tilde{H}(\tilde{x}, \tilde{p}, u) = \tilde{p} \cdot \tilde{f}(t, \tilde{x}, u)$.

En écrivant $\tilde{p} = (p, p^0) \in (\mathbb{R}^n \times \mathbb{R}) \setminus \{0\}$, où p^0 est appelée variable duale du coût, on obtient :

$$(\dot{p}, \dot{p}^0) = -(p, p^0) \begin{pmatrix} \frac{\partial f}{\partial x} & 0 \\ \frac{\partial f^0}{\partial x} & 0 \end{pmatrix}.$$

Or $\tilde{H} = \tilde{p} \cdot \tilde{f}(x, u) = pf + p^0 f^0$, donc :

$$\frac{\partial \tilde{H}}{\partial u} = 0 = p \frac{\partial f}{\partial u} + p^0 \frac{\partial f^0}{\partial u},$$

et donc $\dot{p}^0(t) = 0$, c'est-à-dire $p^0(t)$ est constant sur $[0, T]$. Comme le vecteur $\tilde{p}(t)$ est défini à scalaire multiplicatif près, on choisit $p^0 \leq 0$. Le théorème est démontré.