

Principes fondamentaux de l'Automatique : dynamique et contrôle des systèmes

Nicolas Petit

Pierre Rouchon

MINES ParisTech

PSL Research University

CAS - Centre Automatique et Systèmes

Novembre 2018

Avant propos

Les mécanismes de régulation et d'adaptation des systèmes, largement répandus dans la nature, sont à la base du fonctionnement de nombreuses technologies, machines et inventions créées et utilisées par l'homme. Au temps de la révolution industrielle du XIXème siècle, le régulateur de Watt a été un élément constitutif des machines à vapeur. Puis, les régulateurs ont vu le jour dans de nombreux véhicules (aéroplanes, automobiles, dès les années 1920). Aujourd'hui on retrouve des régulateurs dans les systèmes de communication, les usines de production manufacturière, les transports publics...

Graduellement, une théorie de l'automatique (Control Theory), aussi appelée théorie mathématique des systèmes (Mathematical Systems Theory), s'est constituée. On a vu apparaître les premières formalisations modernes : modélisation (avec les équations différentielles inventées par Newton) et stabilité. Le point de départ de cette théorie remonte ainsi aux travaux du mathématicien et astronome anglais G. Airy. Il fut le premier à tenter une analyse du régulateur de Watt. Ce n'est qu'en 1868, que le physicien écossais J. C. Maxwell publia une première analyse mathématique convaincante et expliqua certains comportements erratiques observés parmi les nombreux régulateurs en service à cet époque.

Ses travaux furent suivis par de nombreux autres sur la stabilité, notion marquée par le travail de H. Poincaré et A. M. Lyapounov, sa caractérisation pour les systèmes linéaires stationnaires ayant été obtenue indépendamment par les mathématiciens A. Hurwitz et E. J. Routh. Les recherches initiées dans les années 1930 aux "Bell Telephone Laboratories" sur les amplificateurs et poursuivies pendant la deuxième guerre mondiale au sein du "Radiation Laboratory" du MIT ont été à l'origine de notions encore enseignées aujourd'hui. Citons par exemple les travaux de Nyquist et de Bode caractérisant à partir de la réponse fréquentielle en boucle ouverte celle de la boucle fermée. L'après deuxième guerre mondiale vit l'essor de la forme d'état, et son application à la conquête spatiale, notamment sous la forme du filtre de Kalman. Les années récentes ont vu l'intégration de calculs embarqués complexes dans des contrôleurs de plus en plus miniaturisés ainsi que l'apparition de plusieurs expériences menées sur des circuits obéissant aux lois de la physique quantique.

Initialement, l'automatique a étudié le cadre des systèmes linéaires ayant une seule commande et une seule sortie. On disposait d'une mesure sous la forme d'un signal électrique. Cette dernière était alors entrée dans un amplificateur qui restituait en sortie un autre signal électrique qu'on utilisait comme signal de contrôle. Ce n'est qu'après les années 1950 que les développements théoriques et technologiques (avec l'invention des calculateurs numériques) permirent le traitement des systèmes multi-variables linéaires et non linéaires ayant plusieurs entrées et plusieurs sorties. Citons comme contributions importantes dans les années 1960 celles de R. Bellmann avec la programmation dynamique, celles de R. Kalman avec la commandabilité, le filtrage et la commande linéaire quadratique ; celles de L. Pontryagin avec la commande optimale. Ces contributions continuent encore aujourd'hui à alimenter les recherches en théorie mathématique des systèmes et restent très utiles pour aborder les systèmes quantiques.

Aujourd'hui l'automatique est omniprésente dans les domaines industriels tels que l'aéronautique, l'aérospatiale, l'automobile, ou le génie des procédés. Les notions clés présentées dans ce cours concernent également des systèmes grand public de l'industrie du jouet (drones miniatures), la photographie (objectifs stabilisés), la mise au point de fusées expérimentales, les systèmes de localisation,....

Ce cours est une introduction aux trois grands thèmes sur lesquels reposent l'automatique et la théorie mathématique des systèmes :

1. Systèmes dynamiques : stabilité, robustesse, théorie de perturbations.
2. Commandabilité : stabilisation par bouclage, planification et suivi de trajectoire.

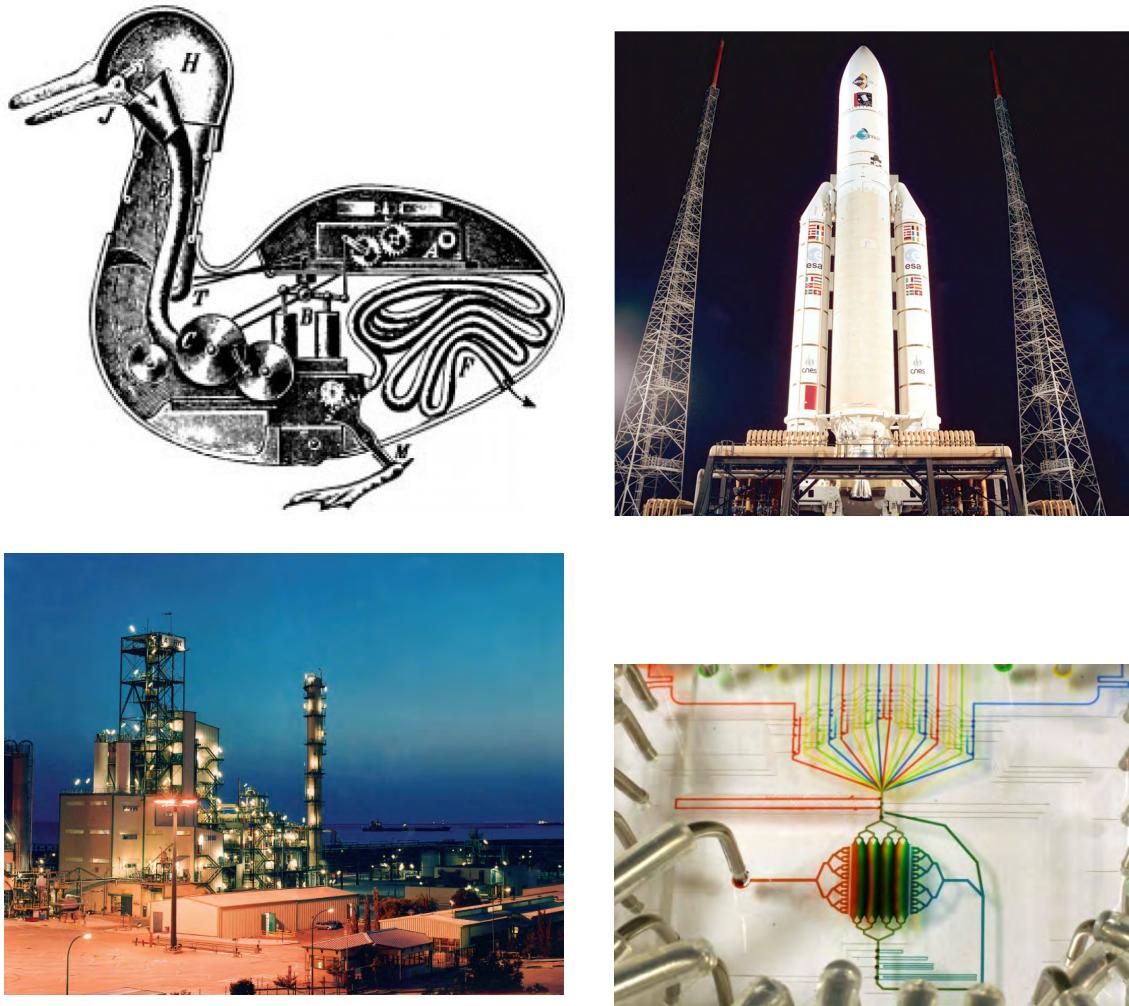


FIGURE 1 – Canard mécanique de Vaucanson, Fusée Ariane 5, Usine PP2 Lavéra, Système microfluide.

3. Observabilité : estimation, observateur asymptotique, filtrage et diagnostic.

Dans bien des domaines scientifiques, une théorie a très souvent pour origine une classe d'exemples représentatifs bien compris et analysés. Nous nous inscrivons dans cette démarche. Le cours expose plusieurs exemples issus du monde industriel ou académique pour motiver et justifier les définitions et résultats abstraits sur lesquels reposent une classe d'algorithmes de contrôle, de filtrage et d'estimation. L'automatique et la théorie mathématique des systèmes étant un domaine actif de recherche scientifique, le cours abordera certaines questions qui n'admettent pas de réponse définitive aujourd'hui bien qu'elles aient de fortes motivations pratiques.

Note de cette édition Nous vous serions reconnaissants de nous faire part de vos critiques et des erreurs que vous auriez découvertes par un message explicatif à

nicolas.petit@mines-paristech.fr ou à
pierre.rouchon@mines-paristech.fr

Table des matières

1 Systèmes dynamiques et régulateurs	7
1.1 Systèmes dynamiques non linéaires	7
1.1.1 Existence et unicité des solutions	10
1.1.2 Sensibilité et première variation	12
1.1.3 Stabilité locale autour d'un équilibre	13
1.2 Systèmes dynamiques linéaires	16
1.2.1 L'exponentielle d'une matrice	16
1.2.2 Forme de Jordan et calcul de l'exponentielle	16
1.2.3 Portraits de phases des systèmes linéaires	19
1.2.4 Polynôme caractéristique	21
1.2.5 Systèmes linéaires instationnaires	23
1.2.6 Compléments : matrices symétriques et équation de Lyapounov	24
1.3 Stabilité des systèmes non linéaires	25
1.3.1 Étude au premier ordre	25
1.3.2 Fonctions de Lyapounov	27
1.3.3 Robustesse paramétrique	32
1.3.4 Compléments : caractère intrinsèque des valeurs propres du système linéarisé tangent	32
1.3.5 Compléments : les systèmes dynamiques dans le plan	33
1.4 Systèmes multi-échelles lents/rapides	38
1.4.1 Perturbations singulières	39
1.4.2 Feedback sur un système à deux échelles de temps	44
1.4.3 Modèle de contrôle et modèle de simulation	45
1.5 Cas d'étude	47
1.5.1 Présentation du système	47
1.5.2 Un régulateur PI	47
1.5.3 Une modélisation simplifiée	49
1.5.4 Passage en temps continu	50
1.5.5 Simulations en boucle ouverte et en boucle fermée	51
1.5.6 Un résultat général : régulateur PI sur un système non linéaire du premier ordre	52
1.5.7 Dynamiques négligées : rôle du contrôle dans l'approximation	56
1.5.8 Intérêt de la pré-compensation (feedforward)	58
1.5.9 Pré-compensation et suivi de trajectoires sur un système linéaire du premier ordre	59
1.6 Cas d'étude	61
2 Fonctions de transfert	65

2.1	Passage à la fonction de transfert	66
2.1.1	Questions de robustesse	66
2.1.2	Principe des calculs	66
2.1.3	Régime asymptotique forcé	68
2.1.4	Simplifications pôles-zéros	68
2.1.5	Formalisme	70
2.1.6	Du transfert vers l'état : réalisation	71
2.2	Schémas blocs et fonctions de transfert	73
2.2.1	De la forme d'état vers le transfert	73
2.2.2	Transfert avec perturbation et bouclage	75
2.3	Marge de robustesse	76
2.3.1	Critère de Nyquist	76
2.3.2	Marge de phase et retard critique	82
2.3.3	Marge de gain	85
2.3.4	Lecture des marges sur le diagramme de Bode	86
2.3.5	Pôles dominants	87
2.4	Compléments	88
2.4.1	Calcul de tous les contrôleurs PID stabilisant un système du premier ordre à retard	88
2.4.2	Méthodes de réglage de Ziegler-Nichols	90
2.4.3	Prédicteur de Smith	92
2.4.4	Systèmes à non minimum de phase	94
3	Commandabilité, stabilisation	97
3.1	Un exemple de planification et de suivi de trajectoires	97
3.1.1	Modélisation de deux oscillateurs en parallèle	98
3.1.2	Planification de trajectoires	98
3.1.3	Stabilisation et suivi de trajectoires	99
3.1.4	Autres exemples	101
3.2	Commandabilité non linéaire	104
3.2.1	Exemple de non commandabilité par la présence d'intégrale première	105
3.3	Commandabilité linéaire	107
3.3.1	Matrice de commandabilité et intégrales premières	107
3.3.2	Invariance	109
3.3.3	Critère de Kalman et forme de Brunovsky	110
3.3.4	Planification et suivi de trajectoires	114
3.4	Commande linéaire quadratique LQR	115
3.4.1	Multiplicateurs de Lagrange en dimension infinie	117
3.4.2	Problème aux deux bouts dans le cas linéaire quadratique	121
3.4.3	Planification de trajectoires	123
3.4.4	Régulateur LQR	125
3.5	Compléments	132
3.5.1	Linéarisation par bouclage	132
3.5.2	Stabilisation par méthode de Lyapounov et backstepping	136
4	Observabilité, estimation et adaptation	139
4.1	Un exemple	139

4.1.1	Un modèle simple de moteur à courant continu	140
4.1.2	Estimation de la vitesse et de la charge	141
4.1.3	Prise en compte des échelles de temps	142
4.1.4	Contraintes sur les courants	143
4.2	Observabilité non linéaire	143
4.2.1	Définition	144
4.2.2	Critère	144
4.2.3	Observateur, estimation, moindres carrés	146
4.3	Observabilité linéaire	147
4.3.1	Le critère de Kalman	147
4.3.2	Observateurs asymptotiques	149
4.3.3	Observateurs réduits de Luenberger	149
4.4	Observateur-contrôleur	150
4.4.1	Version état multi-entrée multi-sortie (MIMO) ¹	150
4.4.2	Version transfert mono-entrée mono-sortie (SISO) ²	151
4.5	Filtre de Kalman	152
4.5.1	Formalisme	153
4.5.2	Hypothèses et définition du filtre	154
4.6	Compléments	160
4.6.1	Estimation de paramètres et commande adaptative	160
4.6.2	Linéarisation par injection de sortie	161
4.6.3	Contraction	162
A	Théorème de Cauchy-Lipchitz	165
B	Fonctions de Lyapounov et stabilité des points d'équilibre	171
C	Moyennisation	177
C.1	Introduction	177
C.2	Le résultat de base	177
C.3	Un exemple classique	179
C.4	Recherche d'extremum (extremum seeking)	180
C.5	Boucle à verrouillage de phase PLL	181
D	Automatique en temps discret	185
D.1	Représentation externe et interne	189
D.1.1	Transformée en z	190
D.1.2	Réalisation canonique d'une fonction de transfert discrète	190
D.2	Analyse de la stabilité	191
D.3	Commandabilité en temps discret	192
D.3.1	Placement de pôles	193
D.3.2	Rendre une matrice nilpotente par feedback	193
D.3.3	Synthèse d'une commande pour aller en temps fini à un point d'équilibre arbitraire	194
D.3.4	Commande linéaire quadratique LQR en temps discret	194

1. MIMO : pour Multi-Input Multi-Output

2. SISO : pour Single-Input Single-Output

D.4 Observabilité, reconstructibilité et filtrage	196
D.4.1 Observabilité en temps discret	196
D.4.2 Filtre de Kalman en temps discret	197
Références	198
Index	203

Chapitre 1

Systèmes dynamiques et régulateurs

1.1 Systèmes dynamiques non linéaires

Nous nous intéressons à des systèmes dynamiques représentés par un nombre fini d'équations différentielles du premier ordre couplées entre elles que nous écrivons

$$\begin{aligned}\frac{d}{dt}x_1 &= f_1(x_1, \dots, x_n, u_1, \dots, u_m, t) \\ \frac{d}{dt}x_2 &= f_2(x_1, \dots, x_n, u_1, \dots, u_m, t) \\ &\vdots \\ \frac{d}{dt}x_n &= f_n(x_1, \dots, x_n, u_1, \dots, u_m, t)\end{aligned}$$

où les grandeurs x_1, \dots, x_n sont appelées *états*, les grandeurs u_1, \dots, u_m sont les *entrées* (ou *commandes*), et n et m sont des entiers. Le temps t est ici considéré de manière générale dans le second membre des équations. Les fonctions $(f_i)_{i=1, \dots, n}$ sont à valeur réelle. En toute généralité, leur régularité peut être très faible, même si, comme on le verra par la suite, de nombreuses propriétés fondamentales proviennent justement de leur régularité.

Il est commode d'écrire le système différentiel précédent sous la forme vectorielle (appelée *forme d'état*)

$$\frac{d}{dt}x = f(x, u, t) \tag{1.1}$$

Pour calculer l'évolution future d'un tel système, il faut connaître les grandeurs $t \mapsto u(t)$ ainsi que la condition initiale de l'état. On dit que l'état x du système représente sa mémoire. Étant donnée l'évolution du système, on s'intéresse souvent à un certain nombre de grandeurs (par exemple car elles ont un intérêt pratique) qu'on nomme *sorties* ou *mesures*. Les *équations de sorties* que nous considérons sont de la forme

$$y = h(x, u, t) \in \mathbb{R}^q \tag{1.2}$$

où q est un entier souvent inférieur à n ¹.

1. Bien souvent seules certaines composantes sont mesurées pour des raisons technologiques, il existe des exceptions notables comme dans les applications de fusion de capteurs (voir le Chapitre 4) où on dispose de mesures redondantes de certaines composantes de l'état x .

Le formalisme (1.1) (1.2) que nous venons de présenter est très général. On trouve un vaste choix d'exemples dans les domaines de la mécanique (les équations d'Euler-Lagrange, voir [46] par exemple, exprimant le principe de moindre action sont des équations d'ordre 2 dans les variables de configurations (positions généralisées), l'état est alors constitué des positions généralisées et leur vitesses), les machines électriques (voir [48]), l'aéronautique (voir l'Exemple 1), la robotique (voir [60]), la chimie, la biochimie, les sciences du vivant (voir [37])².

Ce que nous cherchons à faire dans le domaine de l'Automatique ce n'est pas simplement étudier les propriétés des systèmes d'équations différentielles, mais les contrôler. Dans l'équation (1.1), on peut agir sur l'état x en choisissant u . L'outil principal de l'Automaticien c'est la *rétro-action* $u = k(t, x)$ aussi appelée *feedback*. En spécifiant de la sorte la commande, on change complètement le comportement du système dynamique (1.1) qui devient

$$\frac{d}{dt}x = f(x, k(t, x), t)$$

En pratique, on sera souvent limité à l'utilisation de la mesure définie en (1.2). Dans ce contexte, les rétro-actions envisageables sont de la forme $u = k(t, y)$. On parle alors de *retour de sortie*. C'est un problème difficile pour lequel nous montrerons, au Chapitre 4 qu'il est en fait utile d'utiliser un système dynamique supplémentaire (on parlera d'*observateur*) qui permettra de bien plus intéressantes possibilités.

Que ce soit parce qu'on a choisi une loi de rétroaction particulière ou parce que le système considéré ne possède pas de commande, il est important de comprendre comment étudier les *systèmes libres* de la forme

$$\frac{d}{dt}x = f(x, t) \tag{1.3}$$

On dira qu'un tel système est *stationnaire* (par opposition au cas (1.3) *instationnaire*) lorsque f ne dépend pas explicitement du temps. Dans ce cas l'équation s'écrit

$$\frac{d}{dt}x = f(x)$$

Un des concepts clés dans l'étude des systèmes dynamiques est la notion de *point d'équilibre* (appelé également *point stationnaire*) \bar{x} . On appelle point d'équilibre, un point \bar{x} tel que, si le système différentiel (1.3) est initialisé en ce point, c.-à-d. $x(0) = \bar{x}$, alors le système reste en \bar{x} pour tous les temps futurs. Dans le cas d'un système stationnaire, les points d'équilibre sont simplement³ caractérisés par l'équation $f(\bar{x}) = 0$.

Autour d'un point d'équilibre il est tentant de chercher une expression approchée des équations dynamiques (1.3) qu'on souhaite étudier. En général, lorsque le système est instationnaire, on obtiendra un système linéaire instationnaire (appelé *linéarisé tangent*) de la forme

$$\frac{d}{dt}x = A(t)x + B(t)u \tag{1.4}$$

$$y = C(t)x + D(t)u \tag{1.5}$$

2. Pourtant, on pourra remarquer que certains domaines échappent à ce formalisme. On peut citer quelques exemples tels que l'étude de la turbulence dans les écoulements, les phénomènes thermiques dans les chambres de combustion, ou les ondes de propagation. L'étude des systèmes de dimension infinie est traitée dans de nombreux ouvrages sur les équations différentielles partielles [19] aussi appelés systèmes à paramètres distribués. Ils sont en grande partie absent de ce cours.

3. On notera l'intérêt d'un formalisme du premier ordre. Si on considère l'équation différentielle scalaire du second ordre $\frac{d^2}{dt^2}\theta = 2\theta$, dont l'état est $x = (\theta, \frac{d}{dt}\theta = \omega)^T$, les points d'équilibre sont donnés par $\bar{x} = (\bar{\theta}, \bar{\omega}) = (0, 0)$ et pas uniquement par l'équation $2\bar{\theta} = 0$.

Comme nous le verrons (notamment à la Section 1.3), cette approche est intéressante car elle fournit souvent une information *locale* sur le comportement de (1.3) autour de \bar{x} . Néanmoins, cette information est parfois insuffisante, surtout lorsqu'on recherche des informations sur le comportement du système loin de l'équilibre. En outre, les systèmes linéaires et les systèmes non linéaires sont en fait très différents par nature. Soumis à une superposition d'excitations (par une somme de termes dans la commande), les systèmes linéaires répondent d'après le *principe de superposition* : leur sortie est la somme des sorties correspondant à chacun des termes d'excitation. Ce principe, qui ne s'applique pas aux systèmes non linéaires, est à la base de l'analyse de Fourier des signaux. Il existe de nombreux phénomènes qu'on ne constate que chez les systèmes régis par des équations non linéaires (on pourra se référer à [42] pour une discussion plus détaillée). Les systèmes non linéaires peuvent diverger en temps fini alors que c'est seulement en temps infini qu'un système linéaire peut diverger (c.-à-d. que son état tend vers l'infini). Un système non linéaire peut avoir de multiples points d'équilibre isolés. L'unicité des points d'équilibre n'est pas non plus garantie dans le cas des systèmes linéaires, mais elle apparaît sous la forme de sous espaces vectoriels de points d'équilibre. Enfin, et c'est un des faits les plus marquants, un système non linéaire peut posséder un ou plusieurs cycles limites. C'est une des particularités les plus importantes et aussi l'une des plus fréquemment observées en pratique. Alors que les systèmes linéaires peuvent avoir des trajectoires oscillantes, l'amplitude de celles-ci dépendent de la condition initiale. Dans le cas non linéaire il est possible d'observer des trajectoires oscillantes dont l'amplitude ne dépend pas de la condition initiale. C'est en exploitant cette propriété qu'on a construit les premiers circuits électriques oscillants qui sont à la base de l'électronique (et de la synthèse sonore notamment). À titre d'illustration, on pourra se reporter à l'onde de densité observée dans les puits de pétrole présentés dans l'Exemple 10.

Exemple 1 (Missile dans le plan). *En grande partie à cause des effets aérodynamiques, les équations qui régissent le mouvement d'un missile ou d'une fusée sont non linéaires. Après réduction à un plan, les équations de la dynamique d'un missile s'expriment au moyen de 6 états. Les variables x_m et y_m correspondent aux coordonnées du centre de gravité de l'engin, V_m est sa vitesse, χ_m , α_m et β_m correspondent à des angles, et on utilise deux variables intermédiaires ϕ_A et α_T pour exprimer les seconds membres de la dynamique (voir la Figure 1.1).*

$$\left. \begin{array}{l} \dot{x}_m = V_m \cos \chi_m \\ \dot{y}_m = V_m \sin \chi_m \\ \dot{V}_m = \frac{1}{m} \left(T \cos \alpha_T - \frac{1}{2} \rho S V_m^2 C_D(\alpha_T) \right) \\ \text{où } \sin \alpha_T = \sqrt{1 - \frac{1}{1 + \tan^2 \alpha_m + \tan^2 \beta_m}} \\ \dot{\chi}_m = \tan \phi_A \frac{g}{V_m} \\ \text{où } 1 + \tan^2 \phi_A = \frac{\tan^2 \alpha_T}{\tan^2 \alpha_m} \\ \dot{\alpha}_m = 0.3(\alpha_c - \alpha_m) \\ \dot{\beta}_m = 0.3(\beta_c - \beta_m) \end{array} \right\} \quad (1.6)$$

Les deux commandes sont α_c , β_c , tandis que $m(t)$ et $T(t)$ sont des fonctions données du temps (variation de la masse et de la poussée), ρ et S sont des constantes et $C_D(\cdot)$ est le coefficient de traînée. Ce modèle prend en compte les effets aérodynamiques dus aux angles d'attaque du corps du missile,

la perte de masse au cours du vol (qui est loin d'être négligeable), et les dynamiques des actionneurs (à travers des vérins).

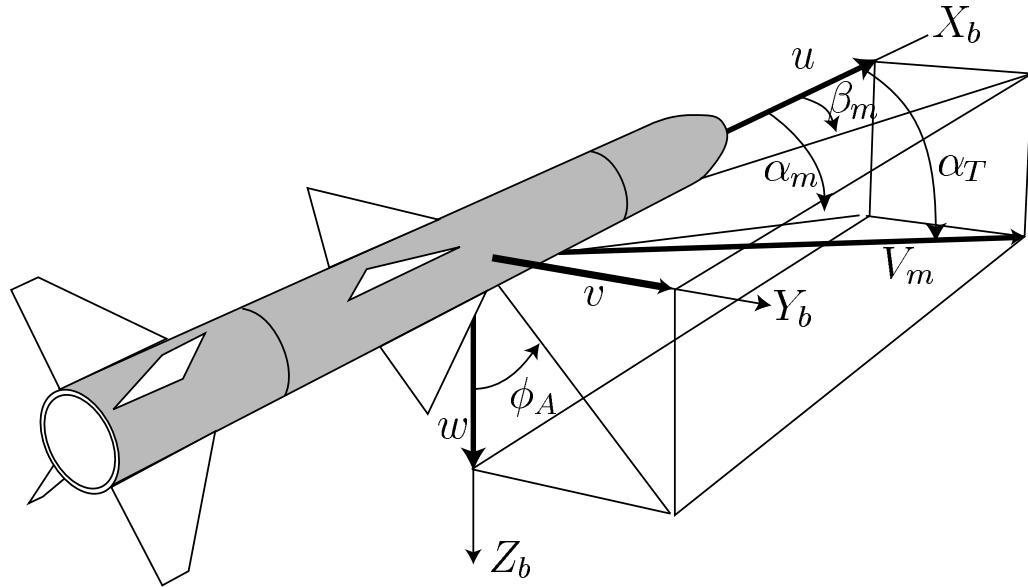


FIGURE 1.1 – Variables décrivant un missile.

1.1.1 Existence et unicité des solutions

L'outil de base de la modélisation mathématique d'un modèle physique est le *problème de Cauchy*. Il consiste en une condition initiale et une équation différentielle

$$x(0) = x^0, \quad \frac{d}{dt}x(t) = f(x(t), t) \quad (1.7)$$

Face à ce problème, on est en droit d'espérer que les propriétés suivantes sont satisfaites

1. qu'une solution avec ces conditions initiales existe
2. que cette solution soit unique
3. que les solutions dépendent continûment des conditions initiales

On trouvera une démonstration élémentaire du théorème de Cauchy-Lipschitz (à partir du schéma d'Euler explicite et de la notion de solution approchante) dans l'Annexe A. Ce théorème dont une version un peu plus générale est donnée ci-dessous garantie l'existence et l'unicité pour des temps courts. En revanche l'existence pour des temps t arbitrairement grands est nettement plus délicate. On peut la garantir sous des hypothèses fortes (voir Théorème 3). Enfin, la continuité de la solution par rapport à sa condition initiale peut être garantie, mais sous une hypothèse plus forte. C'est l'objet du Théorème 2.

Théorème 1 (Cauchy-Lipschitz. Existence et unicité)

Soit $\mathbb{R}^n \times \mathbb{R} \ni (x, t) \mapsto f(x, t) \in \mathbb{R}^n$ une fonction satisfaisant aux deux propriétés suivantes :

1. (localement lipchitzienne par rapport à x) pour tout $(x, t) \in \mathbb{R}^n \times \mathbb{R}$, il existe $\eta_{x,t} > 0$ et $K_{x,t} > 0$, tels que, pour tout $y \in \mathbb{R}^n$ vérifiant $\|x - y\| \leq \eta_{x,t}$ on a $\|f(x, t) - f(y, t)\| \leq K_{x,t} \|x - y\|$
2. (localement intégrable par rapport à t) pour tout $x \in \mathbb{R}^n$, l'application $t \mapsto f(x, t)$ est localemement intégrable^a.

On considère le problème de Cauchy,

$$x(0) = x^0, \quad \frac{d}{dt}x(t) = f(x(t), t),$$

Ce problème possède les propriétés suivantes

1. existence locale en temps : pour toute condition initiale $x^0 \in \mathbb{R}^n$, il existe $\epsilon > 0$ et une fonction $] -\epsilon, \epsilon[\ni t \mapsto x(t) \in \mathbb{R}^n$ dérivable par rapport à t et solution du problème de Cauchy.
2. unicité globale en temps : pour toute condition initiale $x^0 \in \mathbb{R}^n$ et pour tout $a, b > 0$ il existe au plus une fonction $] -a, b[\ni t \mapsto x(t) \in \mathbb{R}^n$ dérivable par rapport à t et solution du problème de Cauchy.

a. À x fixé, chacune de ses composantes est une fonction mesurable du temps dont l'intégrale de la valeur absolue sur tout intervalle borné est bornée, voir [54].

Une démonstration de ce théorème, reposant sur le théorème du point fixe de Picard dans un espace de Banach, figure dans [53, 34]. On peut vérifier grâce aux exemples qui suivent que les hypothèses du Théorème 1 de Cauchy-Lipchitz sont effectivement nécessaires :

- l'équation $\frac{d}{dt}x = \sqrt{|x|}$ admet deux solutions partant de $x^0 = 0$: $x(t) = 0$ et $x(t) = t^2/4$; on remarque que $\sqrt{|x|}$ n'est pas Lipchitz en $x = 0$ bien qu'elle y soit continue. En dehors du point 0, l'unique solution de l'équation différentielle (voir [10, page 149]) est $x(0) = l$, $x(t) = \frac{1}{4}(t + 2\sqrt{l})^2$.
- toute solution de $\frac{d}{dt}x = 1/t$ n'est pas définie en $t = 0$ car $1/t$ n'est pas localemement intégrable autour de 0 alors que $\frac{d}{dt}x = 1/\sqrt{|t|}$ admet des solutions parfaitement définies en 0.

Théorème 2 (Dépendance en la condition initiale)

Soit $\mathbb{R}^n \times \mathbb{R} \ni (x, t) \mapsto f(x, t) \in \mathbb{R}^n$ une fonction continue à dérivées partielles $\frac{\partial f}{\partial x_i}$, $i = 1 \dots n$, continues par rapport à x . Soit $x^0 \in \mathbb{R}^n$ une condition initiale, il existe une unique solution $t \mapsto x(t)$ telle que $x(0) = x^0$. Cette solution, considérée comme une fonction de la condition initiale x^0 , admet des différentielles partielles $\frac{\partial x_i(x^0, t)}{\partial x_j^0}$ continues par rapport à x^0 et à t .

Ce théorème est démontré dans [36, page 29]. Il est illustré par la Figure 1.2 où on a représenté trois solutions de la même équation différentielle scalaire (à second membre C^1) obtenues en faisant

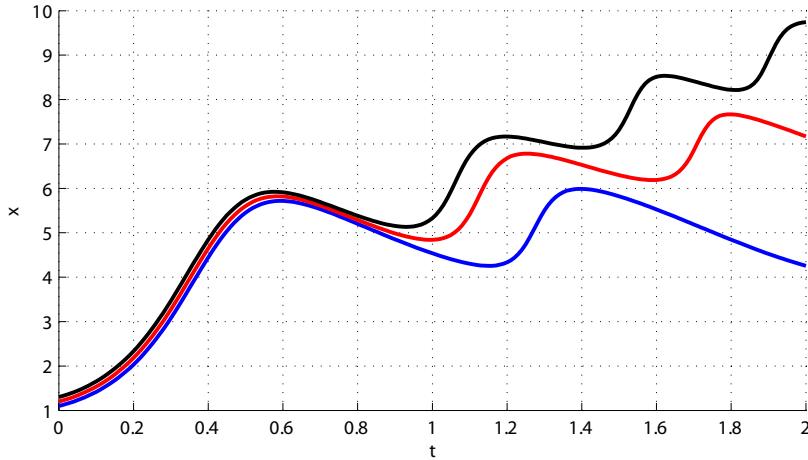


FIGURE 1.2 – Dépendance de la solution d'une équation différentielle par rapport à sa condition initiale. Trois solutions d'une équation scalaire au cours du temps. Elles ne peuvent se croiser en vertu de la propriété d'unicité.

varier la condition initiale. On remarque que les courbes ne se croisent pas (cela est interdit par la propriété d'unicité). Elles semblent s'écartier lorsque t croît.

L'existence pour tout temps $t > 0$ n'est pas une conséquence du Théorème 1 de Cauchy-Lipchitz. En effet, $\frac{d}{dt}x = x^2$ vérifie bien toutes ses hypothèses, mais comme la solution qui part de x^0 est $x(t) = \frac{x^0}{1-tx^0}$, on voit que, pour $x^0 > 0$, la solution explose en $t = 1/x^0$. La solution explose en temps fini.

On peut donner une condition assez générale (mais relativement forte) pour éviter l'explosion en temps fini. C'est l'objet du lemme suivant.

Théorème 3 (existence globale en temps)

En plus des hypothèses du Théorème 1, si on suppose qu'il existe $M_0(t), M_1(t) > 0$ fonctions localement intégrables telles que pour tout $x \in \mathbb{R}^n$, $\|f(x, t)\| \leq M_0(t) + M_1(t)\|x\|$, alors la solution (unique) au problème de Cauchy est définie pour $t \in]-\infty, +\infty[$.

Noter que ce lemme ne dit pas que les trajectoires restent bornées. A priori, elles ne le sont pas comme le montre l'exemple $\frac{d}{dt}x = x$: elles peuvent tendre vers l'infini mais en temps infini.

1.1.2 Sensibilité et première variation

Nous nous intéressons à la solution du système

$$\frac{d}{dt}x(t) = f(x(t), u(t), p), \quad x(0) = x^0$$

où $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $p \in \mathbb{R}^r$, f est une fonction C^1 de x , u et p ($n, m, r \in \mathbb{N}$). On suppose ici que la fonction $\mathbb{R} \ni t \mapsto u(t) \in \mathbb{R}^m$ est continue par morceaux. Ainsi, les hypothèses du théorème 1 sont satisfaites et donc la solution $t \mapsto x(t)$ existe pour t autour de 0 et est unique. On regarde

sa dépendance à des petites variations de l'entrée u , du paramètre p ou de sa condition initiale x^0 . Nous ne présentons pas de façon rigoureuse les résultats de dérivabilité de x par rapport à u , p et x^0 . Ils découlent du théorème 2 et du fait que f est C^1 . Nous présentons simplement la façon de calculer directement les dérivées partielles de x par rapport aux composantes de x^0 et p , ainsi que la différentielle de x par rapport à u .

La méthode est élémentaire : il suffit de différentier les équations. On note δ l'opération de différentiation. Ainsi les petites variations de u , p et x^0 , notées δu , δp et δx^0 , engendrent des petites variations de x , notées δx . Il faut bien comprendre que δp et δx^0 sont des petits vecteurs de \mathbb{R}^r et \mathbb{R}^n alors que δu et δx sont des petites fonctions du temps à valeur dans \mathbb{R}^m et \mathbb{R}^n , respectivement. Ainsi δx est solution de l'équation différentielle

$$\frac{d}{dt}(\delta x)(t) = \left(\frac{\partial f}{\partial x} \right)_t \delta x + \left(\frac{\partial f}{\partial u} \right)_t \delta u(t) + \left(\frac{\partial f}{\partial p} \right)_t \delta p$$

avec comme condition initiale $\delta x(0) = \delta x^0$. Les matrices Jacobiennes $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial u}$ et $\frac{\partial f}{\partial p}$ sont évaluées le long de la trajectoire $(x(t), u(t), p)$. C'est pourquoi nous les notons $\left(\frac{\partial f}{\partial x} \right)_t$, $\left(\frac{\partial f}{\partial u} \right)_t$ et $\left(\frac{\partial f}{\partial p} \right)_t$. Ainsi, δx est solution d'une équation différentielle affine à coefficients dépendant a priori du temps :

$$\frac{d}{dt}(\delta x)(t) = A(t)\delta x + b(t), \quad \delta x(0) = \delta x^0,$$

où $A(t) = \left(\frac{\partial f}{\partial x} \right)_t$ et $b(t) = \left(\frac{\partial f}{\partial u} \right)_t \delta u(t) + \left(\frac{\partial f}{\partial p} \right)_t \delta p$.

Supposons que l'on souhaite calculer la dérivée partielle de x par rapport à p_k , le k -ième paramètre scalaire ($k \in \{1, \dots, r\}$). On note $W(t) \in \mathbb{R}^n$ cette dérivée partielle à l'instant t . Le calcul ci-dessus nous dit que W est la solution du système affine dépendant du temps suivant :

$$\frac{d}{dt}W = \left(\frac{\partial f}{\partial x} \right)_t W + \left(\frac{\partial f}{\partial p_k} \right)_t$$

avec comme condition initiale $W(0) = 0$. De même, pour la dérivée partielle par rapport à x_k^0 notée $V(t) \in \mathbb{R}^n$, on doit résoudre le système linéaire dépendant du temps

$$\frac{d}{dt}V = \left(\frac{\partial f}{\partial x} \right)_t V$$

avec comme condition initiale $V(0) = (\delta_{jk})_{1 \leq j \leq n}$ où ici δ_{jk} est le symbole de Kronecker qui vaut 1 si $j = k$ et 0 sinon.

Noter qu'il n'est pas possible de définir la dérivée partielle de x par rapport à u_k avec un simple vecteur car u_k est une fonction de t . On parle alors de différentielle partielle qui est, à chaque instant, un opérateur linéaire sur des fonctions, une fonctionnelle linéaire donc.

1.1.3 Stabilité locale autour d'un équilibre

La notion de stabilité s'attache à formaliser l'intuition suivante : un point d'équilibre sera dit stable si un petit déséquilibre initial n'entraîne que de petits écarts pour tout temps postérieur; en bref de petites causes n'ont que de petites conséquences.

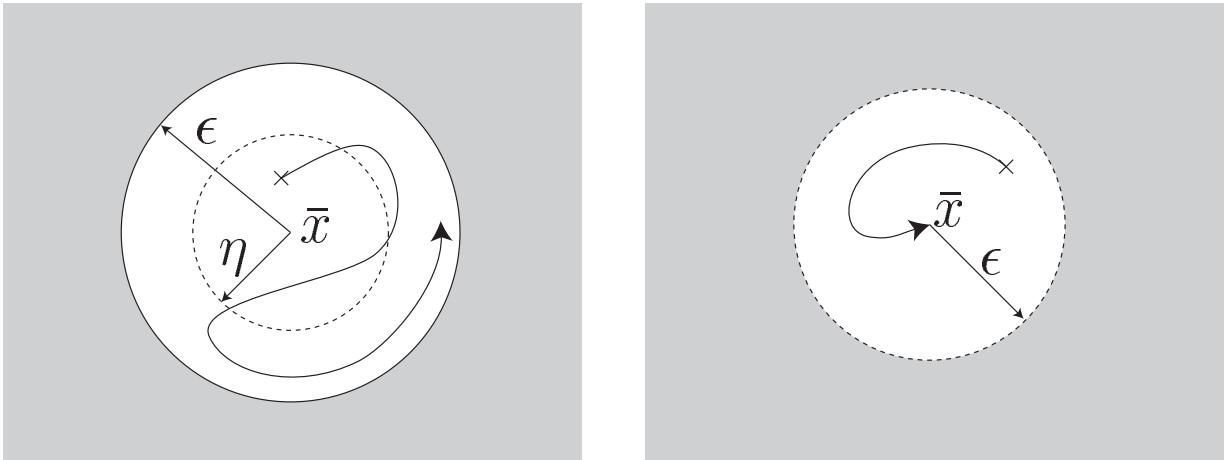


FIGURE 1.3 – Stabilité (gauche) et stabilité asymptotique (droite).

Définition 1 (Stabilité (au sens de Lyapounov) et instabilité). *On reprend les notations et hypothèses du Théorème 1 de Cauchy Lipchitz. On suppose de plus qu'il existe un point d'équilibre $\bar{x} \in \mathbb{R}^n$ caractérisé par $f(\bar{x}, t) = 0$ pour tout $t \in \mathbb{R}$.*

L'équilibre $\bar{x} \in \mathbb{R}^n$ est dit stable (au sens de Lyapounov) si et seulement si pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que pour toute condition initiale x^0 vérifiant $\|x^0 - \bar{x}\| \leq \eta$, la solution de $\frac{d}{dt}x = f(x, t)$ issue de x^0 à $t = 0$, est définie pour tout temps positif et vérifie $\|x(t) - \bar{x}\| \leq \epsilon$ pour tout temps $t \geq 0$. S'il n'est pas stable, il est dit instable.

Définition 2 (Stabilité asymptotique). *Avec les notations de la Définition 1, l'équilibre \bar{x} est dit localement asymptotiquement stable si et seulement s'il est stable et si, de plus, il existe $\eta > 0$ tel que toutes les solutions $x(t)$ de $\frac{d}{dt}x = f(x, t)$, partant en $t = 0$ de conditions initiales x^0 telles que $\|x^0 - \bar{x}\| \leq \eta$, convergent vers \bar{x} lorsque t tend vers $+\infty$.*

Ces notions de stabilité sont illustrées sur la Figure 1.3. Lorsque \bar{x} est asymptotiquement stable, on dit souvent que le système oublie sa condition initiale. En effet, localement, quelle que soit la condition initiale, la trajectoire converge vers \bar{x} . Lorsque, dans la Définition 2, la condition initiale peut être librement choisie, on dit que \bar{x} est *globalement asymptotiquement stable*.

Exemple 2 (Oscillateur harmonique). *Un exemple d'équilibre stable mais non asymptotiquement stable est celui de l'oscillateur harmonique non amorti (le paramètre $\Omega > 0$ est la pulsation)*

$$\frac{d}{dt}x_1 = x_2, \quad \frac{d}{dt}x_2 = -\Omega^2 x_1 \quad (1.8)$$

Le rajout d'un amortissement

$$\frac{d}{dt}x_1 = x_2, \quad \frac{d}{dt}x_2 = -\Omega^2 x_1 - 2\xi\Omega x_2 \quad (1.9)$$

rend alors l'équilibre $(0, 0)$ asymptotiquement stable. Le paramètre sans dimension $\xi > 0$ est le facteur d'amortissement : pour $\xi \in]0, 1[$ le retour à l'équilibre se fait sous la forme d'oscillations amorties ; pour $\xi \geq 1$, le retour à l'équilibre se fait quasiment sans oscillation, i.e., avec un très petit nombre de changements de signe pour x_1 et x_2 . Comme on le verra en détails dans la Section 1.2 consacrée aux systèmes linéaires, cela vient du fait bien connu que les racines s du polynôme caractéristique

$$s^2 + 2\xi\Omega s + \Omega^2 = 0$$

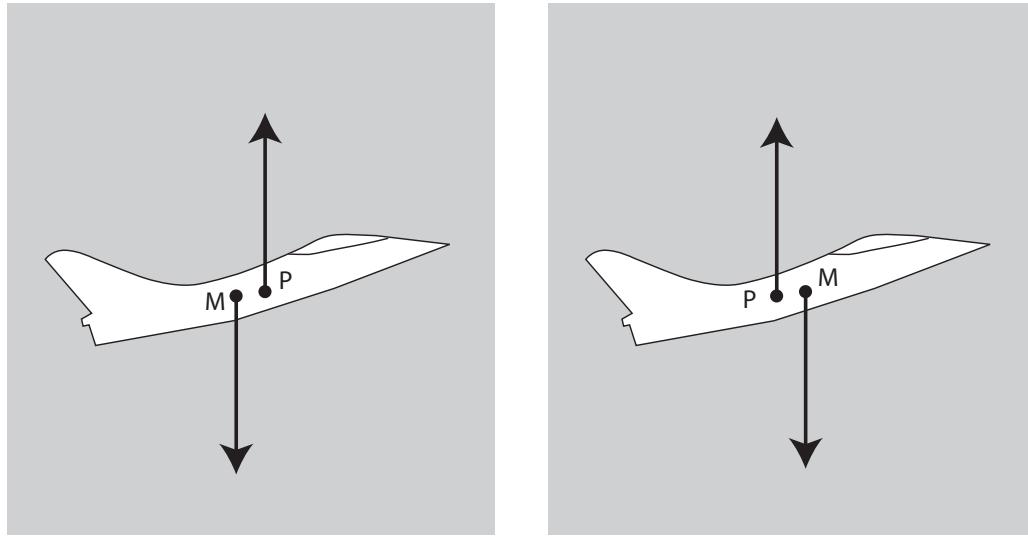


FIGURE 1.4 – Deux configurations possibles d'un avion de combat. Noter la différence de position relative entre le centre de poussée (P) et le centre de masse (M) entre les deux configurations. Le système est stable dans la configuration de droite, et instable dans la configuration de gauche.

de l'équation du second ordre

$$\frac{d^2}{dt^2}x_1 + 2\xi\Omega \frac{d}{dt}x_1 + \Omega^2x_1 = 0$$

qui correspond au système (1.9), sont réelles négatives pour $\xi > 1$ et complexes conjuguées avec une partie réelle négative si $0 < \xi < 1$. Noter enfin qu'aucun des équilibres du double intégrateur, (1.8) avec $\Omega = 0$, n'est stable.

Exemple 3 (Stabilité des avions). *La position relative du centre de masse et du centre de pression aérodynamique conditionnent la stabilité en boucle ouverte des avions. Si l'engin est trop stable, il est difficile à manœuvrer. On doit alors utiliser des surfaces de contrôle (volets) très étendues pour générer les forces requises pour les manœuvres. En régime supersonique, le centre de pression se décale vers l'arrière de l'appareil. Lors de la conception d'un appareil de combat, il est souvent préféré de prévoir un centre de poussée à l'avant du centre de pression lors du régime subsonique, l'avion y est instable en boucle ouverte (notamment pour le décollage et l'atterrissement), alors qu'en régime supersonique il est stable car le centre de pression passe à l'arrière. On se reporterà à l'ouvrage [2] pour de très nombreuses explications et développements sur ce thème.*

Du fait du caractère local de la Définition 2, il est possible, sous certaines conditions, de déduire la stabilité asymptotique des termes du développement limité à l'ordre 1 des équations différentielles autour de l'équilibre (le *système linéarisé tangent*)⁴. Avant d'énoncer un résultat fondamental dans ce cadre (Théorème 10) nous devons d'abord étudier les systèmes linéaires.

4. Ce calcul à l'ordre 1 est même à l'origine de la théorie des matrices et de l'analyse spectrale

1.2 Systèmes dynamiques linéaires

Cette section ne comporte que le strict minimum sur les systèmes linéaires. Pour un exposé complet, nous renvoyons à [34]. Considérons le système linéaire

$$\frac{d}{dt}x(t) = Ax(t) \quad (1.10)$$

avec $x(t) \in \mathbb{R}^n$ et A une matrice $n \times n$ à coefficients réels constants. Une des principales spécificités des systèmes linéaires est que, par homothétie, les propriétés locales sont également globales. C'est en particulier vrai pour la stabilité asymptotique qui peut s'établir, comme nous le verrons au Théorème 4, en étudiant les valeurs propres de A .

1.2.1 L'exponentielle d'une matrice

La matrice dépendant du temps $\exp(tA)$ est définie par la série absolument convergente

$$\exp(tA) = \left[I + tA + \frac{t^2}{2!}A^2 + \dots + \frac{t^k}{k!}A^k + \dots \right] \quad (1.11)$$

où I est la matrice identité. On appelle $t \mapsto \exp(tA)$ l'*exponentielle de la matrice* A . Quel que soit $x^0 \in \mathbb{R}^n$, l'unique solution $x(t)$ du problème de Cauchy $\frac{d}{dt}x(t) = Ax(t)$, $x(0) = x_0$ s'exprime sous la forme

$$x(t) = \exp(tA) x^0$$

Proposition 1 (Propriétés de l'exponentielle de matrice). *L'exponentielle de matrice satisfait les propriétés suivantes*

$$\begin{aligned} \exp(tA) \exp(\tau A) &= \exp((t + \tau)A) \\ \frac{d}{dt}(\exp(tA)) &= \exp(tA) A = A \exp(tA) \\ \exp(PAP^{-1}) &= P \exp(A) P^{-1} \\ \exp(A) &= \lim_{m \rightarrow +\infty} \left(I + \frac{A}{m} \right)^m \\ \det(\exp(A)) &= \exp(\text{tr}(A)) \end{aligned}$$

où t et τ sont des réels, P est une matrice inversible, “ \det ” désigne le déterminant et “ tr ” désigne la trace. En outre, si A et B sont des matrices qui commutent, c.-à-d. $AB = BA$ alors $\exp(A) \exp(B) = \exp(A + B)$.

1.2.2 Forme de Jordan et calcul de l'exponentielle

Le calcul de l'exponentielle de matrice $\exp(tA)$ peut être simplifié en faisant intervenir une transformation P inversible qui diagonalise A , lorsque c'est possible, ou qui transforme A en une matrice diagonale par blocs, dite matrice de Jordan (voir par exemple [34]). Le résultat suivant précise les notations.

Proposition 2 (Réduction de Jordan). *Soit A une matrice $n \times n$ dont le polynôme caractéristique scindé sur \mathbb{C} s'écrit sous la forme*

$$\det(sI - A) = \prod_{i=1}^q (s - \lambda_i)^{\alpha_i}, \quad (\lambda_i \neq \lambda_j, \text{ si } i \neq j)$$

Il existe une matrice inversible T (à coefficients complexes) telle que

$$A = T^{-1}JT \quad \text{où} \quad J = \begin{pmatrix} J_1 & & 0 & & \\ & J_2 & & & \\ & & \ddots & & \\ 0 & & & & J_q \end{pmatrix},$$

avec pour $i = 1, \dots, q$,

$$J_i = \begin{pmatrix} \lambda_i & v_{i,1} & 0 & \dots & 0 \\ 0 & \lambda_i & v_{i,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda_i & v_{i,\alpha_i-1} \\ 0 & \dots & \dots & 0 & \lambda_i \end{pmatrix}$$

où pour tout i, j , $v_{i,j}$ vaut 0 ou 1.

Théorème 4 (CNS de stabilité asymptotique d'un système linéaire stationnaire)

Soit le problème de Cauchy pour le système linéaire stationnaire $\frac{d}{dt}x = Ax$, $x(0) = x^0$ où A est une matrice $n \times n$ et $x^0 \in \mathbb{R}^n$. Ce système possède la propriété que $\lim_{t \rightarrow \infty} x(t) = 0$ quel que soit x^0 (c.-à-d. que le point d'équilibre 0 est *globalement asymptotiquement stable*) si et seulement si toutes les valeurs propres de A ont une partie réelle strictement négative.

Démonstration.

La condition est nécessaire.

Considérons λ valeur propre de A et v un vecteur propre associé. Il vient, pour tout $i \in \mathbb{N}$, $A^i v = \lambda^i v$. On a alors, pour tout $t > 0$,

$$\exp(tA)v = \sum_{i=0}^{\infty} \frac{t^i}{i!} A^i v = \sum_{i=0}^{\infty} (t\lambda)^i v = \exp(\lambda t)v$$

Donc, pour avoir $\lim_{t \rightarrow \infty} x(t) = 0$ quelle que soit la condition initiale, il faut que toutes les valeurs propres aient une partie réelle $\Re(\lambda_i)$ strictement négative.

La condition est suffisante.

Supposons la condition réalisée et notons $\mu = \sup_{i=1, \dots, n} \Re(\lambda_i)$. D'après le Théorème 2, on peut décomposer $A = T^{-1}JT$, avec $J = D + N$ où D est diagonale et N est nilpotente. D'après la Proposition 1, on a

$$\exp(tA) = T^{-1} \exp(tJ)T$$

Les matrice D et N commutent. Intéressons nous donc à

$$\exp(tJ) = \exp(tD) \exp(tN)$$

Par construction, on a $D = \text{diag}(\lambda_i)$ où les λ_i sont les valeurs propres de A . On obtient directement

$$\exp(tD) = \begin{pmatrix} \exp t\lambda_1 & & 0 & & \\ & \exp t\lambda_2 & & & \\ & & \ddots & & \\ 0 & & & & \exp t\lambda_n \end{pmatrix}$$

Notons $\|M\| = \sum_{i=1, \dots, n, j=1, \dots, n} |m_{i,j}|$, pour $M = (m_{i,j})$ matrice $n \times n$. On peut clairement établir, pour tout $t > 0$,

$$\|\exp(tD)\| = \sum_{i=1, \dots, n} |\exp(t\lambda_i)| \leq n \exp(\mu t)$$

D'autre part, N étant nilpotente, on a, pour tout $t > 0$,

$$\exp(tN) = I + tN + \frac{t^2}{2}N^2 + \dots + \frac{t^{n-1}}{(n-1)!}N^{n-1}$$

Donc, on peut majorer

$$\|\exp(tN)\| \leq p(t)$$

par le polynôme $p(t) \triangleq 1 + t\|N\| + \frac{t^2}{2}\|N^2\| + \dots + \frac{t^{n-1}}{(n-1)!}\|N^{n-1}\|$. Il vient alors

$$\|\exp(tJ)\| = \|\exp(tD)\exp(tN)\| \leq \|\exp(tD)\| \|\exp(tN)\| \leq np(t) \exp(\mu t)$$

On en déduit, en utilisant $\|\exp(tA)\| \leq \|T^{-1}\| \|\exp(tJ)\| \|T\|$, que

$$\lim_{t \rightarrow \infty} \|\exp(tA)\| = 0$$

d'où la conclusion. \square

Dans le cas où le système linéaire est effectivement (globalement) asymptotiquement stable, on a en fait *convergence exponentielle* vers zéro de toutes les trajectoires. Le résultat suivant précise ce point, on peut utiliser n'importe quel minorant strict (en valeur absolue) des parties réelles des valeurs propres comme constante σ dans cet énoncé. On trouvera une démonstration dans [12].

Proposition 3 (Estimation de la convergence d'un système linéaire). *Soit le problème de Cauchy pour le système linéaire stationnaire $\frac{d}{dt}x = Ax$, $x(0) = x^0$ où A est une matrice $n \times n$ et $x^0 \in \mathbb{R}^n$. Si toutes les valeurs propres de A sont à partie réelle strictement négative, alors il existe $K > 0$ et $\sigma > 0$ tel que, pour tout $t \geq 0$*

$$\|\exp(tA)\| \leq K \exp(-\sigma t)$$

et donc

$$\|x(t)\| \leq K \|x^0\| \exp(-\sigma t)$$

Théorème 5 (CNS de stabilité d'un système linéaire stationnaire)

Le point d'équilibre $x = 0$ du système linéaire $\frac{d}{dt}x = Ax$ est stable si et seulement si toutes les valeurs propres de A ont une partie réelle négative ou nulle et si toute valeur propre ayant une partie réelle nulle et une multiplicité supérieure ou égale à 2 correspond à un bloc de Jordan d'ordre 1.

L'idée de la démonstration du Théorème 5 repose sur le calcul de l'exponentielle d'une matrice de Jordan d'ordre supérieur à 1. Considérons le cas le plus simple où on a un bloc de Jordan de taille p et d'ordre p

$$J_p = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}$$

Dans ce cas, l'exponentielle de matrice $\exp(tJ_p)$ vaut

$$\exp(tJ_p) = \exp(\lambda t) \begin{pmatrix} 1 & t & \dots & \frac{t^{p-1}}{(p-1)!} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & t \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

Les termes polynomiaux n'apparaissent que dans ce cas. Dans le cas d'une valeur propre à partie réelle nulle, le terme exponentiel ne procure pas d'amortissement et par conséquent, ne domine pas les termes polynomiaux. Ensuite, par les formules de changement de base, on en retrouvera des combinaisons linéaires dans les coordonnées d'origine, prouvant ainsi que les composantes de l'exponentielle de la matrice sont non bornées lorsque t croît. On en déduit l'instabilité du système. Une démonstration détaillée se trouve dans [12]. Une interprétation de ce résultat est que la forme de Jordan d'ordre supérieur à 1 met en évidence un couplage entre les états, qui résultent en une instabilité.

Exemple 4. Considérons une matrice A ayant toutes ses valeurs propres à partie réelle négative ou nulle. Si les valeurs propres à partie réelle nulle sont toutes simples, le système est stable. Le fait que λ valeur propre de A à partie réelle nulle et de multiplicité m corresponde à un bloc de Jordan de taille 1 est équivalent à la condition $\text{rang}(\lambda I - A) = n - m$. On vérifiera ainsi simplement que la matrice A_1 correspond à un système stable alors que la matrice A_2 correspond à un système instable, avec

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

1.2.3 Portraits de phases des systèmes linéaires

Nous allons considérer maintenant les cas les plus intéressants, principalement les cas génériques (i.e. invariants par légères modifications des éléments de la matrice A), qu'on peut rencontrer en dimensions $n = 2$ et $n = 3$.

Dimension $n = 2$

Les principaux cas sont illustrés sur les Figures 1.5, et 1.6. On a noté λ_1 et λ_2 les valeurs propres de A (distinctes ou non, réelles ou complexes conjuguées), ξ_1 et ξ_2 sont des vecteurs propres réels associés quand ils existent. On appelle ici *plan de phases* l'espace \mathbb{R}^n correspondant à l'état x . Pour chaque cas, on a représenté l'emplacement des valeurs propres de A dans le plan complexe, et l'allure générale des trajectoires (notée *portrait de phases*) du système $\frac{d}{dt}x = Ax$. Cette forme générale, comme on l'a vu, ne dépend pas de la condition initiale.

Dimension $n = 3$

La Figure 1.7, montre sur un exemple comment, à partir des portraits de phases en dimension 2, on construit, dans les cas génériques, le portrait de phases en dimension 3 : il suffit de décomposer \mathbb{R}^3 en somme d'espaces propres invariants de dimension 1 ou 2. On a convergence suivant une direction et enroulement avec convergence (typique d'un foyer stable) suivant deux autres directions.

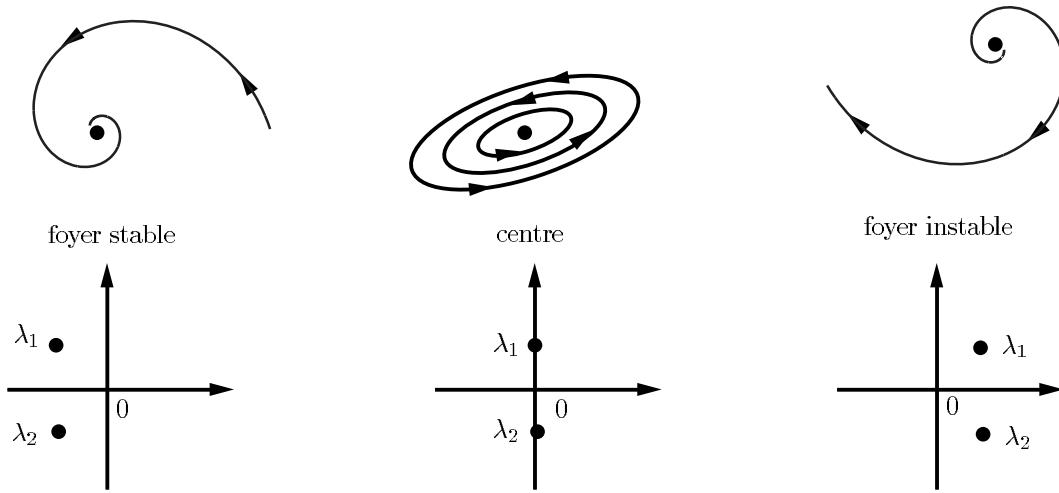


FIGURE 1.5 – *Portraits de phases* plans et linéaires lorsque les valeurs propres de A , λ_1 et λ_2 , ont une partie imaginaire non nulle.

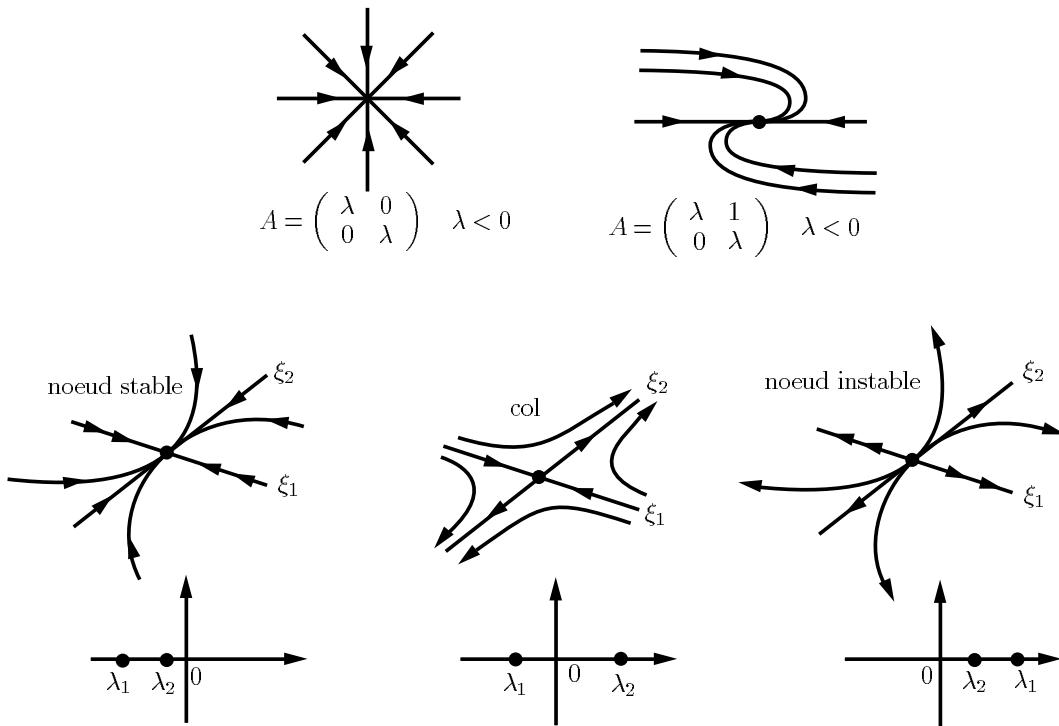


FIGURE 1.6 – *Portraits de phases* plans et linéaires, $\frac{d}{dt}x = Ax$, lorsque les valeurs propres de A , λ_1 et λ_2 , sont réelles (ξ_1 et ξ_2 vecteurs propres de A , lorsqu'ils existent).

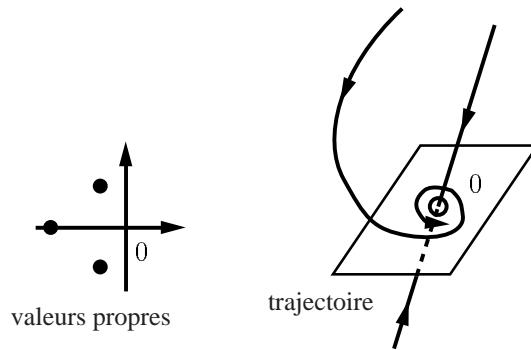


FIGURE 1.7 – Exemple de *portrait de phases* d'un système linéaire de dimension 3 en fonction des valeurs propres de A .

1.2.4 Polynôme caractéristique

Les valeurs propres de A (son spectre) correspondent aux racines de son polynôme caractéristique⁵ $P(s)$

$$P(s) = \det(sI - A) = s^n - \sum_{\nu=0}^{n-1} \sigma_\nu s^\nu = 0$$

À partir des coefficients de la matrice A les coefficients σ_ν de ce polynôme $P(s)$ se calculent simplement.

Exemple 5 (Forme canonique d'un système linéaire). Soit $t \mapsto y(t) \in \mathbb{R}$ solution de

$$y^{(n)} = \sigma_0 y + \sigma_1 y^{(1)} + \dots + \sigma_{n-1} y^{(n-1)}$$

où les σ_i sont des scalaires et où la ν -ième dérivée de y par rapport au temps est notée $y^{(\nu)}$.

En posant $x = (y, y^{(1)}, \dots, y^{(n-1)})^T$ vecteur de \mathbb{R}^n , cette équation scalaire d'ordre n devient un système du premier ordre de dimension n , $\frac{d}{dt}x = Ax$, avec pour A la matrice suivante

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \ddots & 0 & 1 & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ \sigma_0 & \dots & \dots & \dots & \sigma_{n-2} & \sigma_{n-1} \end{pmatrix}$$

Puisque la matrice A a une forme compagnie (aussi appelée forme canonique), son polynôme caractéristique $P(s)$ s'obtient très simplement en partant directement de la forme scalaire d'ordre n . Il suffit de remplacer $\frac{d}{dt}$ par s

$$s^n y = \sigma_0 y + \sigma_1 s y + \dots + \sigma_{n-1} s^{n-1} y$$

La condition pour que cette équation linéaire en y ait des solutions non nulles donne $P(s)$

$$P(s) = s^n - \sigma_0 - \sigma_1 s - \dots - \sigma_{n-1} s^{n-1} = 0$$

5. Ce dernier est obtenu en remplaçant dans $\frac{d}{dt}x = Ax$, l'opérateur $\frac{d}{dt}$ par s la variable de Laplace et en cherchant les conditions sur $s \in \mathbb{C}$ pour le système linéaire de n équations à n inconnues $sx = Ax$ admettre des solutions x non triviales. Ce qui équivaut à dire que la matrice $sI - A$ n'est pas inversible, c'est à dire que son déterminant $P(s)$, un polynôme de degré n , est nul

On dit que la matrice A est *Hurwitz* (stable), lorsque toutes ses valeurs propres sont à partie réelle strictement négative, i.e., les zéros du polynôme $P(s)$ sont dans le demi-plan $\Re(s) < 0$. D'après le Théorème 4, le point d'équilibre 0 est asymptotiquement stable sous cette hypothèse. On dit alors que $P(s)$ est un *polynôme Hurwitz* (stable).

Définition 3 (Polynôme Hurwitz). *Un polynôme à coefficients réels dont toutes les racines résident dans le demi plan complexe ouvert gauche (i.e. sont à partie réelle strictement négative) est un polynôme Hurwitz.*

Les polynômes Hurwitz sont caractérisés par la condition nécessaire et suffisante détaillée dans le théorème suivant.

Théorème 6 (Hermite-Biehler)

Soit $P(s) = a_0 + a_1s + \dots + a_ns^n$ un polynôme de degré n à coefficients réels. On définit $P_p(s^2)$ et $sP_i(s^2)$ les parties paires et impaires de $P(s)$, si bien que $P(s) = P_p(s^2) + sP_i(s^2)$. Ce polynôme est Hurwitz si et seulement si

1. tous les zéros de $w \mapsto P_p(-w^2)$ et de $w \mapsto P_i(-w^2)$ sont réels et distincts
2. a_n et a_{n-1} sont de même signe
3. les racines positives rangées en ordre croissant de $w \mapsto P_i(-w^2)$ (notées w_{i1}, \dots) et les racines positives rangées en ordre croissant de $s \mapsto P_p(-w^2)$ (notées w_{p1}, \dots) satisfont la propriété d'*entrelacement*

$$0 < w_{p1} < w_{e1} < w_{p2} < w_{e2} < \dots$$

Exemple 6. Considérons le polynôme $P(s) = 36 + 34s + 61s^2 + 36s^3 + 29s^4 + 11s^5 + 4s^6 + s^7$. On a alors $P_p(s^2) = 36 + 61s^2 + 29s^4 + 4s^6$ et $P_i(s^2) = 34 + 36s^2 + 11s^4 + s^6$. Les racines positives rangées en ordre croissant des polynômes

$$w \mapsto P_p(-w^2) = 36 - 61w^2 + 29w^4 - 4w^6$$

et

$$w \mapsto P_i(-w^2) = 34 - 36w^2 + 11w^4 - w^6$$

sont $[1 \quad 3/2 \quad 2]$ et $\approx [1.2873 \quad 1.8786 \quad 2.4111]$. Elles satisfont bien la propriété d'*entrelacement*. Tous les zéros de $w \mapsto P_p(-w^2)$ et de $w \mapsto P_i(-w^2)$ sont réels et distincts. Enfin, les coefficients $a_n = 1$ et $a_{n-1} = 4$ sont de même signe. Le polynôme $P(s)$ est donc Hurwitz comme on peut aisément le vérifier numériquement.

De manière générale, ce n'est pas en calculant les racines du polynôme caractéristique qu'on peut vérifier que les parties réelles des valeurs propres sont négatives, mais en analysant ses coefficients. C'est l'objet du critère de Routh explicité dans le Théorème 7. On sait d'ailleurs depuis Galois, qu'il n'existe pas de formule générale utilisant des radicaux donnant les racines d'un polynôme à partir de ses coefficients pour un degré $n \geq 5$.

Théorème 7 (Critère de Routh)

Soit $P(s) = a_0 + a_1s + \dots + a_ns^n$ un polynôme de degré n à coefficients réels. On définit la table de Routh à partir de ces deux premières lignes comme suit

s^n	a_n	a_{n-2}	a_{n-4}	\dots
s^{n-1}	a_{n-1}	a_{n-3}	a_{n-5}	
s^{n-2}	b_{n-1}	b_{n-3}	b_{n-5}	
s^{n-3}	c_{n-1}	c_{n-3}	c_{n-5}	
.	.	.	.	
.	.	.	.	
.	.	.	.	
s^0	g_{n-1}	.	.	

où

$$b_{n-1} = -\frac{1}{a_{n-1}} \begin{vmatrix} a_n & a_{n-2} \\ a_{n-1} & a_{n-3} \end{vmatrix}, \quad b_{n-3} = -\frac{1}{a_{n-1}} \begin{vmatrix} a_n & a_{n-4} \\ a_{n-1} & a_{n-5} \end{vmatrix}, \dots$$

$$c_{n-1} = -\frac{1}{b_{n-1}} \begin{vmatrix} a_{n-1} & a_{n-3} \\ b_{n-1} & b_{n-3} \end{vmatrix} \dots$$

Le nombre de racines de $P(s)$ ayant une partie réelle positive est égal au nombre de changements de signe dans la première colonne de la table de Routh. Le polynôme $P(s)$ est Hurwitz si et seulement si il n'y a pas de changement de signe dans la première colonne de la table de Routh.

On trouvera dans [24, 25] la démonstration de ce résultat. Pour les systèmes d'ordres 2 et 3, $\frac{d}{dt}x = Ax$ de polynôme caractéristique $P(s) = \det(sI - A)$, on en déduit les conditions suivantes de stabilité asymptotique :

— pour $n = 2$ et $P(s) = s^2 - \sigma_1s - \sigma_0$,

$$\sigma_0 = -\det(A) < 0 \quad \text{et} \quad \sigma_1 = \text{tr}(A) < 0. \quad (1.12)$$

— pour $n = 3$ et $P(s) = s^3 - \sigma_2s^2 - \sigma_1s - \sigma_0$,

$$\sigma_0 < 0, \quad \sigma_1 < 0, \quad \sigma_2 < 0 \quad \text{et} \quad -\sigma_0 < \sigma_1\sigma_2. \quad (1.13)$$

1.2.5 Systèmes linéaires instationnaires

La solution générale du système linéaire stationnaire avec terme de forçage $b(t)$

$$\frac{d}{dt}x = Ax + b(t), \quad x \in \mathbb{R}^n, \quad b(t) \in \mathbb{R}^n$$

s'écrit

$$x(t) = \exp(tA)x(0) + \int_0^t \exp((t-\tau)A) b(\tau) d\tau \quad (1.14)$$

Si A dépend du temps t (cas instationnaire), on n'obtient pas, contrairement à ce qu'on pourrait croire, une formule correcte en remplaçant tA et $(t-\tau)A$ par les intégrales $\int_0^t A$ et $\int_\tau^t A$, respectivement. La raison fondamentale est que le produit de deux matrices n'est pas commutatif en général et donc que l'on n'a pas l'identité pourtant fort séduisante suivante : $\frac{d}{dt} \left[\exp \left(\int_0^t A(\tau) d\tau \right) \right] =$

$A(t) \exp\left(\int_0^t A(\tau) d\tau\right)$. C'est cependant vrai si $A(t_1)$ et $A(t_2)$ commutent, c.-à-d. $A(t_1)A(t_2) = A(t_2)A(t_1)$ pour tout couple (t_1, t_2) .

Ainsi cette quadrature instationnaire, fausse en générale pour $n > 1$, n'est valide qu'essentiellement en dimension 1 (où la commutation est trivialement vraie) : la solution générale de l'équation scalaire affine à coefficients variables

$$\frac{d}{dt}x = a(t)x + b(t), \quad x \in \mathbb{R}$$

est

$$x(t) = \exp\left(\int_0^t a(\tau) d\tau\right)x(0) + \int_0^t \exp\left(\int_\tau^t a(\zeta) d\zeta\right)b(\tau) d\tau$$

À partir de la dimension 2, on ne dispose plus de formules explicites et générales pour calculer la solution de $\frac{d}{dt}x = A(t)x + b(t)$, même si $b(t) = 0$. Un exemple est l'équation d'Airy [1] : $\frac{d^2}{dt^2}x = (a+bt)x$ qui n'admet pas de quadrature simple avec des fonctions usuelles (exponentielle, logarithme, ...) et qui définit les fonctions d'Airy, une classe particulière de fonctions spéciales⁶.

Il est faux en général de dire que, si à chaque instant t , $A(t)$ a ses valeurs propres à partie réelle strictement négative, alors les solutions de $\frac{d}{dt}x = A(t)x$ convergent vers 0. Un contre-exemple suffit à s'en convaincre. Considérons

$$A(t) = \begin{pmatrix} -1 + 1.5 \cos^2 t & 1 - 1.5 \sin t \cos t \\ -1 - 1.5 \sin t \cos t & -1 + 1.5 \sin^2 t \end{pmatrix}$$

Pour tout t , les valeurs propres de $A(t)$ sont $-0.25 \pm 0.25\sqrt{7}i$. Or, le système $\frac{d}{dt}x(t) = A(t)x(t)$ a pour solution

$$x(t) = \begin{pmatrix} e^{0.5t} \cos t & e^{-t} \sin t \\ -e^{0.5t} \sin t & e^{-t} \cos t \end{pmatrix} x^0$$

qui, pour des conditions initiales x^0 aussi proches de 0 qu'on le souhaite, diverge lorsque $t \rightarrow +\infty$. Un autre exemple, encore plus simple est fourni par la matrice

$$A(t) = \begin{pmatrix} 0 & 1 \\ -(1+k(t)) & -0.2 \end{pmatrix}$$

correspondant à un système masse-ressort dont le ressort a une raideur variable dans le temps. Si on choisit $k(t) = \cos(2t)/2$, le système devient instable, alors que pour tout temps t fixé, la matrice correspond à un système exponentiellement stable. Une explication intuitive est que si le ressort est raide à la contraction et mou à la dilatation, les déplacements de la masse croissent au lieu de se réduire (voir [65]).

En conclusion, on ne dispose pas de méthode générale pour caractériser, à partir des formules donnant $A(t)$, la stabilité du système différentiel linéaire à coefficients dépendant du temps $\frac{d}{dt}x = A(t)x$, sauf lorsque $\dim(x) = 1$, bien sûr.

1.2.6 Compléments : matrices symétriques et équation de Lyapounov

On donne ici deux résultats utiles qui permettent de caractériser les matrices constantes ayant un polynôme caractéristique Hurwitz.

6. En fait l'immense majorité les fonctions spéciales (fonctions de Bessel, de Jacobi voir [7]) sont solutions d'équations différentielles du second ordre à coefficients polynomiaux en t : elles correspondent donc à des matrices carrées $A(t)$ de dimension 2 dont les coefficients sont simplement des polynômes en t .

Théorème 8 (Sylvester)

Une matrice symétrique de $\mathcal{M}_n(\mathbb{R})$ est définie positive si et seulement si tous ses mineurs principaux sont strictement positifs.

Le théorème suivant est souvent utilisé pour construire une fonction de Lyapounov V (définition 23) d'un système linéaire asymptotiquement stable (P sert pour exhiber $V(x) = x^T P x$)

Théorème 9 (Équation de Lyapounov)

Si A est une matrice *Hurwitz* (as. stable), alors, pour toute matrice Q symétrique définie positive, il existe une matrice symétrique définie positive P solution de l'équation suivante, équation dite de Lyapounov

$$A^T P + P A = -Q \quad (1.15)$$

Réciproquement, s'il existe des matrices symétriques définies positives P et Q telles que (1.15) est vérifiée, alors A est Hurwitz (stable).

1.3 Stabilité des systèmes non linéaires

1.3.1 Étude au premier ordre

Les notions de stabilité de la Section 1.1.3 sont, dans le cadre non linéaire, locales. Autour d'un point d'équilibre, les équations non linéaires de la dynamique sont proches de leur développement limité. Il est naturel de se demander quelle information peut être déduite d'un tel développement au premier ordre.

Comme nous l'avons évoqué à la Section 1.1, un système non linéaire peut avoir plusieurs points d'équilibre isolés. Autour de chacun de ces équilibres, les équations peuvent admettre des systèmes linéarisés tangents très différents et donc localement des propriétés différentes. On pourra se reporter à l'exemple suivant.

Exemple 7. *Le système suivant*

$$\begin{aligned} \frac{d}{dt}x_1 &= -a_1x_1 - x_2x_1 + a_2 \\ \frac{d}{dt}x_2 &= -x_2 + x_1^2 \end{aligned}$$

possède, suivant les valeurs des paramètres a_1, a_2 , un portrait de phase très intéressant. Il est étudié en détail dans [56]. Il s'agit en fait d'un simple système linéaire $\frac{d}{dt}x = -a_1x + u + a_2$ où on choisit une commande proportionnelle $u = -kx$ dont le gain k varie (de manière adaptative⁷) selon l'équation différentielle $\frac{d}{dt}k = -k + x^2$. L'idée est que tant que le x n'a pas convergé vers zéro, le gain k augmente de telle sorte que la commande entraîne effectivement le système vers zéro. La convergence

7. La commande adaptative est un domaine scientifique riche de l'automatique. On pourra se référer à [5].

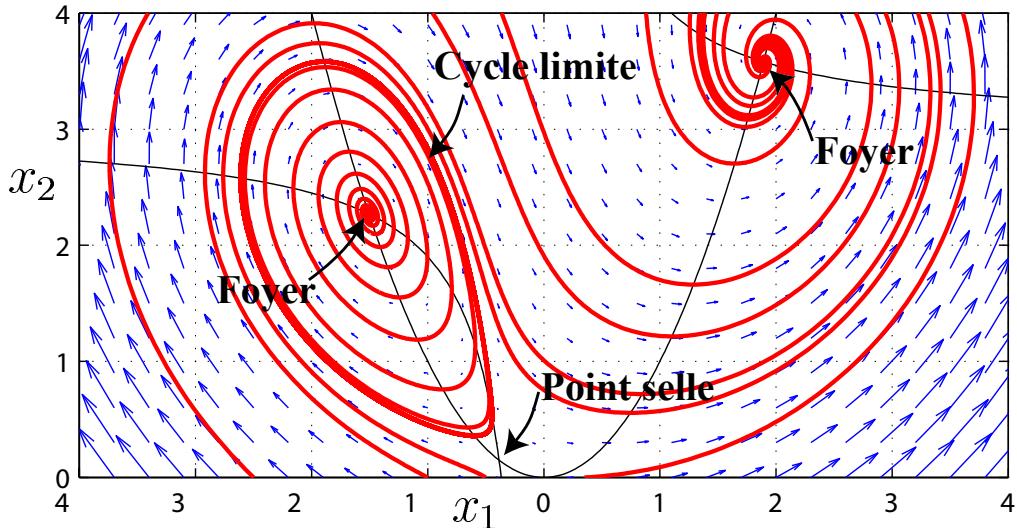


FIGURE 1.8 – Portrait de phase d'un système avec deux foyers stables, un point selle et un cycle limite instable.

espérée de x étant exponentielle, la croissance de k devrait être limitée. On peut montrer que ce n'est pas le cas si le paramètre $a_2 \neq 0$. On a représenté sur la Figure 1.8, le portrait de phase du système pour $a_1 = -3$, $a_2 = 1.1$. Comme cela est montré dans [56], pour ces valeurs, le système possède deux foyers stables, un cycle limite instable et un point selle. Lorsque le système converge, ce n'est pas vers zéro.

D'après le théorème suivant, il est possible de déduire la stabilité asymptotique locale d'un point d'équilibre *hyperbolique* d'après son linéarisé tangent.

Définition 4 (Point d'équilibre hyperbolique). *Un point d'équilibre \bar{x} de l'équation $\frac{d}{dt}x = f(x)$ est dit hyperbolique si les valeurs propres de la matrice Jacobienne*

$$\frac{\partial f}{\partial x}(\bar{x}) = \left(\frac{\partial f_i}{\partial x_j} \right)_{1 \leq i,j \leq n}$$

sont toutes à partie réelle non nulle.

Théorème 10 (Stabilité asymptotique locale d'un point d'équilibre hyperbolique)

Première méthode de Lyapounov. Soit $\mathbb{R}^n \ni x \mapsto f(x) \in \mathbb{R}^n$ continûment dérivable par rapport à x et $\bar{x} \in \mathbb{R}^n$ tel que $f(\bar{x}) = 0$. Le point d'équilibre \bar{x} de

$$\frac{d}{dt}x = f(x)$$

est *localement asymptotiquement stable* (c.f. Définition 1) si les valeurs propres de la matrice Jacobienne

$$\frac{\partial f}{\partial x}(\bar{x}) = \left(\frac{\partial f_i}{\partial x_j} \right)_{1 \leq i,j \leq n}$$

sont toutes à partie réelle strictement négative. Le point d'équilibre \bar{x} est *instable au sens de Lyapounov* si au moins l'une des valeurs propres de la matrice Jacobienne $\frac{\partial f}{\partial x}(\bar{x})$ est à partie réelle strictement positive.

Il est important de bien comprendre ce résultat. On commence par écrire un développement limité autour de \bar{x}

$$\frac{d}{dt}x = \frac{d}{dt}(x - \bar{x}) = f(x) = f(\bar{x}) + \frac{\partial f}{\partial x}(\bar{x})(x - \bar{x}) + \epsilon(x - \bar{x})$$

où $\epsilon(x - \bar{x})$ rassemble les termes d'ordres supérieurs. On ne considère que les termes d'ordre 1 en $\xi = x - \bar{x}$ pour en déduire le système différentiel linéaire à coefficients constants (A ne dépend pas du temps) suivant

$$\frac{d}{dt}\xi = A\xi$$

avec $\xi(t) \in \mathbb{R}^n$ et $A = \frac{\partial f}{\partial x}(\bar{x})$ la matrice Jacobienne de f en \bar{x} . On sait, d'après le Théorème 4, que toutes ses solutions convergent vers 0 lorsque les valeurs propres de A sont toutes à partie réelle strictement négative. Le Théorème 10 énonce que, si le système linéaire du premier ordre est asymptotiquement stable, alors les termes d'ordre supérieur ne peuvent pas être déstabilisateurs pour des conditions initiales proches de l'équilibre. La preuve de ce résultat est reproduite un peu plus loin. Elle utilise la seconde méthode de Lyapounov, un autre outil pour montrer la stabilité d'un point d'équilibre qui s'inspire la notion d'énergie et de dissipation. Nous développons cette méthode maintenant.

1.3.2 Fonctions de Lyapounov

À titre d'introduction, reprenons l'exemple 2 de l'oscillateur harmonique amorti et considérons la fonction

$$V(x_1, x_2) = \frac{\Omega^2}{2}(x_1)^2 + \frac{1}{2}(x_2)^2$$

qui n'est autre que son énergie totale. Un calcul simple montre que pour $t \mapsto (x_1(t), x_2(t))$ solution de (1.9), on a

$$\frac{d}{dt}(V(x_1(t), x_2(t))) = \frac{\partial V}{\partial x_1} \frac{d}{dt}x_1 + \frac{\partial V}{\partial x_2} \frac{d}{dt}x_2$$

Or $\frac{d}{dt}x_1 = x_2$ et $\frac{d}{dt}x_2 = -2\xi\Omega x_2 - \Omega^2 x_1$. Donc

$$\frac{d}{dt}V = -2\xi\Omega(x_2)^2 \leq 0$$

Ainsi, $t \mapsto V(x_1(t), x_2(t))$ est une fonction décroissante, comme elle est positive, elle converge vers une valeur positive. Il est intuitif de penser que $\frac{d}{dt}V$ converge vers 0 et donc que x_2 converge vers 0. Mais si x_2 converge vers 0, il est aussi intuitif de penser que sa dérivée converge aussi vers 0. Or, $\frac{d}{dt}x_2 = -2\xi\Omega x_2 - \Omega^2 x_1$ et donc x_1 tend vers 0 aussi. Le raisonnement ci-dessus n'est pas très rigoureux mais il est correct car on a affaire à des fonctions du temps V, x_1, x_2, \dots uniformément continues, c.-à-d. dont le module de continuité en t ne dépend pas de t ⁸. Cette continuité uniforme vient du fait que, comme V est décroissante le long des trajectoires et que V est infini quand x_1 ou x_2 tend vers l'infini, les trajectoires sont nécessairement bornées et donc sont des fonctions uniformément continues du temps. En effet, leurs dérivées en temps sont uniformément bornées. Or, un lemme classique d'analyse (lemme de Barbalat) dit que, si l'intégrale $\int_0^t h(s)ds$ d'une fonction uniformément continue h de $[0, +\infty[$ dans \mathbb{R} est convergente pour t tendant vers l'infini, alors cette fonction h admet 0 comme limite en $t = +\infty$. Comme $V(t) - V(0) = \int_0^t \frac{d}{dt}V$ on sait que $\frac{d}{dt}V$ converge vers 0 et comme $\frac{d}{dt}V(t) - \frac{d}{dt}V(0) = \int_0^t \frac{d^2}{dt^2}V$, $\frac{d^2}{dt^2}V$ converge aussi vers 0. Ce qui nous donne que x_2 et x_1 convergent vers 0. En fait, nous aurions pu faire le raisonnement à l'identique pour un système général pour peu que nous disposions d'une fonction V , infinie à l'infini (on dira *non bornée radialement*), bornée inférieurement et décroissante le long de toute trajectoire.

Il est remarquable que dans ces calculs nous n'ayons eu besoin d'expliciter la loi horaire des trajectoires $t \mapsto x_1(t)$ et $t \mapsto x_2(t)$. C'est la force des fonctions V dites de Lyapounov : elles fournissent des informations précieuses sur les solutions d'équations différentielles sans requérir la connaissance précise des lois horaires. Un lecteur ayant compris ce qui précède n'aura pas de difficulté à saisir l'utilité de la définition ainsi que l'intérêt du résultat ci-dessous.

Définition 5 (Fonction de Lyapunov). *Soit le système dynamique $\frac{d}{dt}x = f(x)$ défini dans un domaine $\Omega \subset \mathbb{R}^n$ (*f Lipschitz*) et soit une fonction continûment dérivable V de Ω dans \mathbb{R} et bornée inférieurement. On dit que V est une fonction de Lyapounov si elle est décroissante (au sens large) le long des trajectoires, i.e.,*

$$\frac{d}{dt}V(x) = \nabla V(x)f(x) \leq 0$$

pour tout $x \in \Omega$.

Ainsi, une intégrale première $I : \Omega \mapsto \mathbb{R}$ est une fonction de Lyapounov si elle est bornée inférieurement car $\frac{d}{dt}I = 0$.

Définition 6 (Ensemble positivement invariant). *On dit que $E \subset \Omega$ est un sous-ensemble positivement invariant de $\frac{d}{dt}x = f(x)$, si quel que soit $x^0 \in E$ condition initiale du problème de Cauchy $\frac{d}{dt}x = f(x)$, $x(0) = x^0$, tous les points de la trajectoire $[0, +\infty[\ni t \mapsto x(t)$ appartiennent à E .*

8. Une fonction $\mathbb{R} \ni t \mapsto h(t) \in \mathbb{R}^n$ est dite uniformément continue si pour tout $\epsilon > 0$, il existe $\eta > 0$ indépendant de t , tel que $\|h(t + \epsilon) - h(t)\| \leq \eta$ pour tout $t \in \mathbb{R}$.

Théorème 11 (Invariance de LaSalle)

Soit $\Omega \subset \mathbb{R}^n$ un ouvert non vide. $\Omega \ni x \mapsto f(x) \in \mathbb{R}^n$ une fonction continûment dérivable de x . Soit $\Omega \ni x \mapsto V(x) \in \mathbb{R}$ une fonction réelle continûment dérivable de x . On suppose

1. qu'il existe $c \in \mathbb{R}$ tel que le sous-ensemble $E_c = \{x \in \Omega \mid V(x) \leq c\}$ de \mathbb{R}^n soit un compact (fermé borné dans \mathbb{R}^n) non vide.
2. que V décroît le long des trajectoires de $\frac{d}{dt}x = f(x)$, c.-à-d., pour tout $x \in E_c$,

$$\frac{d}{dt}V(x) = \nabla V(x) \cdot f(x) = \sum_{i=1}^n \frac{\partial V}{\partial x_i}(x) f_i(x) \leq 0$$

Alors, pour toute condition initiale $x^0 \in E_c$, la solution de $\frac{d}{dt}x = f(x)$ reste dans E_c , est définie pour tout temps $t > 0$ (pas d'explosion en temps fini) et converge vers la réunion de tous les sous-ensembles positivement invariants inclus dans

$$\left\{ x \in E_c \mid \frac{d}{dt}V(x) = 0 \right\}.$$

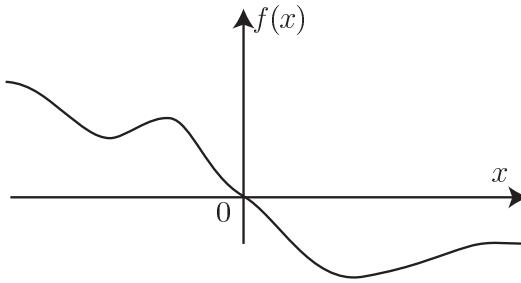
Cette réunion d'ensembles positivement invariants dont il est question dans le théorème 11 est aussi positivement invariant. On l'appelle *ensemble invariant de LaSalle*. Il est caractérisé par le système *sur-déterminé* de $n + 1$ équations

$$\begin{aligned} \frac{d}{dt}\xi_1 &= f_1(\xi) \\ &\vdots \\ \frac{d}{dt}\xi_n &= f_n(\xi) \\ \sum_{i=1}^n \frac{\partial V}{\partial x_i}(\xi) f_i(\xi) &= 0 \end{aligned}$$

et n inconnues $(\xi_1(t), \dots, \xi_n(t)) \in E_c$ qui sont des fonctions continûment dérivables de t . Pour résoudre ce système, i.e., obtenir les équations de l'ensemble invariant de LaSalle, il suffit de dériver un certain nombre de fois la dernière équation et de remplacer, à chaque nouvelle dérivation en temps, les $\frac{d}{dt}\xi_i$ par $f_i(\xi)$. On obtient ainsi une famille d'équations portant uniquement sur ξ qui caractérise cet ensemble limite. Il attire les trajectoires initialisées dans E_c : c'est un *attracteur*.

Dans le cas où $\bar{x} \in E_c$ est un point d'équilibre, $f(\bar{x}) = 0$, il est clair que $\xi(t) = \bar{x}$ est une solution du système sur-déterminé précédent et donc \bar{x} appartient à cet ensemble limite. Si, maintenant, \bar{x} est l'unique solution de ce système sur-déterminé, alors on est sûr que les trajectoires du système qui démarrent dans E_c convergent toutes vers \bar{x} . Des variantes du théorème ci-dessus spécifiques à l'étude de la stabilité des points d'équilibre sont données dans l'Annexe B.

Si maintenant, on suppose dans le théorème 11 que $\Omega = \mathbb{R}^n$ et que le paramètre $c \in \mathbb{R}$ peut être choisi aussi grand que l'on veut, alors d'une part, V tend vers $+\infty$ quand $\|x\|$ tend vers l'infini c.-à-d. que V est non bornée radialement et, d'autre part, toute trajectoire $x(t)$ est bornée pour les temps $t > 0$ car contenue dans E_c avec $c = V(x(0))$. Enfin, si on suppose en plus que le point d'équilibre \bar{x} est

FIGURE 1.9 – Forme générale de la fonction f .

l’unique solution $\xi = \bar{x}$ du système sur-déterminé ci-dessus, alors toutes les trajectoires convergent vers l’équilibre \bar{x} qui est dit *globalement asymptotiquement stable*.

Exemple 8. Considérons un système scalaire, $\dim(x) = 1$, de la forme

$$\frac{d}{dt}x = f(x(t))$$

où f (fonction Lipschitz) satisfait les deux propriétés suivantes :

1. pour tout $x \neq 0$ on a $xf(x) < 0$
2. $\int_0^{+\infty} f(x)dx = -\infty$, et $\int_{-\infty}^0 f(x)dx = +\infty$

La forme générale de f est donnée sur la Figure 1.9. Une fonction candidate à être de Lyapounov est

$$V(x) = - \int_0^x f(\tau)d\tau$$

Par construction, d’après le point 1, V est continue et possède une dérivée continue. On a

$$\frac{d}{dt}V(x) = -f(x)^2$$

Donc, pour tout x , $\frac{d}{dt}V(x) \leq 0$. Enfin, d’après le point 2, $V(x) \rightarrow +\infty$ lorsque $|x| \rightarrow +\infty$. D’après le Théorème 11, on peut prendre pour c n’importe quelle valeur positive. Comme $\frac{d}{dt}V = 0$ implique $x = 0$, on en déduit que 0 est globalement asymptotiquement stable. À titre d’exemple, on obtient, par ce raisonnement, la stabilité asymptotique de l’origine pour le système

$$\frac{d}{dt}x = x - \sinh x$$

alors qu’en considérant le linéarisé tangent autour de 0, on ne peut pas conclure (le point d’équilibre 0 n’étant pas hyperbolique).

Montrons en quoi la preuve du Théorème 10 repose sur le Théorème 11 d’invariance de LaSalle. Avec les notations du Théorème 10, considérons un point d’équilibre \bar{x} . On suppose que les valeurs propres du linéaire tangent, c.-à-d. les valeurs propres de la matrice Jacobienne $A = \left(\frac{\partial f_i}{\partial x_j}(\bar{x}) \right)_{1 \leq i,j \leq n}$ sont toutes à partie réelle strictement négative. Il nous faut construire une fonction de Lyapounov \tilde{V} autour de \bar{x} . On sait que $\exp(tA)$ est une matrice qui converge exponentiellement vers 0 quand t tend vers $+\infty$ (voir Proposition 3). On pose

$$P = \int_0^{+\infty} \exp(sA^T) \exp(sA) ds$$

où A^T est la transposée de A . Par cette construction, P est une matrice symétrique définie positive. On pose

$$V(x) = (x - \bar{x})^T P(x - \bar{x}).$$

Ainsi on a

$$V(x) = \int_0^{+\infty} [\exp(sA)(x - \bar{x})]^T \exp(sA)(x - \bar{x}) ds$$

On a en dérivant par rapport au temps sous le signe somme

$$\begin{aligned} \frac{d}{dt} V(x) &= \\ &\int_0^{+\infty} \left([\exp(sA)(x - \bar{x})]^T \exp(sA)f(x) + [\exp(sA)f(x)]^T \exp(sA)(x - \bar{x}) \right) ds \end{aligned}$$

car $\frac{d}{dt}x = f(x)$. Comme $f(x) = A(x - \bar{x}) + o(\|x - \bar{x}\|)$, on a

$$\begin{aligned} \frac{d}{dt} V(x) &= \int_0^{+\infty} [\exp(sA)(x - \bar{x})]^T \exp(sA)A(x - \bar{x}) ds \\ &\quad + \int_0^{+\infty} [\exp(sA)A(x - \bar{x})]^T \exp(sA)(x - \bar{x}) ds + o(\|x - \bar{x}\|^2) \end{aligned}$$

Ainsi,

$$\frac{d}{dt} V(x) = (x - \bar{x})^T Q(x - \bar{x}) + o(\|x - \bar{x}\|^2)$$

avec la matrice symétrique Q , définie par l'intégrale

$$Q = \int_0^{+\infty} \left([\exp(sA)A]^T \exp(sA) + [\exp(sA)]^T \exp(sA)A \right) ds$$

Comme $\frac{d}{ds} \exp(sA) = \exp(sA)A$ (voir Proposition 1), on a

$$\frac{d}{ds} [(\exp(sA))^T \exp(sA)] = [\exp(sA)A]^T \exp(sA) + [\exp(sA)]^T \exp(sA)A$$

et donc l'intégrale définissant Q se calcule explicitement, pour donner simplement

$$Q = [(\exp(sA))^T \exp(sA)]_{s=0}^{s=+\infty} = -I$$

où I est la matrice identité de taille n . Ainsi, on a

$$\frac{d}{dt} V(x) = -\|x - \bar{x}\|^2 + o(\|x - \bar{x}\|^2)$$

Donc $\frac{d}{dt} V \leq 0$ pour x assez proche de \bar{x} et ne s'annule qu'en $x = \bar{x}$. Le Théorème 11 avec $c > 0$ assez petit permet alors de conclure : le point d'équilibre \bar{x} est localement asymptotiquement stable.

Ainsi, lorsque les valeurs propres au point d'équilibre \bar{x} de $\frac{d}{dt}x = f(x)$ sont toutes à partie réelle négative, des petits écarts à l'équilibre sont naturellement amortis : le système oublie sa condition initiale et converge vers \bar{x} . Le système n'est que transitoirement affecté par une petite perturbation de conditions initiales.

1.3.3 Robustesse paramétrique

Imaginons que les équations dépendent en fait de paramètres notés $p = (p_1, \dots, p_r) \in \mathbb{R}^r$ plus ou moins bien connus : $\frac{d}{dt}x = f(x, p)$, avec f fonction continûment dérivable par rapport à x et p . On suppose que, pour une valeur nominale du paramètre \bar{p} , le système admet un point d'équilibre \bar{x} , $f(\bar{x}, \bar{p}) = 0$ et que les valeurs propres du *système linéarisé tangent* en ce point \bar{x} sont toutes à partie réelle strictement négative. Qu'en est-il des systèmes voisins obtenus pour des valeurs de p proche de \bar{p} ? En fait, pour p proche de \bar{p} , $\frac{d}{dt}x = f(x, p)$ admet aussi un point d'équilibre $\phi(p)$ qui dépend continûment de p , $f(\phi(p), p)$, $\phi(\bar{p}) = \bar{x}$. De plus, les valeurs propres en $\phi(p)$ sont toutes à partie réelle strictement négative. Ainsi, qualitativement, le système ne change pas de comportement lorsqu'on bouge un peu les paramètres : localement autour de \bar{x} , les trajectoires convergent vers un point d'équilibre unique qui reste proche de \bar{x} . La situation est dite *robuste* au sens où elle ne change que très peu lorsque l'on bouge un peu la condition initiale et les paramètres.

L'existence de $\phi(p)$ résulte en fait du théorème des fonctions implicites. Le point d'équilibre est défini implicitement par $f(x, p) = 0$; pour $p = \bar{p}$ on a une solution \bar{x} ; pour cette solution \bar{x} , la matrice Jacobienne $\frac{\partial f_i}{\partial x_j}(\bar{x}, \bar{p})$ est inversible (pas de valeur propre nulle puisqu'elles sont toutes par hypothèse à partie réelle strictement négative); f dépend continûment de p . Le fait que les valeurs propres en $\phi(p)$ restent à partie réelle strictement négative, vient du fait que les valeurs propres d'une matrice dépendent de façon continue de ses coefficients⁹.

1.3.4 Compléments : caractère intrinsèque des valeurs propres du système linéarisé tangent

Les valeurs propres du linéaire tangent associé à un équilibre sont des nombres intrinsèques, i.e. ils ne dépendent pas des systèmes de coordonnées x . Lorsque qu'on pose $\frac{d}{dt}x = f(x)$, on considère un jeu particulier de variables pour écrire les équations du système. Ici, on a choisi, pour représenter notre système les variables $(x_1, \dots, x_n) \in \mathbb{R}^n$. Ce choix est clairement arbitraire : on pourrait tout aussi bien prendre d'autres variables, par exemple les variables $z = (z_1, \dots, z_n)$ définies par une correspondance bi-univoque avec les variables x : $z = \psi(x)$ et $x = \chi(z)$ où les fonctions ψ et χ sont des applications inverses l'une de l'autre. Par exemple, pour repérer un point dans le plan on doit définir un jeu de deux nombres. On peut prendre les coordonnées cartésiennes avec l'abscisse et l'ordonnée mais on peut aussi prendre les coordonnées polaires avec un angle et la distance à l'origine. Il est évident que les résultats de nature qualitative ne doivent pas dépendre du choix des variables. C'est bien le cas de la propriété de stabilité.

Si dans les variables x , la dynamique s'écrit $\frac{d}{dt}x = f(x)$ et qu'elle admet un point d'équilibre \bar{x} asymptotiquement stable, alors dans les variables z , les équations vont bien-sûr changer, mais le point d'équilibre \bar{x} aura son correspondant \bar{z} via les transformations inversibles ψ et χ , et \bar{z} sera bien-sûr lui aussi asymptotiquement stable. Supposons les transformations ψ et ϕ régulières (difféomorphismes continûment dérivables). Alors, les équations dans les variables z se déduisent de celles dans les variables x par un simple calcul de fonctions composées

$$\frac{d}{dt}x = \frac{d}{dt}\chi(z) = \left(\frac{\partial \chi}{\partial z} \right) \frac{d}{dt}z = f(\chi(z))$$

9. On n'a pas en général de dépendance plus régulière que C^0 : prendre par exemple $\begin{pmatrix} 0 & 1 \\ p & 0 \end{pmatrix}$ avec p proche de 0; les valeurs propres sont $\pm\sqrt{p}$.

Ainsi,

$$\frac{d}{dt}z = g(z) = \left[\frac{\partial \chi}{\partial z} \right]^{-1} f(\chi(z))$$

Comme $f(\bar{x}) = 0, g(\bar{z}) = 0$. Un calcul montre alors que la matrice Jacobienne de f en \bar{x} est semblable à celle de g en \bar{z}

$$\frac{\partial \chi}{\partial z}(\bar{z}) \frac{\partial g}{\partial z}(\bar{z}) = \frac{\partial f}{\partial x}(\bar{x}) \frac{\partial \chi}{\partial z}(\bar{z})$$

car $\frac{\partial \chi}{\partial z}(\bar{z})$ est une matrice inversible. Il suffit de dériver par rapport à z la relation

$$\frac{\partial \chi}{\partial z}(z) g(z) = f(\chi(z))$$

et de remarquer, puisque $g(\bar{z}) = 0$, qu'il n'est pas utile de calculer la dérivée seconde $\frac{\partial^2 \chi}{\partial z^2}$ car elle est en facteur de $g(\bar{z})$. Ainsi les valeurs propres obtenues avec les variables x sont les mêmes que celles obtenues avec les variables z . Le spectre de la matrice du linéaire tangent autour d'un point d'équilibre est indépendant du choix des variables que l'on choisit pour faire les calculs. Ces valeurs propres ont pour unité l'inverse d'un temps : ce sont des indicateurs intrinsèques caractérisant les échelles du système. On les appelle aussi exposants caractéristiques.

1.3.5 Compléments : les systèmes dynamiques dans le plan

Dans bien des cas pratiques, on se trouve confronté à des systèmes de dimension 2. Il est notable que, dans ces cas, on peut établir un certain nombre de résultats théoriques très informatifs. Nous les présentons dans ce qui suit.

Théorème 12 (Poincaré)

Le système autonome plan $\frac{d}{dt}x = f(x)$ avec $x \in \mathbb{R}^2$, ne peut admettre que quelques types de régimes asymptotiques possibles. Étant donnée une condition initiale $x^0 \in \mathbb{R}^2$, considérons $t \mapsto x(t)$ la solution de $\frac{d}{dt}x = f(x)$ qui démarre en $t = 0$ en x^0 . Alors, pour des temps $t \geq 0$, on ne peut avoir que les cas de figure suivants (voir illustrations sur la Figure 1.10) :

1. Si $x(t)$ ne reste pas borné, alors soit $x(t)$ explose en temps fini, soit $x(t)$ n'explose pas en temps fini mais alors $x(t)$ est défini pour tout $t > 0$ et $\lim_{t \rightarrow +\infty} \|x(t)\| = +\infty$. En résumé : *si la trajectoire n'est pas bornée pour les temps positifs, elle converge vers l'infini en temps fini ou en temps infini.*
2. Si $x(t)$ reste borné pour les temps positifs alors elle est définie pour tout temps $t > 0$ et on distingue trois cas :
 - (a) soit $x(t)$ converge vers un point d'équilibre (en temps infini si x^0 n'est pas un point d'équilibre)
 - (b) soit $x(t)$ converge vers une trajectoire périodique (*cycle limite*)
 - (c) soit $x(t)$ s'enroule autour d'une courbe fermée du plan formée de trajectoires qui partent en $t = -\infty$ d'un point d'équilibre et qui arrive en $t = +\infty$ vers, a priori, un autre point d'équilibre (orbite *hétérocline* si les deux points d'équilibres sont différents et orbite *homocline* si les deux points d'équilibres sont identiques).

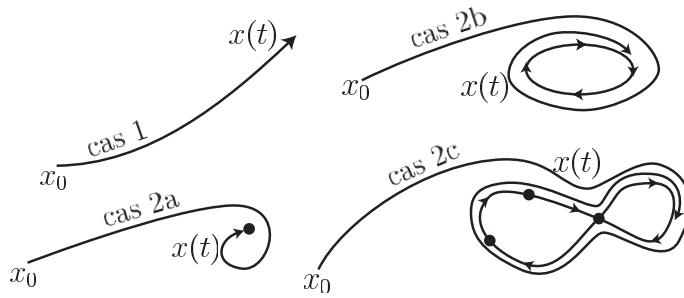


FIGURE 1.10 – Les quatre comportements asymptotiques possibles pour une trajectoire d'un système dynamique autonome défini dans le plan.

En résumé, lorsque la trajectoire reste bornée elle converge soit vers un point soit vers une courbe fermée du plan, courbe qui est tangente au *champ de vecteurs* $f(x)$. Ainsi en dimension 2, on ne peut pas avoir de comportements asymptotiques très compliqués : il est d'usage de dire qu'en dimension deux, il ne peut pas y avoir de chaos. Il faut bien comprendre que cela n'est vrai que pour le plan (sur le tore $\mathbb{S}^1 \times \mathbb{S}^1$ ce n'est plus vrai, les trajectoires peuvent être partout denses), et que pour les systèmes continus autonomes, i.e., définis par deux équations différentielles scalaires ne dépendant pas du temps. L'argument essentiel de démonstration est l'unicité des trajectoires. Dans le plan deux trajectoires ne peuvent se couper sans être confondues.

Si l'on rajoute par exemple une dépendance périodique en temps $f(x, t) \equiv f(x, t + 2\pi)$ alors ce n'est plus vrai. Le système autonome sous-jacent est de dimension 3 : $(x_1, x_2, \theta) \in \mathbb{R}^2 \times \mathbb{S}^1$ avec

$$\frac{d}{dt}x_1 = f_1(x_1, x_2, \theta), \quad \frac{d}{dt}x_2 = f_2(x_1, x_2, \theta), \quad \frac{d}{dt}\theta = 1.$$

Ce n'est plus vrai non plus avec les systèmes discrets même de dimension un. Par exemple les solutions de l'équation logistique $x_{k+1} = 4x_k(1 - x_k)$ qui démarrent en $x_0 \in [0, 1]$ restent toujours dans $[0, 1]$ mais elles sont partout denses dans $[0, 1]$. Pour comprendre ce phénomène il faut étudier des itérées de la formule de récurrence. On a représenté les applications correspondantes sur la Figure 1.11. Les régimes asymptotiques pour les indices k grands sont complexes.

Sous une hypothèse supplémentaire, on peut spécifier la nature de la limite.

Théorème 13 (Critère de Bendixon)

Soit $\mathbb{R}^2 \ni x \mapsto f(x) \in \mathbb{R}^2$ une fonction continue et dérivable. On suppose que $\text{div}(f)(x) = \frac{\partial f_1}{\partial x_1}(x) + \frac{\partial f_2}{\partial x_2}(x) < 0$ pour presque tout $x \in \mathbb{R}^2$. Soit $t \mapsto x(t)$ une solution de $\frac{d}{dt}x = f(x)$ qui reste bornée pour les temps t positifs. Alors, sa limite quand t tend vers $+\infty$ est un point d'équilibre, i.e., une solution $\bar{x} \in \mathbb{R}^2$ de $f(\bar{x}) = 0$.

Ce résultat n'est plus du tout vrai en dimension supérieure à deux. Il suffit de prendre le système chaotique de Lorenz (représenté sur la Figure 1.12) de l'exemple suivant.

Exemple 9 (Système de Lorenz).

$$\begin{cases} \frac{dx_1}{dt} = s(-x_1 + x_2) \\ \frac{dx_2}{dt} = r x_1 - x_2 - x_1 x_3 \\ \frac{dx_3}{dt} = -bx_3 + x_1 x_2, \end{cases}$$

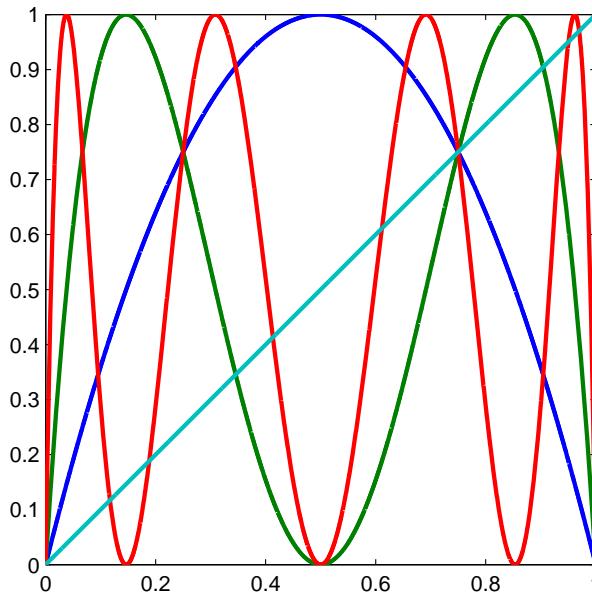


FIGURE 1.11 – Première bissectrice et graphes de la fonction logistique $[0, 1] \ni x \mapsto f(x) = 4x(1 - x) \in [0, 1]$ et de ses deux premières itérées, $f \circ f$ et $f \circ f \circ f$.

avec $s = 10$, $r = 28$ et $b = 8/3$. Toutes les trajectoires sont bornées, la divergence du champ de vecteurs, $-s - 1 - b < 0$, et les trajectoires ont des comportements asymptotiques complexes et encore aujourd’hui assez mal compris.

De la caractérisation des régimes asymptotiques illustrés sur la Figure 1.10, on peut montrer directement le résultat suivant, résultat utile pour montrer l’existence d’une *orbite périodique* (cycle limite) tel qu’illustrée sur la Figure 1.13.

Théorème 14 (Poincaré-Bendixon)

Soit $\mathbb{R}^2 \ni x = f(x) \in \mathbb{R}^2$ une fonction de classe C^1 . On considère le système dynamique $\frac{dx}{dt} = f(x)$. On suppose qu’il existe dans le plan un ensemble compact Ω tel que

- toute trajectoire ayant sa condition initiale dans Ω reste dans Ω pour les temps $t > 0$ (Ω est positivement invariant).
- soit Ω ne contient aucun point d’équilibre, soit Ω contient un unique point d’équilibre dont toutes les valeurs propres sont à partie réelle strictement positive.

alors Ω contient une orbite périodique (cycle limite).

L’idée derrière cet énoncé est que les trajectoires bornées dans le plan doivent approcher des orbites périodiques ou des points d’équilibre lorsque le temps tend vers l’infini. Si Ω ne contient aucun point d’équilibre alors il doit contenir une orbite périodique. Si Ω contient un unique *point d’équilibre hyperbolique* instable dans toutes les directions, alors au voisinage de ce point, les trajectoires sont repoussées par ce point. Il est alors possible de définir une courbe fermée permettant d’exclure ce point et de se ramener au cas précédent.

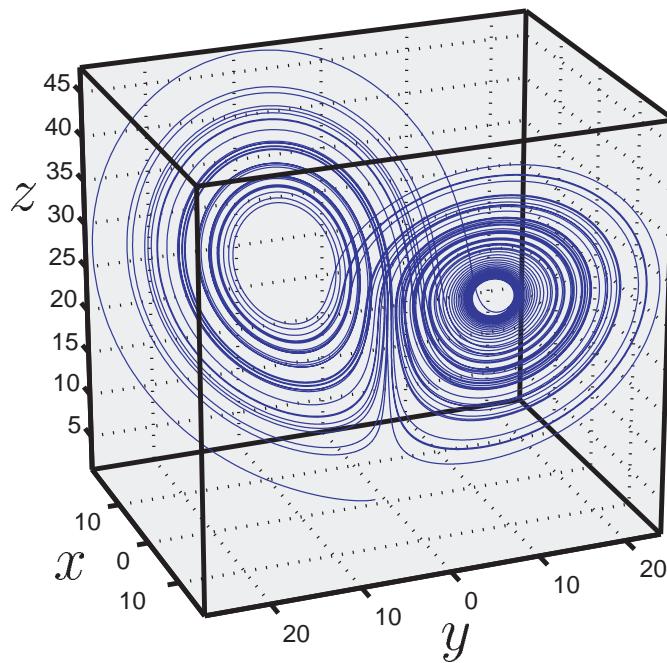


FIGURE 1.12 – Système chaotique de Lorenz.

Exemple 10 (Exemple de cycle limite : l’onde de densité). *Un exemple de cycle limite observé en pratique sur site industriel est donné par l’onde de densité. Ce phénomène apparaît sur les puits de pétrole activés par gas-lift. La technique d’activation par gas-lift des puits de pétrole (reliant un réservoir situé en profondeur aux installations de surface) permet de produire des hydrocarbures à partir de champs matures. Au début de la production d’un puits la pression du réservoir suffit fréquemment à propulser les hydrocarbures jusqu’à la surface où le pétrole est récupéré. C’est une phase de production dite “naturelle” qui, suivant les caractéristiques du réservoir, peut durer de quelques à de nombreuses années. Malheureusement en expulsant les effluents vers la surface, le réservoir tend à se dépressuriser jusqu’à n’être plus capable de contrebalancer le poids de la colonne de liquide dans le puits. Il faut alors recourir à des moyens d’activation. Le gaz est injecté au fond du puits, il peut alors être utilisé pour pousser le liquide ou pour s’y mêler de façon à diminuer la masse volumique moyenne. On peut se reporter à la Figure 1.14 pour voir les différents éléments permettant la production (tubing, partie centrale) et l’injection de gaz (casing, partie annulaire). On a représenté sur la Figure 1.15 le cycle limite observé sur site. Ces oscillations nonlinéaires font sensiblement diminuer le débit de production et donc la rentabilité du champ pétrolier (elles concourent également à endommager les installations par les à-coups de pression –coups de bâlier– dûs à l’inhomogénéité de l’écoulement). Comme cela a été mentionné dans la Section 1.1, on constate en pratique une parfaite répétabilité de l’établissement du cycle limite, et que sa forme (mais pas sa phase) ne dépend pas des conditions initiales. Il s’agit effectivement d’un cycle limite. On pourra se référer à [69] pour plus de détails.*

Exemple 11 (Les systèmes de Liénard et leurs cycles limites). *On appelle système de Liénard les*

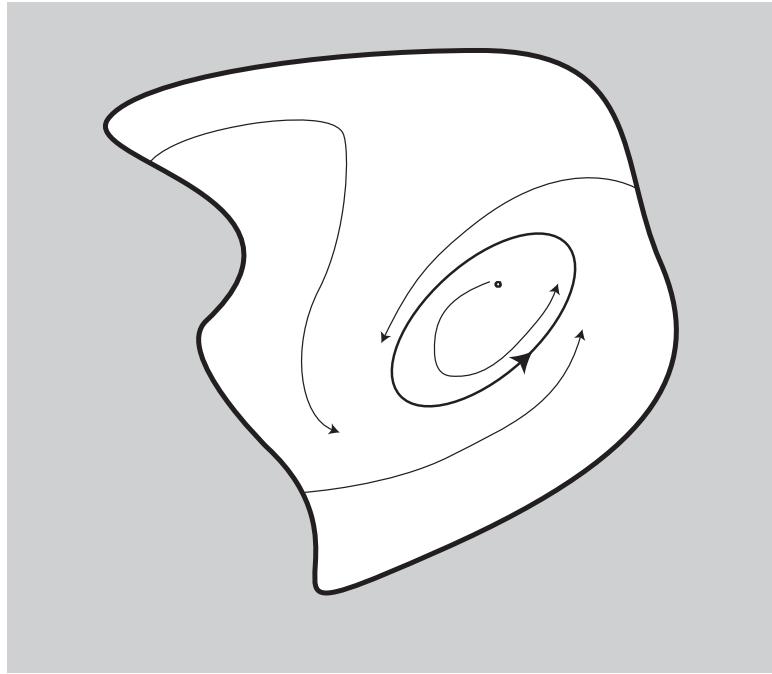


FIGURE 1.13 – Existence d'un cycle limite dans un compact positivement invariant et qui ne comporte qu'un seul point d'équilibre type *foyer instable* ou *nœud instable*.

systèmes de dimension 2 d'état $(x, \dot{x})^T$

$$\frac{d^2}{dt^2}x + f(x)\frac{d}{dt}x + g(x) = 0$$

où f et g sont des fonctions C^1 satisfaisant les conditions suivantes

1. f est paire et g est impaire et strictement positive sur \mathbb{R}^{+*}
2. La primitive de f nulle en zéro $F(x) = \int_0^x f(\tau)d\tau$ est négative sur un intervalle $0 < x < a$, nulle en a , jamais décroissante pour $x > a$ et tend vers $+\infty$ lorsque $|x| \rightarrow \infty$.

Ces systèmes possèdent une unique solution périodique. Ce résultat est connu sous le nom de théorème de Liénard (on pourra se référer à [12]). Un exemple de tel système est l'oscillateur de Van der Pol

$$\frac{d^2}{dt^2}x + \epsilon(x^2 - 1)\frac{d}{dt}x + x = 0$$

Ces équations représentent un circuit électrique oscillant à résistance négative (telle qu'on peut les reproduire avec des lampes de puissance). Ce circuit augmente naturellement l'amplitude des faibles oscillations tandis qu'elle atténue celle des oscillations trop fortes. On a représenté sur la Figure 1.16 le portrait de phases d'un oscillateur de Van der Pol. Pour $\epsilon > 0$, le cycle limite prévu par le théorème de Liénard est attractif, tandis que l'origine est un point d'équilibre instable. Plus ϵ est grand plus les trajectoires convergent rapidement vers ce cycle limite. La période à laquelle est parcourue de manière asymptotique le cycle limite est approximativement égale pour ϵ grand à $\epsilon(3 - 2 \ln(2))$.

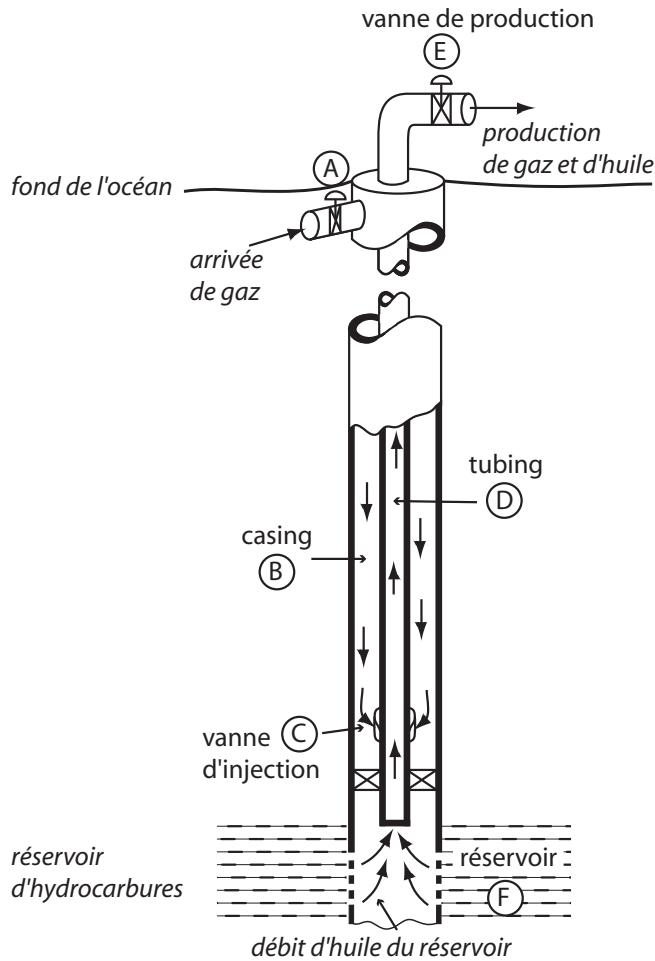


FIGURE 1.14 – Schéma d'un puits de pétrole activé par gas-lift.

1.4 Systèmes multi-échelles lents/rapides

Toutes les analyses mathématiques que nous proposons reposent sur un modèle de la physique du système que nous avons à étudier. Dans ce contexte, toute simplification préalable des équations peut sembler la bienvenue, mais on peut légitimement se demander si ces simplifications ne risquent pas de remettre en cause la validité des conclusions qu'elles ont permis d'établir. Le système réel est en général nettement plus complexe que la modélisation qu'on en fait. Une tendance naturelle est de proposer des modèles de plus en plus compliqués et de fait inextricables, ce qui est, en y réfléchissant bien assez facile et ne mène pas très loin dans la compréhension des phénomènes. Par exemple, on ne peut pas en général mener d'analyse théorique de stabilité comme nous l'avons fait sur un système de grande dimension. Il est en revanche nettement plus difficile de proposer un modèle de complexité minimale tenu des phénomènes que l'on souhaite comprendre et contrôler. Le but de cette section est de donner quelques résultats généraux sur les systèmes multi-échelles et leur approximation, sous certaines hypothèses, par des systèmes moins “complexes” et ne comportant essentiellement qu'une seule échelle de temps. C'est une des voies possibles pour justifier la pertinence de modèles réduits sur lesquels on sait prouver mathématiquement des résultats de stabilité et de robustesse. Les modèles plus compliqués, plus “proches” en quelque sorte de la réalité au sens platonicien du terme et très utiles comme modèles de simulation, sont alors vus comme des perturbations, prenant en compte

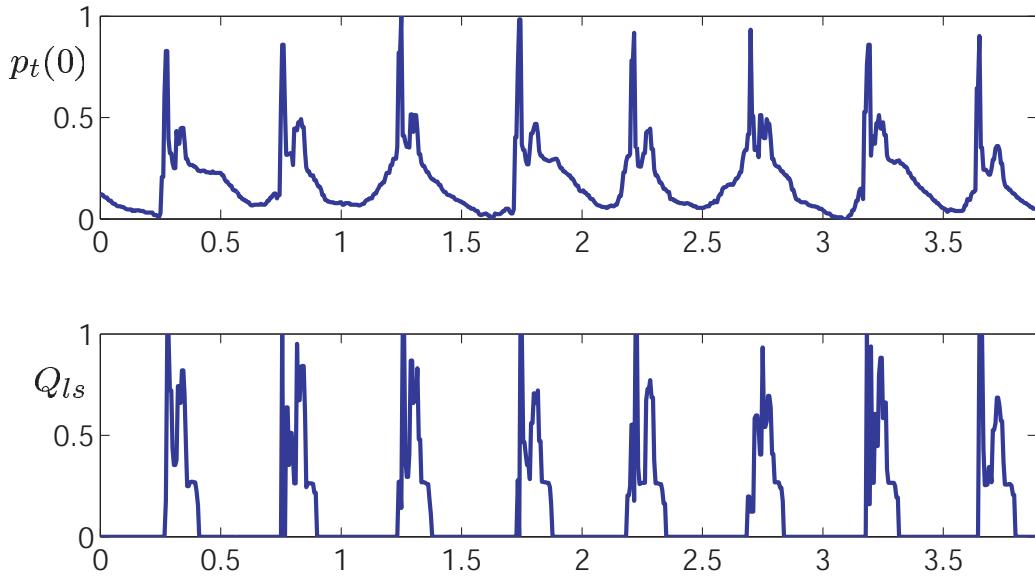


FIGURE 1.15 – Onde de densité dans un puits activé en gas-lift. On peut clairement observer un cycle limite suivi par la pression en tête de tubing (courbe du haut) et une production par bouchons (courbe du bas) très dommageable pour le débit d'huile produit (données normalisées TOTAL 2006).

des dynamiques rapides et donc des phénomènes à hautes fréquences, de modèles plus simples, de petites dimensions et très souvent utilisés en contrôle.

1.4.1 Perturbations singulières

Le premier outil que nous présentons est la théorie des *perturbations singulières*. Cette théorie a pour origine l'étude des phénomènes de couches limites dans les écoulements près des parois d'un fluide avec une faible viscosité. Certaines terminologies en sont directement inspirées.

La théorie des perturbations permet de relier les trajectoires de deux systèmes ayant des espaces d'état de dimensions différentes. Dans ce cadre, le système perturbé possède un nombre d'états plus grand que le système réduit. Plus précisément, cette théorie vise à éliminer les effets à court terme et à ne conserver que les effets à long terme. C'est un outil précieux pour la construction de *modèles réduits* résumant l'essentiel des comportements qualitatifs à long terme.

De manière générale, on distingue deux cas illustrés par la Figure 1.17 :

- premier cas : les effets rapides se stabilisent très vite et on parle alors de *perturbations singulières*, d'*approximation quasi-statique*, ou encore d'*approximation adiabatique* ;
- second cas : les effets rapides ne sont pas asymptotiquement stables mais restent d'amplitude bornée ; ils sont donc oscillants et l'on parle alors indifféremment de *moyennisation* ou d'*approximation séculaire*.

Seul le premier cas est abordé ici. Le second est traité dans l'Annexe C lorsque la dynamique rapide est périodique. Les cas plus généraux où la dynamique rapide n'est pas périodique sont nettement plus difficiles à formaliser : il faut passer par la théorie ergodique des systèmes dynamiques pour obtenir la dynamique lente en prenant des moyennes faisant intervenir la mesure asymptotique du rapide : la dynamique rapide est alors vue comme un bruit à haute fréquence dont il faut connaître la loi de probabilité (la mesure asymptotique) qui dépend en général des variables lentes.

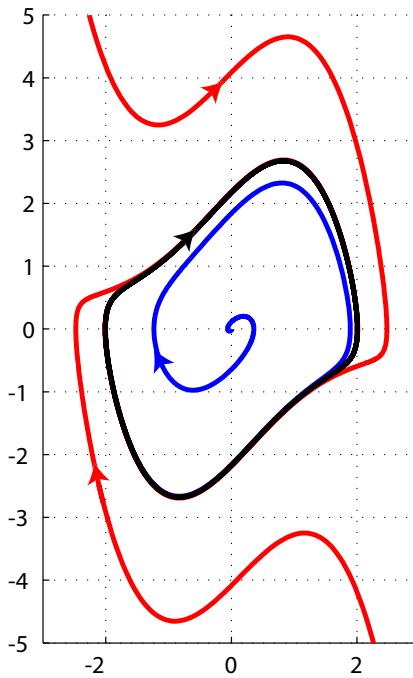


FIGURE 1.16 – Oscillations du système de Van der Pol et attraction du cycle limite.

On considère les systèmes continus du type

$$(\Sigma^\varepsilon) \begin{cases} \frac{dx}{dt} = f(x, z, \varepsilon) \\ \varepsilon \frac{dz}{dt} = g(x, z, \varepsilon) \end{cases}$$

avec $x \in \mathbb{R}^n$, $z \in \mathbb{R}^p$, où $0 < \varepsilon \ll 1$ est un petit paramètre, f et g sont des fonctions régulières. L'état partiel x correspond aux variables dont l'évolution est lente (variation significative sur une durée en t de l'ordre 1) et z correspond aux variables dont l'évolution est rapide (variation significative sur une durée en t de l'ordre de ε). On dit que $t \approx \varepsilon$ correspond à l'échelle de temps rapide et $t \approx 1$ à l'échelle de temps lente.

Considérons pour commencer l'exemple suivant¹⁰

$$\begin{cases} \frac{d}{dt}x = z \\ \varepsilon \frac{d}{dt}z = x - z \end{cases}$$

avec $0 < \varepsilon \ll 1$. Intuitivement, on voit que x est une variable lente (sa vitesse est petite et d'ordre 1), tandis que z est une variable rapide (sa vitesse est d'ordre $1/\varepsilon$). On a donc envie de dire que z

10. Comme autre exemple caractéristique citons la cinétique chimique où les constantes de vitesses de certaines réactions peuvent être nettement plus grandes que d'autres (cinétiques lentes et cinétiques très rapides).

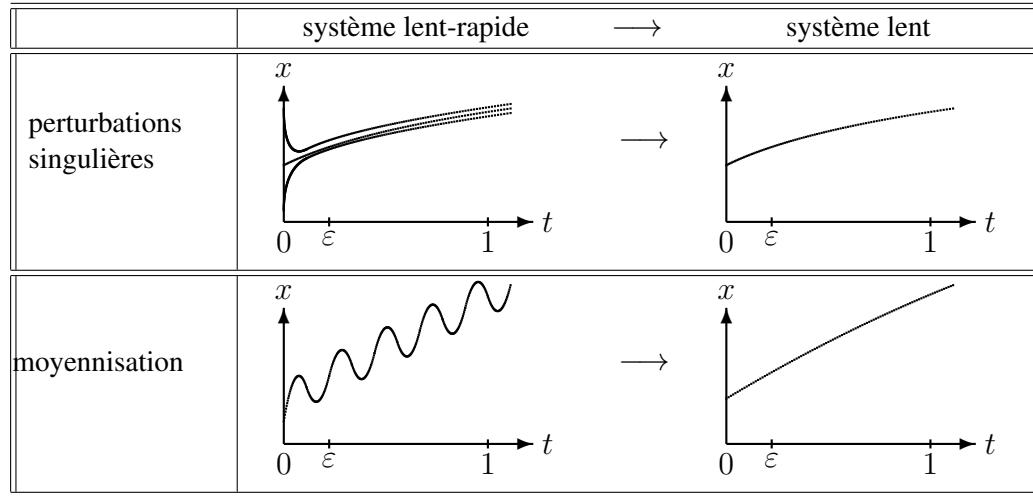


FIGURE 1.17 – La théorie des perturbations consiste à éliminer les effets à court terme, $t \sim \varepsilon$, qu'ils soient asymptotiquement stables ou oscillants, afin de ne conserver que les effets à long terme, $t \sim 1$ ($0 < \varepsilon \ll 1$).

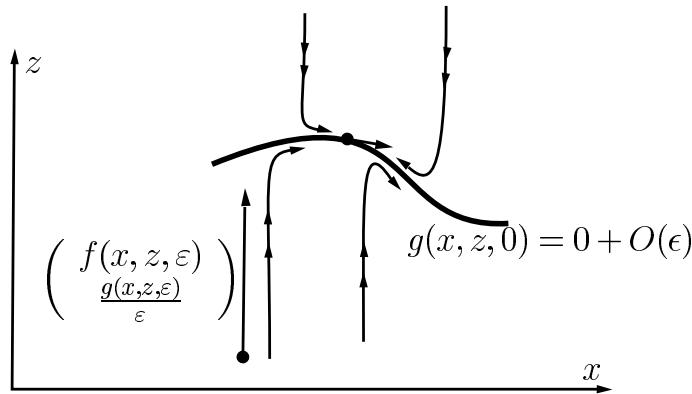


FIGURE 1.18 – Le champ des vitesses est quasi-vertical pour la forme normale de Tikhonov (Σ^ε).

atteint rapidement son point d'équilibre $z = x$ et que, par suite, x évolue selon $\frac{dx}{dt} = x$. Cette idée est fondamentalement correcte dès que les effets rapides sont asymptotiquement stables.

La situation géométrique est donnée par la Figure 1.18 : pour $\varepsilon > 0$ assez petit et localement autour de $g(x, z, 0) = 0$, les trajectoires du système sont quasi-verticales et convergent toutes vers le sous-ensemble de l'espace d'état (*sous-variété*) donné à l'ordre 0 en ε par l'équation $g(x, z, 0) = 0$.

Les résultats ci-dessous justifient alors, sous essentiellement l'hypothèse de stabilité asymptotique à x constant de la dynamique rapide en z , $\varepsilon \frac{dz}{dt} = g(x, z, 0)$, l'approximation des trajectoires du système perturbé (Σ^ε) par celles du système lent (Σ^0) obtenu en faisant $\varepsilon = 0$ dans les équations

$$(\Sigma^0) \left\{ \begin{array}{lcl} \frac{dx}{dt} & = & f(x, z, 0) \\ 0 & = & g(x, z, 0) \end{array} \right.$$

On néglige ainsi les convergences rapides vers la sous-variété donnée approximativement par les équations statiques $g(x, z, 0) = 0$. Toute trajectoire du système (Σ^ε) démarrant en (x, z) est proche, après une durée en t de l'ordre de ε , de la trajectoire du système lent (Σ^0) démarrant avec le même x (projection selon la verticale, voir Figure 1.18).

Nous voyons que cette approximation s'accompagne d'une diminution de la dimension de l'état. En fait, *la réduction n'est qu'une restriction à une sous-variété invariante attractive*, les équations de cette sous-variété étant approximativement données par $g(x, z, 0) = 0$. On a le premier résultat général suivant (sa démonstration figure dans [75])

Théorème 15 (Tikhonov)

Soit le système (Σ^ε) . Supposons que

H1 l'équation $g(x, z, 0) = 0$ admet une solution, $z = \rho(x)$, avec ρ fonction régulière de x telle que la matrice Jacobienne partielle

$$\frac{\partial g}{\partial z}(x, \rho(x), 0)$$

est une matrice dont toutes les valeurs propres sont à partie réelle strictement négative (Hurwitz) ; on dit alors que le système est sous *forme standard* ;

H2 le système réduit

$$\left\{ \begin{array}{lcl} \frac{dx}{dt} & = & f(x, \rho(x), 0) \\ x_{(t=0)} & = & x^0 \end{array} \right. \quad (1.16)$$

admet une unique solution $x_0(t)$ pour $t \in [0, T]$, $0 < T < +\infty$.

Alors, pour ε suffisamment proche de 0, le système complet (Σ_ε) admet une unique solution $(x_\varepsilon(t), z_\varepsilon(t))$ sur $[0, T]$ dès que la condition initiale z^0 appartient au *bassin d'attraction* du point d'équilibre $\rho(x^0)$ du sous-système rapide

$$\varepsilon \frac{d\zeta}{dt} = g(x^0, \zeta, 0).$$

De plus on a

$$\lim_{\varepsilon \rightarrow 0^+} x_\varepsilon(t) = x_0(t) \quad \text{et} \quad \lim_{\varepsilon \rightarrow 0^+} z_\varepsilon(t) = z_0(t)$$

uniformément pour t dans tout intervalle fermé de la forme $[a, T]$ avec $a > 0$.

Par le Théorème 4, l'hypothèse **H1** implique que, à x fixé, la dynamique de ζ

$$\varepsilon \frac{d\zeta}{dt} = g(x, \zeta, 0).$$

est localement asymptotiquement stable autour du point d'équilibre $\rho(x)$. Remarquons aussi que (Σ_0) s'écrit ainsi

$$\frac{d}{dt}x = f(x, \rho(x), 0)$$

avec $z = \rho(x)$, la fonction $x \mapsto \rho(x)$ étant définie implicitement par $g(x, \rho, 0) = 0$.

Sans hypothèses supplémentaires, l'approximation du Théorème 15 n'est valable, en général, que sur des intervalles de temps t de longueur bornée T . L'hypothèse supplémentaire, qu'il convient alors d'utiliser pour avoir une bonne approximation pour tous les temps positifs, concerne le comportement asymptotique du système réduit (1.16) : si ce dernier admet un point d'équilibre dont le linéaire tangent est asymptotiquement stable, l'approximation est alors valable pour tous les temps positifs (pourvu que la condition initiale x^0 soit dans le bassin d'attraction de l'équilibre de la dynamique lente).

Théorème 16 (Préservation de la stabilité)

Supposons en plus des hypothèses du Théorème 15 que le système réduit (1.16) admet un point d'équilibre \bar{x} : $f(\bar{x}, \rho(\bar{x}), 0) = 0$ et que les valeurs propres de la matrice

$$\left[\frac{\partial f}{\partial x} + \frac{\partial f}{\partial z} \cdot \frac{\partial \rho}{\partial x} \right]_{(\bar{x}, \rho(\bar{x}), 0)}$$

sont à partie réelle strictement négative. Alors, pour tout $\varepsilon \geq 0$ assez proche de 0, le système perturbé (Σ^ε) admet un point d'équilibre proche de $(\bar{x}, \rho(\bar{x}))$ et dont le linéaire tangent est asymptotiquement stable.

Preuve L'existence du point stationnaire pour le système perturbé est laissée en exercice (il suffit d'utiliser le théorème des fonctions implicites pour $g = 0$ et ensuite pour $f = 0$). Quitte à faire, pour chaque ε une translation, nous supposons que $(0, 0)$ est point stationnaire du système perturbé

$$f(0, 0, \varepsilon) = 0, \quad g(0, 0, \varepsilon) = 0$$

Notons, pour x proche de 0, $z = \rho_\varepsilon(x)$, la solution proche de 0 de $g(x, z, \varepsilon) = 0$. Suite à la translation précédente, on a $\rho_\varepsilon(0) = 0$. Considérons le changement de variables $(x, z) \mapsto (x, w = z - \rho_\varepsilon(x))$. Dans les coordonnées (x, w) , le système perturbé admet $(x, w) = (0, 0)$ comme point d'équilibre et ses équations ont la forme suivante

$$\frac{d}{dt}x = F(x, w, \varepsilon), \quad \varepsilon \frac{d}{dt}w = G(x, w, \varepsilon)$$

avec

$$F(x, w, \varepsilon) = f(x, w + \rho_\varepsilon(x), \varepsilon), \quad G(x, w, \varepsilon) = g(x, w + \rho_\varepsilon(x), \varepsilon) - \varepsilon \frac{\partial \rho_\varepsilon}{\partial x} F(x, w, \varepsilon).$$

Sa matrice Jacobienne en $(0, 0)$ est alors donnée par

$$J_\varepsilon = \begin{pmatrix} \frac{\partial F}{\partial x}(0, 0, \varepsilon) & \frac{\partial F}{\partial w}(0, 0, \varepsilon) \\ \frac{1}{\varepsilon} \frac{\partial G}{\partial x}(0, 0, \varepsilon) & \frac{1}{\varepsilon} \frac{\partial G}{\partial w}(0, 0, \varepsilon) \end{pmatrix} = \begin{pmatrix} A_\varepsilon & B_\varepsilon \\ C_\varepsilon A_\varepsilon & \frac{1}{\varepsilon} D_\varepsilon + E_\varepsilon \end{pmatrix}$$

où

$$\begin{aligned} A_\varepsilon &= \frac{\partial f}{\partial x}(0, 0, \varepsilon) + \frac{\partial f}{\partial z}(0, 0, \varepsilon) \frac{\partial \rho_\varepsilon}{\partial x}(0), & B_\varepsilon &= \frac{\partial f}{\partial z}(0, 0, \varepsilon), & C_\varepsilon &= -\frac{\partial \rho_\varepsilon}{\partial x}(0), \\ D_\varepsilon &= \frac{\partial g}{\partial z}(0, 0, \varepsilon), & E_\varepsilon &= -\frac{\partial \rho_\varepsilon}{\partial x}(0) \frac{\partial f}{\partial z}(0, 0, \varepsilon) \end{aligned}$$

On a utilisé le fait que $g(x, 0 + \rho_\varepsilon(x), \varepsilon) \equiv 0$ et que $F(0, 0, \varepsilon) = 0$. Les matrices $A_\varepsilon, B_\varepsilon, C_\varepsilon, D_\varepsilon$ et E_ε dépendent régulièrement de ε et convergent vers A_0, B_0, C_0, D_0 et E_0 quand ε tend vers 0 avec, par hypothèse A_0 et D_0 stables. On va montrer que pour $\varepsilon > 0$ assez petit, $\varepsilon J_\varepsilon$ est stable. Soient λ_ε et Y_ε une valeur propre et vecteur propre de $\varepsilon J_\varepsilon$. On peut toujours supposer Y_ε de norme 1 et que $\varepsilon \mapsto (\lambda_\varepsilon, Y_\varepsilon)$ est continue (les valeurs propres, vecteurs propres dépendent continûment des coefficients de la matrice). On décompose alors Y_ε en $(X_\varepsilon, W_\varepsilon)$ selon les coordonnées (x, w) . On a alors

$$A_\varepsilon X_\varepsilon + B_\varepsilon W_\varepsilon = \frac{\lambda_\varepsilon}{\varepsilon} X_\varepsilon, \quad C_\varepsilon A_\varepsilon X_\varepsilon + \left(\frac{1}{\varepsilon} D_\varepsilon + E_\varepsilon\right) W_\varepsilon = \frac{\lambda_\varepsilon}{\varepsilon} W_\varepsilon.$$

Soit $W_0 \neq 0$ est alors λ_0 est valeur propre de D_0 donc $\Re(\lambda_\varepsilon) < 0$ pour $\varepsilon > 0$ assez petit. Soit $W_0 = 0$ est alors $X_0 \neq 0$ et avec $A_0 X_0 = \left(\lim_{\varepsilon \rightarrow 0^+} \frac{\lambda_\varepsilon}{\varepsilon}\right) X_0$ on déduit aussi que pour $\varepsilon > 0$ assez petit, $\Re(\lambda_\varepsilon) < 0$.

Ainsi les valeurs propres de J_ε sont à parties réelles strictement négatives pour $\varepsilon > 0$ assez petit. ■

Cette preuve peut être améliorée pour montrer que l'approximation du Théorème 15 devient valide, localement autour de $(\bar{x}, \rho(\bar{x}))$ et pour tous les temps t positifs, dès que ε est assez petit (le caractère local étant alors indépendant de ε tendant vers zéro).

1.4.2 Feedback sur un système à deux échelles de temps

En théorie des systèmes, le Théorème 16 est utilisé constamment¹¹ de la manière suivante. Rajoutons une commande u à (Σ^ε) et supposons, à commande u fixée, que les hypothèses du Théorème 15 de Tikhonov soient valables. Ainsi

$$\begin{cases} \frac{dx}{dt} = f(x, z, u, \varepsilon) \\ \varepsilon \frac{dz}{dt} = g(x, z, u, \varepsilon) \end{cases} \quad (1.17)$$

avec $\rho(x, u)$ le point d'équilibre asymptotiquement stable de la partie rapide $\frac{d\zeta}{dt} = g(x, \zeta, u, 0)$. Le système lent est alors $\frac{dx}{dt} = f(x, \rho(x, u), u, 0)$.

Supposons que nous ayons un retour d'état lent $u = k(x)$ tel que le système lent bouclé soit asymptotiquement stable autour du point d'équilibre $(\bar{x}, \bar{u} = k(\bar{x}))$. Alors pour tout $\varepsilon > 0$ assez petit, le système perturbé (1.17) avec le bouclage lent $u = k(x)$, admet un *point d'équilibre hyperbolique* proche de $(\bar{x}, \bar{z} = \rho(\bar{x}, \bar{u}))$. Cela veut simplement dire que l'on peut, pour la synthèse d'un bouclage, ignorer des dynamiques asymptotiquement stables et assez rapides. On parle alors de *robustesse par*

11. C'est un peu comme Monsieur Jourdain dans le Bourgeois Gentil-Homme de Molière lorsqu'il réalise qu'il parle constamment en prose.

rapport aux dynamiques négligées. Noter que le retour d'état ne porte que sur la partie lente, x . Un bouclage du type $u = k(x, z)$ où z apparaît directement est envisageable mais alors il faut vérifier que la dépendance en z de u ne détruisse pas la stabilité de la dynamique rapide

$$\varepsilon \frac{d}{dt} \zeta = g(x, \zeta, k(x, \zeta), 0)$$

C'est toujours le cas lorsque g ne dépend pas de u : $\frac{\partial g}{\partial u} = 0$.

1.4.3 Modèle de contrôle et modèle de simulation

Pour les besoins de simulation ou de conception d'un régulateur, on n'a pas besoin d'utiliser le même modèle. Ainsi il n'y a pas de modèle universel. Pour ε petit il suffit, pour concevoir le feedback $u = k(x)$ de prendre par exemple comme modèle, l'approximation lente de (1.17)

$$\frac{d}{dt} x = f(x, \rho(x, u), u, 0).$$

Pour la simulation, il peut cependant être utile de vérifier que le feedback $u = k(x)$ injecté dans les équations (1.17) donne des résultats proches de ceux que l'on aurait avec le modèle lent en boucle fermée

$$\frac{d}{dt} x = f(x, \rho(x, k(x)), k(x), 0)$$

Cela permet de vérifier si ε est assez petit pour que le théorème soit applicable. En pratique, on remarque que ε n'a pas besoin d'être très petit : on constate qu'un rapport de 2 entre les constantes de temps de la dynamique rapide et celles de la dynamique lente donne bien souvent une bonne approximation. Cela veut dire que $\varepsilon = 1/2$ est déjà petit.

Voyons en détail la conception d'un contrôleur. On part d'un modèle de contrôle issu d'une modélisation même très grossière du système pour décrire les corrélations entre les entrée (u, w) et les mesures y . Ce modèle est le modèle d'état

$$\frac{d}{dt} x = f(x, u, w, p), \quad y = h(x)$$

dépendant des paramètres p . On élabore à partir de ce modèle de contrôle et pour une valeur nominale des paramètres \bar{p} , une loi de contrôle sous la forme d'un retour dynamique de sortie (comme l'est un *contrôleur PI*)

$$\frac{d}{dt} \xi = \bar{a}(y, \xi, v), \quad u = \bar{k}(y, \xi, v)$$

où v est le nouveau contrôle (pour le PI c'est la consigne). On est alors sûr du comportement du système bouclé nominal

$$\frac{d}{dt} x = f(x, \bar{k}(h(x), \xi, v), w, \bar{p}), \quad \frac{d}{dt} \xi = \bar{a}(h(x), v)$$

Par exemple, on sait que lorsque les perturbations w et le nouveau contrôle v sont constant, le système est asymptotiquement stable.

On peut bien sûr résoudre numériquement le système différentiel ci-dessus pour vérifier que nous ne nous sommes pas trompés dans les choix de \bar{k} et \bar{a} . On peut aussi, dans une seconde étape, simuler le système précédent mais avec un paramètre p différent de \bar{p} , pour vérifier que notre contrôleur n'est

pas trop sensible aux erreurs de paramètres et surtout pour quantifier les erreurs paramétriques au delà desquelles le contrôleur conduit à des comportements non acceptables. On quantifie en simulation la *robustesse paramétrique du contrôleur*. On voit donc ici que nous avons encore deux modèles : le modèle de contrôle de paramètre \bar{p} et le modèle de simulation de paramètre p . Les méthodes usuelles de contrôle garantissent alors que pour p assez proche de \bar{p} , tout se passe bien : de petites erreurs de paramètres n'engendrent que de petits écarts sur les trajectoires. En général, il n'existe pas de méthode simple qui permet de quantifier analytiquement les valeurs critiques de p au delà desquelles le système change de comportement. C'est l'objet de la théorie des bifurcations et des catastrophes, théories mathématiquement assez techniques (on pourra se reporter à [30]).

Il convient aussi de faire quelques simulations pour quantifier la robustesse du contrôleur par rapport à des dynamiques négligées. Le but est en autre de vérifier que les gains ne sont pas irréalistes (en général trop grands) : pour simplifier, il ne faut pas que, dans les formules servant à calculer \bar{k} et \bar{a} , il y ait de trop grands coefficients et/ou de petits diviseurs.

Les dynamiques négligées sont de trois ordres qualitativement : celles liées au processus de mesure et donc aux capteurs donnant y ; celles liées au processus de contrôle et donc aux actionneurs assurant des valeurs arbitraires pour u ; celles liées au système lui-même, indépendamment des actionneurs et des capteurs.

Par exemple, on peut associer à y et à u des premiers ordres avec des petites échelles de temps $\varepsilon_y > 0$ et $\varepsilon_u > 0$ pour prendre en compte le fait que la réponse des capteurs n'est pas instantanée et que les actionneurs ne sont pas parfaits. On simule alors le système étendu

$$\left\{ \begin{array}{l} \frac{d}{dt}x = f(x, u^m, w, p) \\ \frac{d}{dt}\xi = \bar{a}(y^m, v) \\ \varepsilon_y \frac{d}{dt}y^m = h(x) - y^m \\ \varepsilon_u \frac{d}{dt}u^m = \bar{k}(y^m, \xi, v) - u^m \end{array} \right.$$

et on teste pour diverses valeurs ε_y et ε_u les performances en quantifiant les seuils critiques au delà desquels les performances se dégradent notablement pour conduire à des instabilités¹².

Si, par exemple, on s'aperçoit que la constante de temps ε_y que l'on avait négligée au départ, est en fait trop grande compte tenu des performances demandées au contrôleur, alors il nous faut changer le modèle de contrôle et l'enrichir avec une partie de la dynamique du capteur. Cela signifie qu'on a pris comme modèle de contrôle de départ, un modèle trop grossier, compte tenu des objectifs.

Il ne faudrait pas en conclure qu'il faille dès le départ prendre le modèle le plus complet possible comme modèle de contrôle. Comme en physique, il est en fait bien plus efficace de partir d'une vision trop synthétique qui s'appuie sur des a priori très forts mais dont on est conscient¹³. D'en déduire un modèle simple et ultra-simplifié, dont on sait pertinemment qu'il est faux mais dont, si nécessaire,

12. Une autre façon de prendre en compte des dynamiques rapides négligées consiste à introduire un petit retard de l'ordre de $\varepsilon > 0$ en posant, par exemple, $y_m(t) = y(t - \varepsilon)$. Comme $y_m = e^{-\varepsilon \frac{d}{dt}} y$ on voit que l'on est aussi en face d'un phénomène de perturbation singulière (petit paramètre positif multipliant $\frac{d}{dt}$) mais cette fois en dimension infinie d'état. Avec l'approximation $e^{-\varepsilon \frac{d}{dt}} \approx \frac{1}{1 + \varepsilon \frac{d}{dt}}$, $y_m = e^{-\varepsilon \frac{d}{dt}} y$ devient $y_m = \frac{1}{1 + \varepsilon \frac{d}{dt}} y$ et on retrouve le filtre du premier ordre $y = y_m + \varepsilon \frac{d}{dt} y_m$.

13. Citons l'exemple des calculs perturbatifs de trajectoires spatiales pour lesquels on remet en cause les hypothèses simplificatrices au fur et à mesure : trajectoire plane puis dans l'espace, anomalie des potentiels de gravité, non-sphéricité des planètes, effets atmosphériques en altitude basse,...

on identifiera les faiblesses et que l'on corrigera au bout de quelques itérations de la méthodologie résumée ci-dessus. On aura alors obtenu un modèle de complexité réduite qui représentera l'essentiel de la dynamique que l'on souhaite contrôler.

Moralité : les choses sont déjà assez compliquées naturellement, gardons nous de les rendre encore plus obscures et utilisons des modèles de complexité aussi réduite que possible compte tenu des objectifs à atteindre.

1.5 Cas d'étude : PI et thermostat

L'objet de cette section est l'étude d'un régulateur proportionnel-intégral (*régulateur PI*) sur un système non linéaire du premier ordre. On prend en compte les contraintes sur le contrôle u par une gestion du terme intégral grâce à un algorithme d'anti-emballement (“*anti-windup*” en anglais). L'immense majorité des systèmes industriels de contrôle sont construits à partir de tels régulateurs PI élémentaires. Il est donc naturel de les étudier en détails.

Ce cas d'étude est aussi l'occasion d'utiliser un grand nombre des notions importantes issues de la théorie des équations différentielles ordinaires que nous avons présentée dans ce chapitre : stabilité asymptotique (voir Définition 2), relations entre la stabilité et les valeurs propres du système linéaire tangent autour d'un point d'équilibre (voir Théorème 10), systèmes avec des échelles de temps très différentes (voir la Section 1.4), liens avec les cascades de régulateurs et robustesse par rapport aux dynamiques négligées. Nous ferons aussi appel aux résultats fondamentaux de Poincaré et de Bendixon (voir Théorèmes 12 et 13) sur les systèmes stationnaires dans le plan. Enfin, nous montrerons l'intérêt de rajouter à la rétro-action PI (*feedback*) de la *pré-compensation (feedforward)* pour anticiper et ainsi mieux gérer des transitoires importants en particulier ceux liés aux changements de consignes.

1.5.1 Présentation du système

Considérons un bâtiment équipé d'une chaudière. Notre objectif est d'ajuster la marche de la chaudière en fonction d'une consigne de température $\bar{\theta}$ (librement choisie par un habitant) et d'une mesure en temps réel de la température ambiante θ à l'intérieur du bâtiment. Intuitivement, si l'écart $\theta - \bar{\theta}$ est positif il convient de baisser la chaudière, dans le cas contraire il convient de l'augmenter. Afin d'affranchir l'habitant des incessants ajustements nécessaires pour que la température θ reste le plus proche possible de la consigne $\bar{\theta}$, une question intéressante est : comment faire cela de façon automatique et fiable ?

1.5.2 Un régulateur PI

Comme l'illustre la Figure 1.19, l'algorithme usuellement utilisé avec une vanne proportionnelle 3 voies est le suivant. La marche de la chaudière est directement reliée à la position de la vanne trois voies notée u : pour $u = 0$, l'eau qui arrive des radiateurs ne passe pas par la chaudière et repart directement avec la même température ; pour $u = 1$, toute l'eau arrivant des radiateurs passe par la chaudière et ressort à la température de la chaudière θ_{ch} (température constante), température bien sûr nettement plus haute que $\bar{\theta}$; pour les valeurs intermédiaires de $u \in]0, 1[$, l'eau qui arrive des radiateurs n'est que partiellement réchauffée et repart avec une température intermédiaire. Le régulateur (thermostat) électronique qui se trouve dans une pièce de vie, comporte une sonde de température θ . L'habitant peut fixer une température de consigne $\bar{\theta}$. Le régulateur pilote, via un signal

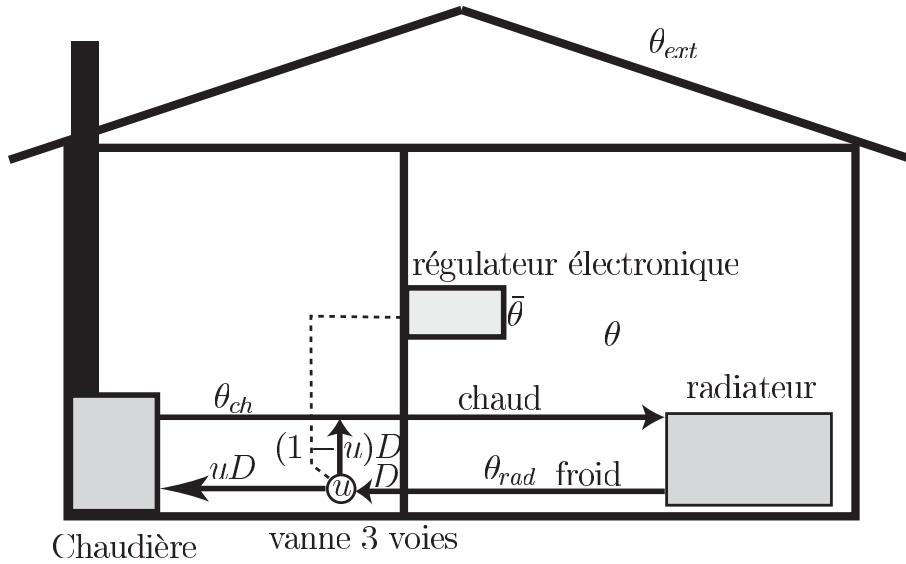


FIGURE 1.19 – Régulation PI de la température θ à sa consigne $\bar{\theta}$ avec une vanne trois voies de position $u \in [0, 1]$.

électrique, la position de la vanne u .

L'algorithme usuellement rencontré dans ce type d'équipement est celui d'un régulateur proportionnel/intégral, régulateur dit PI, qui s'écrit, entre deux instants séparés par $\Delta t > 0$,

$$u_{k+1} = K_p(\bar{\theta} - \theta_k) + I_k$$

où u_{k+1} est la nouvelle position de la vanne, à l'instant $(k + 1)\Delta t$, calculée en fonction de la température θ_k , de la consigne $\bar{\theta}$ et de la valeur d'un terme intégral I_k à l'instant $k\Delta t$. La période d'échantillonnage Δt est de l'ordre de la seconde. Le gain proportionnel K_p est un paramètre positif. Enfin, le terme intégral I_k se calcule par une récurrence

$$I_{k+1} = I_k + \Delta t K_i(\bar{\theta} - \theta_k)$$

où K_i est le *gain intégral*. Contrairement à u , le terme I est donc gardé en mémoire entre deux calculs successifs. C'est une valeur interne au contrôleur (état).

En somme, la commande u est calculée récursivement par l'algorithme

$$\begin{aligned} u_{k+1} &= K_p(\bar{\theta} - \theta_k) + I_k \\ I_{k+1} &= I_k + \Delta t K_i(\bar{\theta} - \theta_k) \end{aligned}$$

où, en même temps qu'on calcule u_{k+1} , on met à jour le terme intégral I_{k+1} pour le calcul suivant (celui de u_{k+2}). Tel quel, cet algorithme ne marche pas. La raison est brutale mais imparable : u doit rester entre 0 et 1. Or un tel algorithme ne garantit nullement ces conditions. Pour prendre en compte ces limitations, on sature u et on obtient l'algorithme modifié suivant

$$\begin{aligned} u_{k+1} &= S^{\text{at}}(K_p(\bar{\theta} - \theta_k) + I_k) \\ I_{k+1} &= I_k + \Delta t K_i(\bar{\theta} - \theta_k) \end{aligned}$$

où $S^{\text{at}}(x) = x$ si $x \in [0, 1]$, $S^{\text{at}}(x) = 0$ pour $x < 0$ et $S^{\text{at}}(x) = 1$ pour $x > 1$. Un tel algorithme ne marche pas non plus. La raison est en plus sophistiquée. Imaginons que nous soyons l'été par

temps de canicule. Dans ce cas θ sera toujours nettement supérieur à $\bar{\theta}$ (une chaudière n'est pas un climatiseur) et donc le terme intégral I deviendra au cours du temps de plus en plus négatif, on parle d'emballement ou windup, u restera du fait de la fonction S^{at} à 0. À la fin de l'été I aura une valeur négative très grande en valeur absolue. Lorsque les premiers froids arrivent, I sera tellement négatif que, même si θ descend largement en dessous de la consigne, u restera toujours à zéro. Il faudra attendre que I remonte au dessus de $K_p(\bar{\theta} - \theta)$ pour avoir enfin un peu de chauffage. Cela peut prendre des semaines. On comprend donc qu'il faut modifier aussi la façon dont on met à jour I . On va rajouter un terme témoignant de la saturation de la commande. Si u_{k+1} ne sature pas, ce terme sera nul, sinon ce terme cherchera à s'opposer à l'emballement dont on vient de parler. C'est le mécanisme d'anti-emballement (ou *anti-windup*) qui conduit à l'algorithme suivant

$$\begin{aligned} u_{k+1} &= S^{\text{at}}(K_p(\bar{\theta} - \theta_k) + I_k) \\ I_{k+1} &= I_k + \Delta t [K_i(\bar{\theta} - \theta_k) + K_s(u_{k+1} - K_p(\bar{\theta} - \theta_k) - I_k)] \end{aligned} \quad (1.18)$$

avec un gain K_s choisi assez grand de façon à avoir $K_s K_p > K_i$. Tout régulateur industriel PI comporte un tel mécanisme d'anti-windup pour prendre en compte les contraintes sur le contrôle u . Cette version est la bonne, nous allons l'étudier en détails.

Dans un premier temps, demandons-nous à quoi sert le terme (intégral) I et pourquoi n'avons nous pas tout simplement considéré le régulateur proportionnel P

$$u_{k+1} = S^{\text{at}}(K_p(\bar{\theta} - \theta_k))$$

Comme on pourrait le voir par exemple avec des simulations (et conformément à l'intuition) un tel algorithme a tendance à limiter les variations de température θ . Cependant, en régime stabilisé, il n'y a aucune raison pour que θ soit égal à sa consigne $\bar{\theta}$. Cet algorithme est incapable d'assurer, en régime stationnaire le fait que θ rejoigne sa consigne $\bar{\theta}$. En effet si $\theta = \bar{\theta}$ alors $u = 0$ et donc il ne faut pas chauffage : ce qui est certes très économique mais assez inconfortable et complètement idiot. Rapidement, on va s'éloigner de ce point qui ne correspond pas à une situation d'équilibre. En revanche avec le PI, en régime stabilisé où u reste constant et à l'intérieur (non strict) des contraintes 0 et 1 (et où θ et $\bar{\theta}$ restent eux aussi constants) on voit très simplement que I ne peut être que constant. En effet, la seconde équation de (1.18) implique nécessairement $\theta = \bar{\theta}$. Comme on le voit, le terme I garantit que, si on atteint un équilibre, c'est $\theta = \bar{\theta}$. Le terme I annule l'erreur statique.

1.5.3 Une modélisation simplifiée

Si on relève les signaux $t \mapsto u(t)$ et $t \mapsto \theta(t)$ sur une période de temps assez longue (par exemple une journée), on s'aperçoit qu'ils sont fortement corrélés. La modélisation consiste à établir ces corrélations. On peut à partir de telles mesures identifier un modèle. On peut aussi utiliser les lois de conservation de la physique. Bien sûr, il n'est pas question ici d'écrire un modèle détaillé prenant en compte un maximum de phénomènes. Notre objectif consiste à réguler une température moyenne, telle que mesurée par une seule sonde, en utilisant un seul actionneur (la chaudière). Si, en revanche, on s'intéresse à la température en différents points, à son inhomogénéité spatiale, on aura recours à des modèles plus sophistiqués tels que ceux utilisés pour la simulation fine des transferts thermiques.

Dans notre situation, nous avons besoin d'un modèle minimal nous permettant de comprendre pourquoi un régulateur PI fonctionne¹⁴, et ce sans avoir besoin de régler de façon très pointue les

14. Il faut savoir que pour l'immense majorité des installations domestiques, un algorithme PI pilotant une vanne 3 voies donne entièrement satisfaction avec quelques réglages standards pour K_p et K_i choisis selon la taille et l'inertie thermique de l'installation.

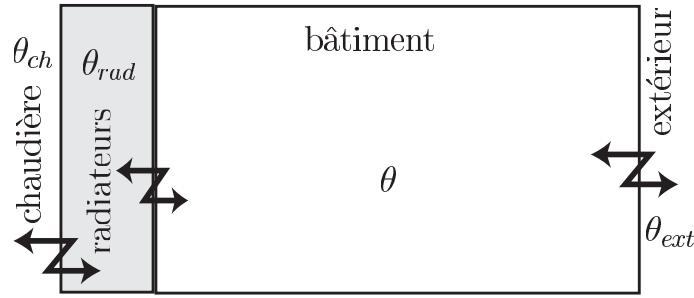


FIGURE 1.20 – Modélisation simplifiée avec un seul compartiment de température uniforme θ qui échange de la chaleur avec l’extérieur de température θ_{ext} et avec les radiateurs de température θ_{rad} dont une fraction de l’eau $u \in [0, 1]$ passe par la chaudière et ressort à la température sortie chaudière θ_{ch} .

gains K_p et K_i . Ce modèle sera bien-sûr grossier et donc faux. En d’autres termes, il ne représentera qu’une vision en réduction de la réalité physique. Il ne prendra en compte que des effets moyens en espace et en temps, moyennes relatives aux échelles qui nous intéressent. Ici l’échelle spatiale est définie par la taille du bâtiment, alors que l’intervalle temporel le plus court pris en considération est la demi-heure. Un tel modèle simplifié a l’immense avantage d’éclaircir l’analyse : de bien cerner les configurations pour lesquelles l’algorithme PI fonctionne ; de bien cerner aussi les cas où des instabilités (oscillations, pompages, ...) peuvent apparaître pour donner des pistes de solutions et ainsi traiter ces difficultés.

1.5.4 Passage en temps continu

Nous allons raisonner en temps continu pour suivre la formulation habituelle des lois de conservation de la physique.

Avec $\frac{d}{dt} I|_k \approx \frac{I_{k+1} - I_k}{\Delta t}$, on obtient la version en temps continu de (1.18)

$$\begin{cases} u(t) = S^{\text{at}}(K_p(\bar{\theta} - \theta(t)) + I(t)) \\ \frac{d}{dt} I(t) = K_i(\bar{\theta} - \theta(t)) + K_s(u(t) - K_p(\bar{\theta} - \theta(t)) - I(t)). \end{cases} \quad (1.19)$$

Pour construire un modèle reliant la position de la vanne u et la température θ , nous allons faire, en accord avec notre choix des échelles de temps et d’espace, les hypothèses simplificatrices illustrées sur la Figure 1.20.

On suppose le bâtiment homogène en température θ . Il échange des calories avec les radiateurs de température uniforme θ_{rad} , et avec l’extérieur dont la température est θ_{ext} . En notant $\Lambda_{rad} > 0$ et $\Lambda_{ext} > 0$, les coefficients d’échanges thermiques avec les radiateurs et l’extérieur, le bilan global d’énergie donne le modèle suivant

$$MC_p \frac{d}{dt} \theta = \Lambda_{rad}(\theta_{rad} - \theta) + \Lambda_{ext}(\theta_{ext} - \theta)$$

où M est la masse du bâtiment, C_p sa capacité calorifique spécifique (J/kg/K). Nous pouvons faire un bilan thermique stationnaire autour des radiateurs en supposant que le flux de chaleur perdue par les radiateurs $\Lambda_{rad}(\theta_{rad} - \theta)$ est instantanément compensé par la chaudière via la fraction du débit d’eau des radiateurs qui passe par la chaudière et en ressort à la température θ_{ch} . Cela donne la relation

statique suivante

$$u\Lambda(\theta_{ch} - \theta_{rad}) = \Lambda_{rad}(\theta_{rad} - \theta)$$

où $\Lambda > 0$ est un coefficient donné.

On peut alors calculer θ_{rad} comme barycentre de θ_{ch} et θ , avec des coefficients étant fonction de u

$$\theta_{rad} = \frac{u\Lambda}{u\Lambda + \Lambda_{rad}}\theta_{ch} + \frac{\Lambda_{rad}}{u\Lambda + \Lambda_{rad}}\theta$$

Ainsi, nous obtenons le modèle suivant

$$MC_p \frac{d}{dt}\theta = \frac{u\Lambda\Lambda_{rad}}{u\Lambda + \Lambda_{rad}}(\theta_{ch} - \theta) + \Lambda_{ext}(\theta_{ext} - \theta) \quad (1.20)$$

1.5.5 Simulations en boucle ouverte et en boucle fermée

On va utiliser les notations qui suivent

- x est l'*état* du système, i.e. les grandeurs satisfaisant des équations différentielles. Ici $x = \theta$.
- $y = h(x)$ est la partie de l'état qui est mesurée (sortie), ici $y = x$, on mesure tout l'état.
- u représente les contrôles (commandes), i.e., les variables d'entrée qu'on peut choisir comme on le souhaite.
- w désigne les *perturbations*, i.e., les variables d'entrée qu'on ne peut pas choisir car elles sont imposées par l'environnement ; ici $w = \theta_{ext}(t)$.
- p est un ensemble de paramètres constants apparaissant dans le modèle (ce sont des entrées particulières qui sont des constantes) : ici $p = (\Lambda, \Lambda_{rad}, \Lambda_{ext}, MC_p)$.

Ces notations nous permettent d'écrire le modèle sous la *forme d'état*

$$\frac{d}{dt}x = f(x, u, w, p), \quad y = h(x)$$

où f et h sont des fonctions issues de la modélisation ; ici

$$f(x, u, w, p) = \frac{u\Lambda\Lambda_{rad}}{(MC_p)(u\Lambda + \Lambda_{rad})}(\theta_{ch} - \theta) + \frac{\Lambda_{ext}}{MC_p}(\theta_{ext} - \theta), \quad h(x) = x$$

La simulation d'un tel système consiste à résoudre numériquement l'équation différentielle $\frac{d}{dt}x = f(x, u, w, p)$ à partir d'une condition initiale $x(0) = x^0$, d'un scénario de perturbations $t \mapsto w(t)$ donné à l'avance et de certaines valeurs des paramètres p . Les valeurs du contrôle sont également requises. Lorsque le contrôle est une loi horaire donnée à l'avance $t \mapsto u(t)$, on parle de simulation en *boucle ouverte*. Lorsque le contrôle u est au contraire calculé en fonction de x ou de y (i.e. par bouclage d'état ou de sortie), on parle de simulation en *boucle fermée*.

Le contrôle u donné par (1.19) fait apparaître une équation différentielle avec un état I , c'est un *bouclage dynamique*. Pour calculer l'évolution de la température, il faut aussi fixer la condition initiale I_0 sur l'état supplémentaire I . Enfin, pour simuler, il faut aussi se donner le scénario pour la consigne de température $t \mapsto \bar{\theta}(t)$ et connaître les paramètres K_p , K_i et K_s de réglage du régulateur.

Considérons une simulation en boucle fermée. Le modèle correspondant est

$$\left\{ \begin{array}{l} \frac{d}{dt}\theta(t) = \frac{u(t)\Lambda\Lambda_{rad}}{(MC_p)(u(t)\Lambda + \Lambda_{rad})}(\theta_{ch} - \theta(t)) + \frac{\Lambda_{ext}}{MC_p}(\theta_{ext} - \theta(t)) \\ \frac{d}{dt}I(t) = K_i(\bar{\theta} - \theta(t)) + K_s(u(t) - K_p(\bar{\theta} - \theta(t)) - I(t)) \\ \text{avec } u(t) = S^{\text{at}}(K_p(\bar{\theta} - \theta(t)) + I(t)) \end{array} \right. \quad (1.21)$$

avec $w = (\theta_{ext}, \theta_{ch})$ constant. Si on choisit des gains $K_p > 0$, $K_i > 0$ et $K_s > 0$ tels que $K_s K_p > K_i$, on constate que les trajectoires $t \mapsto (\theta(t), I(t))$ convergent toutes vers le même point, indépendamment de leur condition initiale. Ce point est donc globalement asymptotiquement stable. Le contrôle $u(t)$ converge vers une valeur \bar{u} dans $[0, 1]$. De plus, si $0 < \bar{u} < 1$ alors nécessairement la limite de $\theta(t)$ est la consigne $\bar{\theta}$.

Avec ces réglages, on n'a pas besoin de connaître précisément les valeurs des coefficients d'échanges thermiques ni celles de l'inertie thermique du bâtiment ni la valeur de la température extérieure. Par ces simulations, on se rend donc compte qu'avec un algorithme de contrôle élémentaire de type PI, il est possible de contrôler la température d'une très large gamme de bâtiments dont les transitoires thermiques peuvent être décrits de façon approximative par une équation bilan du type de celle construite ci-dessus. En fait, nous allons voir que cette constatation expérimentale correspond à un résultat général.

Considérons un système avec un état de dimension 1. Il suffit que $\frac{d}{dt}x = f(x, u, w, p)$ soit une fonction décroissante de $x = \theta$ et croissante de u pour qu'un tel régulateur fonctionne.

1.5.6 Un résultat général : régulateur PI sur un système non linéaire du premier ordre

Résultat

On considère le système (mono-dimensionnel) du premier ordre

$$\frac{d}{dt}x = f(x, u) \quad (1.22)$$

d'état $x(t) \in \mathbb{R}$ avec le contrôle scalaire $u(t) \in \mathbb{R}$ soumis aux contraintes $u \in [u^{\min}, u^{\max}]$ ($u^{\min} < u^{\max}$). On suppose que $\mathbb{R} \times [u^{\min}, u^{\max}] \ni (x, u) \mapsto f(x, u) \in \mathbb{R}$ est une fonction continue et dérivable par morceaux. Les seules informations que nous avons sur notre modèle, i.e. sur f , sont de nature qualitative : pour tout $(x, u) \in \mathbb{R} \times [u^{\min}, u^{\max}]$,

$$\frac{\partial f}{\partial x}(x, u) \leq 0 \text{ et } \frac{\partial f}{\partial u}(x, u) > 0$$

On suppose en outre qu'il existe un régime stationnaire respectant strictement les contraintes sur le contrôle : il existe $(\bar{x}, \bar{u}) \in \mathbb{R} \times [u^{\min}, u^{\max}]$ tel que $f(\bar{x}, \bar{u}) = 0$.

Nous allons montrer que le régulateur PI avec anti-emballlement

$$u = S^{\text{at}}[K_p(\bar{x} - x) + I], \quad \frac{d}{dt}I = K_i(\bar{x} - x) + K_s(u - K_p(\bar{x} - x) - I) \quad (1.23)$$

où la fonction saturation S^{at} est définie par

$$\mathbb{R} \ni u \mapsto S^{\text{at}}(u) = \begin{cases} u^{\min}, & \text{si } u \leq u^{\min}; \\ u, & \text{si } u^{\min} \leq u \leq u^{\max}; \\ u^{\max}, & \text{si } u^{\max} \leq u. \end{cases} \quad (1.24)$$

rend le point d'équilibre (\bar{x}, \bar{u}) globalement asymptotiquement stable pour le système d'état étendu (x, I) dès que ses gains, K_p le gain proportionnel, K_i le gain intégral et K_s le *gain d'anti-emballlement*, vérifient

$$K_p > 0, \quad K_i > 0, \quad K_s > 0, \quad \text{et} \quad K_s K_p \geq K_i$$

Dynamique à étudier

Nous cherchons à montrer la stabilité asymptotique globale du point d'équilibre (\bar{x}, \bar{u}) . Nous cherchons donc à établir que, pour toute condition initiale (x^0, I^0) , la solution $t \mapsto (x(t), I(t))$ du système bouclé

$$\begin{aligned}\frac{d}{dt}x &= f(x, S^{\text{at}}[K_p(\bar{x} - x) + I]) \\ \frac{d}{dt}I &= K_i(\bar{x} - x) + K_s(S^{\text{at}}[K_p(\bar{x} - x) + I] - K_p(\bar{x} - x) - I)\end{aligned}\tag{1.25}$$

existe pour tout temps $t \geq 0$ et que sa limite quand t tend vers $+\infty$ est (\bar{x}, \bar{u}) .

Dans tout ce qui suit, nous considérons l'état étendu $X = (x, I)$ et réécrivons (1.25) sous la forme $\frac{d}{dt}X = F(X)$. La fonction $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ est une fonction continue et dérivable par morceaux de X . On note aussi $\bar{X} = (\bar{x}, \bar{u})$ le point d'équilibre de $F : F(\bar{X}) = 0$.

Les trajectoires sont bornées

Nous allons montrer que les trajectoires du système (1.25) restent bornées pour $t > 0$. Elles seront donc automatiquement définies pour tout temps $t > 0$. Pour cela nous allons considérer le *plan de phases* $X = (x, I)$ et construire une famille de rectangles emboîtés les uns dans les autres. Chacun de ces rectangles est *positivement invariant* par la dynamique (voir Définition 6) : si la condition initiale appartient à l'intérieur d'un de ces rectangles alors il est impossible à la trajectoire d'en sortir. Pour établir ce point, il suffit de regarder la direction du vecteur vitesse $F(X)$ quand X parcourt le bord du rectangle : si $F(X)$ pointe constamment vers l'intérieur, alors il est impossible de sortir du rectangle en intégrant selon les temps positifs. L'existence de ces ensembles repose sur les hypothèses déjà évoquées $K_p > 0$, $K_i > 0$, $K_s > 0$, $K_s K_p \geq K_i$, f croissante par rapport à u et décroissante par rapport à x .

La fonction S^{at} qui intervient dans (1.24) conduit à un découpage en trois zones du plan de phases (x, I) (voir Figure 1.21) selon la position de $K_p(\bar{x} - x) + I$ par rapport à u^{\min} et u^{\max} : les deux zones saturées $K_p(\bar{x} - x) + I \geq u^{\max}$ et $K_p(\bar{x} - x) + I \leq u^{\min}$ et la zone sans saturation $u^{\min} \leq K_p(\bar{x} - x) + I \leq u^{\max}$.

Posons $L = \max(u^{\max} - \bar{u}, \bar{u} - u^{\min}) > 0$. Comme représenté sur la Figure 1.21, considérons, pour $\lambda \geq L$, le rectangle R_λ de centre (\bar{x}, \bar{u}) , de cotés parallèles aux axes et coupant l'axe des x en $\bar{x} \pm \lambda/K_p$ et l'axe des I en $\bar{u} \pm \lambda$. Étudions maintenant, en détaillant ses composantes F^x et F^I , la direction, pour (x, I) sur le bord de R_λ , du vecteur tangent à la trajectoire

$$F(x, I) = \begin{cases} F^x(x, I) = f(x, S^{\text{at}}[K_p(\bar{x} - x) + I]) \\ F^I(x, I) = K_i(\bar{x} - x) + K_s(S^{\text{at}}[K_p(\bar{x} - x) + I] - K_p(\bar{x} - x) - I) \end{cases}$$

— (coté horizontal supérieur) pour $I = \bar{u} + \lambda$ et $x \in [\bar{x} - \lambda/K_p, \bar{x} + \lambda/K_p]$, on voit que $F^I(x, \bar{u} + \lambda) \leq 0$; en effet si $x \in [\bar{x} - \lambda/K_p, \bar{x}]$ on a

$$F^I = \overbrace{(K_i - K_s K_p)(\bar{x} - x)}^{\leq 0} + \overbrace{K_s(u^{\max} - \bar{u} - \lambda)}^{\geq 0} \leq 0;$$

si $x \in [\bar{x}, \bar{x} + (\bar{u} + \lambda - u^{\max})/K_p]$ on a

$$F^I = K_i \overbrace{(\bar{x} - x)}^{\leq 0} + K_s \overbrace{(u^{\max} - \bar{u} - \lambda - K_p(\bar{x} - x))}^{\leq 0} \leq 0;$$

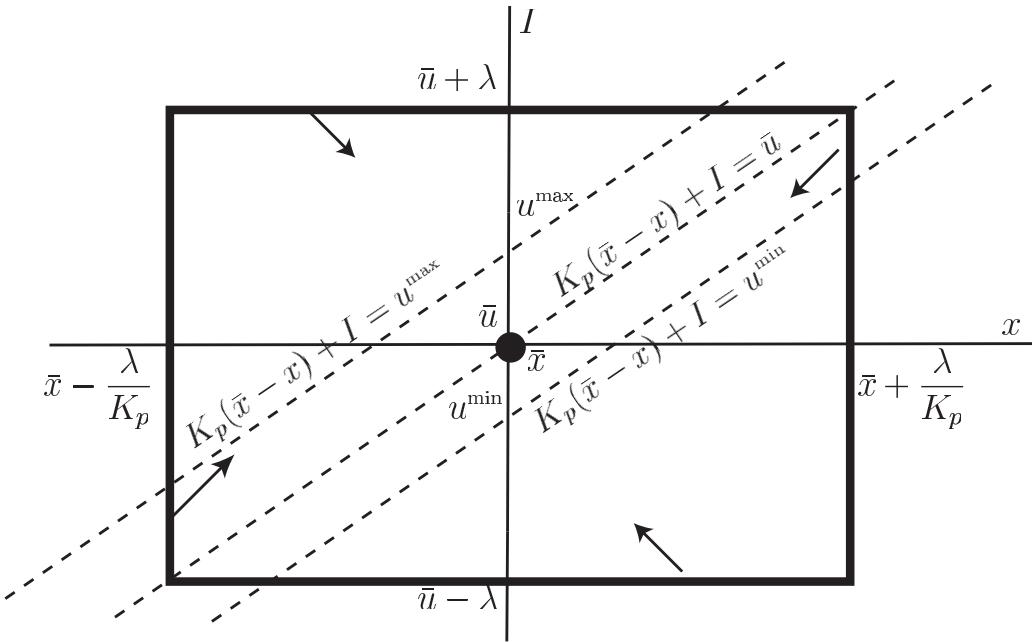


FIGURE 1.21 – Le rectangle R_λ est positivement invariant car le champ de vecteurs $X \mapsto F(X)$ pointe vers l'intérieur lorsque $X = (x, I)$ parcourt le bord de R_λ .

si $x \in [\bar{x} + (\bar{u} + \lambda - u^{\max})/K_p, \bar{x} + \lambda/K_p]$ on a

$$F^I = K_i \underbrace{(\bar{x} - x)}_{\leq 0} \leq 0$$

Ainsi, le vecteur $F = \begin{pmatrix} F^x \\ F^I \end{pmatrix}$ pointe vers le bas le long du côté horizontal supérieur.
— (côté vertical droit) pour $x = \bar{x} + \lambda/K_p$ et $I \in [\bar{u} - \lambda, \bar{u} + \lambda]$ on voit que

$$F^x(\bar{x} + \lambda/K_p, I) \leq 0;$$

en effet si $I \in [\bar{u} - \lambda, u^{\min} + \lambda]$ on a

$$F^x = f(\bar{x} + \lambda/K_p, u^{\min}) \leq f(\bar{x}, u^{\min}) \leq f(\bar{x}, \bar{u}) = 0$$

car $\frac{\partial f}{\partial x} \leq 0$ et $\frac{\partial f}{\partial u} > 0$; si $I \in [u^{\min} + \lambda, \bar{u} + \lambda]$ on a de même

$$F^x = f(\bar{x} + \lambda/K_p, I - \lambda) \leq f(\bar{x}, I - \lambda) \leq f(\bar{x}, \bar{u}) = 0$$

Ainsi, le vecteur F pointe vers la gauche le long du côté vertical droit.

On démontre, de la même façon, que F pointe vers le haut le long du côté horizontal inférieur et pointe vers la droite le long du côté vertical gauche. En conclusion, pour tout $\lambda \geq L$, le vecteur tangent à la trajectoire $F(X)$ pointe vers l'intérieur du rectangle R_λ . Cet ensemble est donc positivement invariant.

Prenons maintenant une condition initiale $X^0 = (x^0, I^0)$ arbitraire. Il existe toujours au moins un λ^0 (il suffit de le prendre assez grand en fait) pour que X^0 soit à l'intérieur de R_{λ^0} . Ainsi, la solution de $\frac{d}{dt}X = F(X)$ avec $X(0) = X^0$ reste dans R_{λ^0} pour tout temps positif : elle est donc définie pour tout temps $t > 0$ et reste bornée.

Stabilité asymptotique

Nous cherchons maintenant à étudier la convergence des trajectoires. On suppose que l'équilibre (\bar{x}, \bar{I}) réside strictement à l'intérieur des contraintes, c.-à-d.

$$u^{\min} < \bar{I} = \bar{u} < u^{\max}$$

Pour (x, I) autour de (\bar{x}, \bar{I}) , le système s'écrit

$$\begin{cases} \frac{d}{dt}x = f(x, K_p(\bar{x} - x) + I) \\ \frac{d}{dt}I = K_i(\bar{x} - x) \end{cases} \quad (1.26)$$

Le calcul du linéaire tangent fait apparaître la matrice Jacobienne suivante

$$\begin{pmatrix} \frac{\partial f}{\partial x}\big|_{(\bar{x}, \bar{u})} - K_p \frac{\partial f}{\partial u}\big|_{(\bar{x}, \bar{u})} & \frac{\partial f}{\partial u}\big|_{(\bar{x}, \bar{u})} \\ -K_i & 0 \end{pmatrix} \quad (1.27)$$

La trace de cette matrice 2×2 est strictement négative car $K_p > 0$, $K_i > 0$, $\frac{\partial f}{\partial x} \leq 0$ et $\frac{\partial f}{\partial u} > 0$ par hypothèse. Son déterminant est strictement positif pour les mêmes raisons. Donc ses deux valeurs propres sont à partie réelle strictement négative. Ainsi, quels que soient les choix de $K_p > 0$ et $K_i > 0$, le linéaire tangent est toujours localement asymptotiquement stable (voir par exemple la Section 1.2.3).

C'est bien un système continu et stationnaire dans le plan. On peut donc utiliser le Théorème 12 de Poincaré. Nous avons vu que les trajectoires restent bornées. Donc, les trajectoires convergent soit vers un équilibre, soit vers une courbe fermée du plan qui délimite un domaine borné.

Or, il n'existe pas d'autre équilibre que l'équilibre $(\bar{x}, \bar{I} = \bar{u})$ dont on a supposé l'existence à l'intérieur strict des contraintes. En effet, (\bar{x}, \bar{I}) est l'unique solution de

$$\begin{cases} f(x, S^{\text{at}}[K_p(\bar{x} - x) + I]) = 0 \\ K_i(\bar{x} - x) + K_s(S^{\text{at}}[K_p(\bar{x} - x) + I] - K_p(\bar{x} - x) - I) = 0 \end{cases}$$

L'unicité résulte du raisonnement suivant

- Si $u^{\min} \leq K_p(\bar{x} - x) + I \leq u^{\max}$, on a $x = \bar{x}$ par la seconde équation, et alors la première équation donne I comme solution de $f(\bar{x}, I) = 0$. Or, f est strictement croissante par rapport à son second argument, donc $I = \bar{I}$.
- Si $K_p(\bar{x} - x) + I > u^{\max}$, la première équation $f(x, u^{\max}) = 0$ implique que $x > \bar{x}$, car $f(\bar{x}, u^{\max}) > 0$ et f est décroissante par rapport à x . La seconde équation donne alors $K_i(x - \bar{x}) = K_s(u^{\max} - (K_p(\bar{x} - x) + I)) < 0$, car $K_s > 0$ par hypothèse. Ceci contredit le fait que x est plus grand que \bar{x} .
- Si $K_p(\bar{x} - x) + I < u^{\min}$, la première équation $f(x, u^{\min}) = 0$ implique que $x < \bar{x}$ et la seconde le contraire.

Pour montrer la convergence globale vers l'unique équilibre (\bar{x}, \bar{I}) de (1.25), il suffit de montrer qu'il n'existe pas de courbe fermée du plan, formée à partir de trajectoires du système. Il ne restera que la convergence vers l'unique équilibre comme régime asymptotique possible.

Pour montrer qu'il n'existe pas de telles courbes fermées, nous procédons par contradiction. Admettons qu'à partir des trajectoires de (1.25), on puisse former par raccordement une courbe fermée du plan. On note Ω l'intérieur de cette courbe fermée. Calculons l'intégrale de la divergence du champ

F sur le domaine borné Ω . Cette intégrale est égale à l'intégrale sur le bord $\partial\Omega$ de son flux sortant¹⁵. Avec $\langle \cdot, \cdot \rangle$ le produit scalaire usuel, il vient

$$\int \int_{\Omega} \operatorname{div} F = \oint_{\partial\Omega} \langle F, n \rangle \, ds$$

où n est la normale extérieure. Par construction, le champ de vecteur associé à (1.25)

$$F = \begin{pmatrix} f(x, S^{\text{at}}[K_p(\bar{x} - x) + I]) \\ K_i(\bar{x} - x) + K_s(S^{\text{at}}[K_p(\bar{x} - x) + I] - K_p(\bar{x} - x) - I) \end{pmatrix}$$

est tangent au bord $\partial\Omega$ de Ω : son flux sortant est donc nul. Pourtant, on remarque que sa divergence (trace de la matrice Jacobienne) est toujours strictement négative

$$\frac{\partial f}{\partial x} - K_p \frac{\partial f}{\partial u} (S^{\text{at}})' + K_s ((S^{\text{at}})' - 1) < 0$$

car $(S^{\text{at}})'$ vaut soit 0 soit 1. On obtient la contradiction recherchée. En conclusion, il n'existe pas de telle courbe fermée, et le point d'équilibre (\bar{x}, \bar{I}) est globalement asymptotiquement stable.

En fait nous avons montré le point no 1 du résultat suivant :

Théorème 17 (PI avec anti-emballlement sur un premier ordre non-linéaire et stable)

Soit le système du premier ordre non-linéaire $\frac{d}{dt}x = f(x, u)$ avec, $x \in \mathbb{R}$, $u \in [u^{\min}, u^{\max}]$, f continue, dérivable par morceau vérifiant $\frac{\partial f}{\partial x}(x, u) \leq 0$, $\frac{\partial f}{\partial u}(x, u) > 0$ pour tout avec $(x, u) \in \mathbb{R} \times [u^{\min}, u^{\max}]$.

Soit le régulateur PI avec anti-emballlement de consigne $\bar{x} \in \mathbb{R}$:

$$u = S^{\text{at}}[K_p(\bar{x} - x) + I], \quad \frac{d}{dt}I = K_i(\bar{x} - x) + K_s(u - K_p(\bar{x} - x) - I)$$

avec $K_p, K_i > 0$, $K_s K_p > K_i$ et $S^{\text{at}}(v) \equiv \max(u^{\min}, \min(u^{\max}, v))$. Alors on a les deux cas suivants :

1. soit il existe $\bar{u} \in]u^{\min}, u^{\max}[$ tel que $f(\bar{x}, \bar{u}) = 0$ et alors l'équilibre $(x, I) = (\bar{x}, \bar{u})$ du système bouclé est unique et globalement asymptotiquement stable au sens de Lyapounov
2. sinon :
 - soit $\forall v \in [u^{\min}, u^{\max}], f(\bar{x}, v) < 0$ (resp. > 0) alors $\lim_{t \rightarrow +\infty} u(t) = u^{\max}$ (resp. u^{\min}) et $\lim_{t \rightarrow +\infty} x(t)$ existe, est éventuellement infinie dans $[-\infty, \bar{x}[$ (resp. $]bar{x}, +\infty]$)
 - soit $f(\bar{x}, u^{\max}) = 0$ (resp. $f(\bar{x}, u^{\min}) = 0$) alors $\lim_{t \rightarrow +\infty} u(t) = u^{\max}$ (resp. u^{\min}) et $\lim_{t \rightarrow +\infty} x(t)$ existe, est finie dans $]-\infty, \bar{x}[$ (resp. $]bar{x}, +\infty[$).

Le point no 2 signifie simplement que, compte tenu des contraintes, le régulateur PI fait au mieux, i.e., tente de rapprocher au maximum x de sa consigne \bar{x} . La preuve de ce point est laissée en exercice.

1.5.7 Dynamiques négligées : rôle du contrôle dans l'approximation

Les simulations en boucle fermée du thermostat que nous pouvons déduire du modèle (1.21) ne sont qu'une représentation approximative de la réalité. Cela est dû, entre autres, à l'écriture par un

15. Il s'agit du théorème de Gauss souvent utilisé en électrostatique.

modèle continu (1.19) du régulateur PI qui est à l'origine régi par des équations discrètes (1.18). Si la période d'échantillonnage Δt est petite nous ne faisons qu'une petite erreur. D'autres erreurs proviennent des nombreux phénomènes dynamiques négligés comme par exemple l'hétérogénéité des températures θ et θ_{rad} , les fluctuations de la chaudière dues à sa propre régulation de température, ...

Bref, nous avons implicitement supposé, en accord avec les résultats de la Section 1.4, que si d'autres phénomènes transitoires existent, ils sont rapides et stables. En conséquence, notre modèle de simulation ne prend en compte que les dynamiques les plus lentes (disons celles qui sont plus lentes que 30 minutes).

Est-ce gênant en pratique ? Bien qu'en théorie nous avons montré que l'équilibre de (1.21) est globalement asymptotiquement stable pour tout gain K_p , K_i et K_s positifs vérifiant $K_s K_p > K_i$, il apparaît, lorsqu'on passe aux expérimentations, un phénomène qui semble contredire ce résultat. Le résumé des différentes expériences qu'on peut réaliser est qu'il n'est pas possible de choisir ces gains aussi grands que nous le souhaiterions. Lorsque les gains sont petits on constate un bon accord entre la théorie proposée et la pratique. Lorsque les gains sont grands, on constate des écarts importants, voire des instabilités.

Ceci est en fait en accord avec la théorie, si on en revient aux hypothèses que nous avons formulées plus ou moins explicitement. L'explication est que des gains très grands rendent le système (1.21) très rapide. Pour s'en convaincre il suffit de voir que les valeurs propres du système linéaire tangent autour de l'équilibre (\bar{x}, \bar{I}) sont celles de la matrice Jacobienne (1.27). Au moins une des valeurs propres de cette matrice tend vers l'infini (en module) quand K_p et/ou K_i tendent vers $+\infty$. S'il est possible, en théorie, de rendre le système (1.21) aussi rapide qu'on le désire grâce à un régulateur PI avec des grands gains, en pratique, les gains du régulateur sont limités par l'apparition d'instabilités liées à toutes les dynamiques rapides négligées et qui sont alors excitées par le contrôle. Les hypothèses d'application des théorèmes d'approximation ne sont plus vérifiées. Par exemple, l'approximation du PI discret par un PI continu n'est plus valable, si on prend K_p et K_i tels que le temps intégral $T_i = K_p/K_i$ est du même ordre de grandeur que la période d'échantillonnage Δt . On va se concentrer sur un des possibles problèmes qui est caractéristique de ce phénomène : la dynamique négligée de la sonde thermique.

Exemple : le rôle de la sonde Rajoutons au modèle de simulation en boucle fermée (1.21), la dynamique de la sonde de température. Le modèle le plus simple ayant une base physique évidente est une dynamique linéaire stable du premier ordre (filtre passe-bas) avec une constante de temps très petite car la sonde a très peu d'inertie thermique et donc répond très rapidement (en quelques secondes pour fixer l'ordre de grandeur) aux changements de la température de l'air ambiant. Ce modèle est vérifiable expérimentalement, par analyse de la réponse transitoire de la sonde thermique à des signaux d'excitation type échelon. En notant $\varepsilon > 0$ cette petite constante de temps, la température qu'on peut utiliser pour le calcul de u n'est plus θ mais sa mesure notée θ^m reliée à θ via le filtre du premier ordre suivant

$$\frac{d}{dt} \theta^m = \frac{\theta - \theta^m}{\varepsilon}$$

Cette dynamique supplémentaire complète le modèle en boucle fermée à étudier. Celui-ci est alors

$$\left\{ \begin{array}{l} \frac{d}{dt} \theta(t) = \frac{u(t)\Lambda\Lambda_{rad}}{(MC_p)(u(t)\Lambda + \Lambda_{rad})} (\theta_{ch} - \theta(t)) + \frac{\Lambda_{ext}}{MC_p} (\theta_{ext} - \theta(t)) \\ \frac{d}{dt} I(t) = K_i (\theta - \theta^m(t)) + K_s (u(t) - K_p(\bar{\theta} - \theta^m(t)) - I(t)) \\ \varepsilon \frac{d}{dt} \theta^m(t) = \theta(t) - \theta^m(t) \\ \text{avec } u(t) = S^{\text{at}}(K_p(\bar{\theta} - \theta^m(t)) + I(t)) \end{array} \right. \quad (1.28)$$

En posant $X = (\theta, I)$ et $Z = \theta^m$, ce modèle admet la structure suivante

$$\frac{d}{dt}X = F(X, Z), \quad \varepsilon \frac{d}{dt}Z = G(X, Z),$$

alors que le modèle initial (1.21) correspond en fait à $\varepsilon = 0$

$$\frac{d}{dt}X = F(X, Z), \quad 0 = G(X, Z).$$

En application du Théorème 15 de Tikhonov, si les valeurs propres de $\frac{\partial G}{\partial Z}(X, Z)$ pour (X, Z) vérifiant $G(X, Z) = 0$ sont toutes à partie réelle négative (ce qui est trivialement le cas ici), alors, pour $\varepsilon > 0$ assez petit, les trajectoires des deux modèles sont très proches. Autrement dit, si la dynamique de la sonde thermique est stable et rapide devant les autres dynamiques, on peut la négliger.

On comprend également pourquoi on ne peut pas choisir les gains K_p et/ou K_i trop grands. Si tel est le cas, K_p et/ou K_i sont alors du même ordre de grandeur que $\frac{1}{\varepsilon}$. Donc $\frac{d}{dt}X$ comportera des termes en $\frac{1}{\varepsilon}$. L'écriture sous la forme $\frac{d}{dt}X = F(X, Z)$ avec $\varepsilon \frac{d}{dt}Z = G(X, Z)$ n'est plus valable et l'approximation précédente peut être mise en défaut. On ne sait plus déduire de la stabilité de (1.21) une information sur celle de (1.28). Dans ce cas, la dynamique de la sonde, bien que stable isolément, n'est plus rapide devant les autres dynamiques. L'approximation n'est pas valable.

1.5.8 Intérêt de la pré-compensation (feedforward)

Pour améliorer les performances de la régulation de température (et donc le confort), on peut compléter le régulateur précédent en utilisant les techniques de *pré-compensation* et de *feedforward*.

Très souvent, on utilise aussi pour ajuster la position de la vanne trois voies, u , la température extérieure, θ_{ext} . Ainsi, u ne dépend pas seulement de la température intérieure θ et de sa consigne $\bar{\theta}$, mais aussi de θ_{ext} . Souvent, on substitue alors aux équations (1.19) les équations suivantes

$$\begin{cases} u(t) = S^{at}(K_p(\bar{\theta} - \bar{\theta}(t)) + I(t) - K_{ext}\theta_{ext}) \\ \frac{d}{dt}I(t) = K_i(\bar{\theta} - \theta(t)) + K_s(u(t) - K_p(\bar{\theta} - \theta(t)) - I(t) + K_{ext}\theta_{ext}) \end{cases}$$

avec K_{ext} un coefficient positif. Ainsi, même si $\theta = \bar{\theta}$, lorsque la température extérieure baisse, u a tendance à augmenter pour anticiper la prochaine baisse de θ . Le rajout de θ_{ext} permet d'anticiper les variations de température extérieure. C'est le terme de *pré-compensation*.

Une autre amélioration un peu plus subtile est celle qu'on utilise lors de changement de la consigne $\bar{\theta}$. Il est assez fréquent de programmer une consigne de température $\bar{\theta}$ un peu plus basse dans la journée, pendant les heures où les habitants sont absents, et un peu plus haute en début et en fin de journée. Ainsi $\bar{\theta}$ est une fonction constante par morceaux. Il est clair que la température θ ne peut pas suivre $\bar{\theta}$ à chacune de ses discontinuités : la réponse de notre système n'est pas instantanée ; notre modèle (1.20) implique que θ est une fonction dérivable de t et dont la dérivée est bornée. Notons $\dot{\theta}^{\max} > 0$ cette borne

$$\left| \frac{d}{dt}\theta(t) \right| \leq \dot{\theta}^{\max}$$

Il est naturel de modifier localement la fonction $\bar{\theta}(t)$ pour la transformer en une fonction continue dérivable par morceaux $\theta_r(t)$: là où $\bar{\theta}$ est constante, θ_r est aussi constante ; là où $\bar{\theta}$ admet un saut, alors θ_r est la fonction affine par morceau de pente inférieure en module à $\alpha\dot{\theta}^{\max}$ ($\alpha \leq 1$) qui remplace le saut par une rampe comme illustré sur la Figure (1.22). Ainsi, $\theta_r(t)$ est une approximation continue

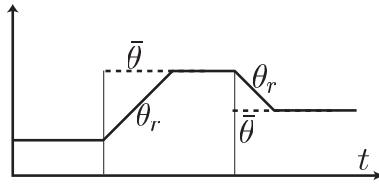


FIGURE 1.22 – Approximation de la consigne brute $t \mapsto \bar{\theta}(t)$, fonction constante par morceaux impossible à suivre à cause des discontinuités, par la référence $t \mapsto \theta_r(t)$, fonction continue et dérivable par morceaux correspondant à une trajectoire réalisable par le système.

de $\bar{\theta}(t)$, dérivable par morceaux et dont la dérivée reste toujours plus petite que $\dot{\theta}^{\max}$ en valeur absolue. En quelque sorte, $\theta_r(t)$ est une consigne de température variable dans le temps, proche de $\bar{\theta}$ tout en étant réellement réalisable par le système. On parle de *filtrage de consigne*. Il est facile de calculer en temps réel la fonction $\theta_r(t)$ qui lisse la consigne brute $\bar{\theta}(t)$.

Avec $\theta_r(t)$ au lieu de $\bar{\theta}(t)$, on rajoute alors au PI (1.19) la dérivée $\dot{\theta}_r$

$$\begin{cases} u(t) = S^{\text{at}}(K_p(\theta_r(t) - \theta(t)) + I(t) + K_r \dot{\theta}_r(t)) \\ \frac{d}{dt} I(t) = K_i(\theta_r(t) - \theta(t)) + K_s(u(t) - K_p(\theta_r(t) - \theta(t)) - I(t) - K_r \dot{\theta}_r(t)) \end{cases}$$

avec K_r un coefficient positif. L'effet de ce terme est le suivant. Supposons que pour $t < 0$, $\theta = \theta_r$. À partir de $t = 0$, θ_r commence une rampe de pente constante positive résultant d'une augmentation de la consigne brute $\bar{\theta}$ en $t = 0$. Alors, le terme en $\dot{\theta}_r$, dit de *feedforward* aura tendance à "booster" le contrôle u avec un surplus de chauffage pour compenser l'inertie thermique du bâtiment et ainsi avoir une température θ qui suivra mieux la référence réalisable. Il s'agit d'une anticipation sur la rampe de montée. De même, à la fin de la rampe de montée pour θ_r , $\dot{\theta}_r$ retourne à 0 et alors, le contrôle u aura un saut vers le bas, une sorte de freinage pour éviter que θ ne dépasse en fin de montée la valeur de θ_r (évite "l'over-shoot"). C'est le *feedforward*.

La combinaison de ses deux types d'anticipation, l'une sur la perturbation mesurée θ_{ext} , l'autre sur la consigne variable $\bar{\theta}$, conduit à l'algorithme *PI avec anticipation* suivant

$$\begin{cases} u(t) = S^{\text{at}} \left[K_p(\theta_r(t) - \theta(t)) + I(t) + K_r \dot{\theta}_r(t) - K_{ext} \theta_{ext} \right] \\ \frac{d}{dt} I(t) = K_i(\theta_r(t) - \theta(t)) \\ \quad + K_s \left[u(t) - K_p(\theta_r(t) - \theta(t)) - I(t) - K_r \dot{\theta}_r(t) + K_{ext} \theta_{ext} \right] \end{cases} \quad (1.29)$$

Des versions plus ou moins sophistiquées de cet algorithme sont utilisées dans les thermostats vendus dans le commerce.

1.5.9 Pré-compensation et suivi de trajectoires sur un système linéaire du premier ordre

Pour comprendre de façon plus formelle et plus précise le rôle de ces deux types d'anticipation, prenons comme modèle de contrôle, le modèle linéaire tangent autour d'un équilibre $(\bar{\theta}, \bar{u}, \bar{\theta}_{ext})$. On note $\delta\theta$ et δu les écarts à l'équilibre. La première variation de (1.20) autour de l'équilibre $\bar{\theta}$ et \bar{u} , donne un système du premier ordre stable. On va la calculer. Le modèle (1.20) est de la forme $\frac{d}{dt}\theta =$

$f(\theta, u, \theta_{ext})$ avec f fonction régulière de ces arguments et $f(\bar{\theta}, \bar{u}, \bar{\theta}_{ext}) = 0$. Il est d'usage de noter $\delta\theta$ par y , δu par u et $\delta\theta_{ext}$ par w .

$$\frac{d}{dt}y = -ay + bu + dw \quad (1.30)$$

avec

$$a = -\left.\frac{\partial f}{\partial \theta}\right|_{\bar{\theta}, \bar{u}, \bar{\theta}_{ext}}, \quad b = \left.\frac{\partial f}{\partial u}\right|_{\bar{\theta}, \bar{u}, \bar{\theta}_{ext}}, \quad d = \left.\frac{\partial f}{\partial \theta_{ext}}\right|_{\bar{\theta}, \bar{u}, \bar{\theta}_{ext}}$$

On remarque que $a > 0$, $b > 0$ et $d > 0$. Ainsi, le système est asymptotiquement stable en boucle ouverte. On peut interpréter la température y comme la sortie d'un filtre du premier ordre de constante de temps $\tau = 1/a$, excité par le signal $\tau(bu + dw)$ en notant

$$\frac{d}{dt}y = \frac{1}{\tau}(\tau bu + \tau dw - y)$$

Pour simplifier, nous considérons ici : $u^{\min} = -\infty$ et $u^{\max} = +\infty$.

Supposons qu'on souhaite suivre une référence $t \mapsto y_r(t)$ continue et dérivable par morceaux tout en connaissant à chaque instant la perturbation $w(t)$. On définit le contrôle de référence comme étant la fonction $t \mapsto u_r(t)$ continue par morceaux et telle que $\frac{d}{dt}y_r = -ay_r + bu_r + dw$, c'est à dire

$$u_r(t) = \frac{\dot{y}_r(t) + ay_r(t) - dw(t)}{b}$$

Pour obtenir u , on rajoute alors à u_r une correction de type PI sur l'erreur de suivi, $e = y - y_r$

$$u(t) = u_r(t) + K_p(y_r(t) - y(t)) + I(t), \quad \frac{d}{dt}I(t) = K_i(y_r(t) - y(t))$$

La partie anticipation (*feedforward*) correspond à u_r : elle compense, de façon cohérente avec le modèle, les effets transitoires dus aux variations de y_r et w . On obtient ainsi en boucle fermée un système différentiel autonome pour les variables $e = y - y_r$ et I

$$\frac{d}{dt}e = -(bK_p + a)e + bI, \quad \frac{d}{dt}I = -K_ie$$

Ce système s'écrit alors sous la forme d'un système du second ordre

$$\frac{d^2}{dt^2}e = -2\xi\omega_0 \frac{d}{dt}e - (\omega_0)^2 e \quad (1.31)$$

avec les notations usuelles

$$2\xi\omega_0 = (bK_p + a), \quad (\omega_0)^2 = bK_i$$

où $\omega_0 > 0$ est la *pulsation de coupure* et $\xi > 0$, le *facteur d'amortissement*. Ainsi, même si w et y_r varient au cours du temps, leurs variations sont pré-compensées par u_r qui est une combinaison linéaire de y_r , \dot{y}_r et w . On comprend mieux l'intérêt des termes $K_r\dot{y}_r$ et $K_{ext}w$ dans (1.29).

En pratique, les coefficients a , b et d ne sont connus qu'avec une certaine approximation. Si l'on note η_a , η_b et η_d les écarts entre leurs vraies valeurs (celles du modèle (1.30)) et celles utilisées pour le calcul de u_r par $\dot{y}_r = -(a - \eta_a)y_r + (b - \eta_b)u_r + (d - \eta_d)w$, alors la dynamique de l'erreur e devient

$$\frac{d^2}{dt^2}e = -2\xi\omega_0 \frac{d}{dt}e - (\omega_0)^2 e + \eta_b\dot{u}_r + \eta_d\dot{w} + \eta_a\dot{y}_r$$

En absence de terme de feedforward, le simple contrôleur PI

$$u = K_p(y_r - y) + I, \quad \frac{d}{dt}I = K_i(y_r - y)$$

conduit à un système du second ordre similaire pour $e = y - y_r$

$$\frac{d^2}{dt^2}e = -2\xi\omega_0 \frac{d}{dt}e - (\omega_0)^2 e - \frac{d^2}{dt^2}y_r + d\dot{w} + a\dot{y}_r \quad (1.32)$$

mais avec un terme source $-\frac{d^2}{dt^2}y_r + d\dot{w} - a\dot{y}_r$ a priori beaucoup plus grand que $\eta_b\dot{u}_r + \eta_d\dot{w} + \eta_a\dot{y}_r$ dans (1.31), puisque les erreurs paramétriques sont supposées peu importantes. Ainsi, même avec des erreurs de modélisation, le rajout de u_r conduit en général à une amélioration notable du suivi : e reste plus près de 0.

En l'absence de feedforward, on voit que la partie rétro-action, le “feedback”, doit alors corriger après coup les variations de la référence y_r et les perturbations w . Il en résulte une moins bonne précision dans le suivi de la référence y_r et des performances dégradées car on demande au feedback de compenser non seulement les erreurs de modèle, ce qui est son rôle premier, mais aussi les variations de y_r et w , ce qui est mieux réalisé par le contrôle feedforward u_r .

1.6 Cas d'étude : contrôle hiérarchisé et régulateurs en cascade

La théorie des systèmes lents/rapides (en particulier le Théorème 15 de Tikhonov) permet de justifier une pratique courante pour la synthèse de boucle de régulation : le *contrôle hiérarchisé* qui consiste à imbriquer les régulateurs. Un exemple suffira pour comprendre ce dont il s'agit. Cet exemple est non trivial car il s'agit d'un système du second ordre avec dynamique inconnue et des contraintes à la fois sur l'état et sur le contrôle. La synthèse d'un contrôleur prenant explicitement les contraintes d'état est un problème largement ouvert et pour lequel on ne dispose pas, à l'heure actuelle, de solution systématique et simple. Des solutions plus numériques fondées sur les techniques de commande prédictive peuvent être considérées mais elles sont coûteuses en temps de calculs et peuvent poser problème en particulier en ce qui concerne l'existence de solutions et la convergence numérique¹⁶. Le contrôleur que nous proposons ci-dessus s'appuie sur le paradigme des systèmes lents/rapides, sa structure est facile à comprendre. Le réglage des gains est simple et s'effectue à partir des contraintes. Nous donnons des formules explicites qui peuvent être utilisées.

Considérons un système du second ordre de la forme suivante

$$\frac{d^2}{dt^2}x = f(x, \frac{d}{dt}x) + u$$

où x et u sont dans \mathbb{R} . De tels modèles se rencontrent naturellement pour les systèmes mécaniques (par formulation variationnelle) : il s'agit par exemple de l'équation de Newton qui relie l'accélération aux forces extérieures parmi lesquelles se trouve le contrôle u . On suppose qu'on ne connaît pas bien $f(x, \frac{d}{dt}x)$.

16. La technique MPC (Model Predictive Control) propose une sélection optimale, au sens d'un critère choisi par l'utilisateur, parmi des trajectoires satisfaisant des contraintes. Elle suppose l'existence de telles trajectoires et l'unicité d'un optimum caractérisable numériquement. Ce sont ces hypothèses qui sont bien souvent difficiles à garantir sur des exemples concrets généraux. Si ces hypothèses sont vérifiées, la technique MPC consiste en la résolution itérée du problème d'optimisation avec mise à jour à chaque nouvelle période d'échantillonnage de la condition initiale. La fonction coût optimal est alors une fonction de Lyapounov, elle garantit la stabilité.

Il est possible de se ramener, au moins approximativement, au cas de deux systèmes du premier ordre. Le principe de la stratégie de contrôle hiérarchisé comporte deux étapes : un premier régulateur proportionnel (régulateur P) assure la convergence rapide de la vitesse $v = \frac{d}{dt}x$ vers une consigne \bar{v} . Un second régulateur PI cherche à stabiliser la position x à \bar{x} et, pour ce faire, met lentement à jour la consigne du régulateur de vitesse \bar{v} . Comme nous allons le voir, il est assez aisé de prendre en compte des contraintes sur la vitesse v et le contrôle u .

Écrivons l'équation du second ordre sous la forme d'un système de deux équations du premier ordre

$$\begin{cases} \frac{d}{dt}x = v \\ \frac{d}{dt}v = f(x, v) + u \end{cases} \quad (1.33)$$

Les contraintes sont $u \in [-u^{\max}, u^{\max}]$ et $v \in [-v^{\max}, v^{\max}]$ ($u^{\max}, v^{\max} > 0$). On considère un régime stationnaire caractérisé par un état admissible $(\bar{x}, \bar{v} = 0)$ et un contrôle admissible \bar{u} ; soit $|\bar{u}| < u^{\max}$ et $f(\bar{x}, 0) + \bar{u} = 0$. On cherche par un bouclage u à stabiliser (x, v) en (\bar{x}, \bar{u}) , en respectant les contraintes sur u et aussi sur v . Précisément, si la vitesse initiale v_0 appartient à $[-v^{\max}, v^{\max}]$ alors on souhaite que pour $t > 0$ la vitesse $v(t)$ du système en boucle fermée reste aussi dans $[-v^{\max}, v^{\max}]$. Cet objectif doit être atteint de manière robuste aux incertitudes sur la fonction $f(x, v)$.

Sans hypothèse supplémentaire sur f , il est difficile de donner une réponse un peu générale à ce problème. Nous allons supposer qu'il existe $0 < \alpha < 1$ tel que pour tout (x, v) ,

$$|f(x, v)| \leq \alpha u^{\max} \quad (1.34)$$

Cela veut dire que pour toute valeur de (x, v) , on peut choisir un contrôle u qui domine $f(x, v)$ et ainsi imposer le signe de $\frac{d^2}{dt^2}x$. Cette hypothèse est essentielle.

La cascade de deux régulateurs, évoquée ci-dessus, donne l'algorithme de contrôle suivant

$$\begin{cases} \frac{d}{dt}I = K_i(\bar{x} - x) + K_s(\bar{v} - K_p(\bar{x} - x) - I) \\ \bar{v} = S_v^{\text{at}}[K_p(\bar{x} - x) + I] \\ u = S_u^{\text{at}}[K_p(\bar{v} - v)/\epsilon] \end{cases} \quad (1.35)$$

avec $K_p > 0$, $K_i > 0$, $K_s > 0$, $K_s K_p > K_i$, et ϵ un paramètre positif. Les fonctions S_v^{at} et S_u^{at} sont les fonctions de saturation usuelles (telles que celle utilisée en (1.24)) associée aux contraintes sur v et u (projections sur les convexes $[-v^{\max}, v^{\max}]$ et $[-u^{\max}, u^{\max}]$).

La dynamique en boucle fermée est le système de trois équations différentielles non linéaires suivant

$$\begin{cases} \frac{d}{dt}x = v \\ \frac{d}{dt}v = f(x, v) + S_u^{\text{at}}[K_p(S_v^{\text{at}}[K_p(\bar{x} - x) + I] - v)/\epsilon] \\ \frac{d}{dt}I = K_i(\bar{x} - x) + K_s(S_v^{\text{at}}[K_p(\bar{x} - x) + I] - K_p(\bar{x} - x) - I) \end{cases}$$

Pour (x, v, I) proche de $(\bar{x}, 0, \bar{u})$, les contraintes ne sont pas actives et donc les fonctions S_v^{at} et S_u^{at}

sont les fonctions identités. Ainsi, le système bouclé ci-dessus devient

$$\begin{cases} \frac{d}{dt}x = v \\ \epsilon \frac{d}{dt}v = \epsilon f(x, v) + K_p(K_p(\bar{x} - x) + I) - v \\ \frac{d}{dt}I = K_i(\bar{x} - x) \end{cases}$$

Ce système est bien sous la *forme standard* du théorème 15 de Tikhonov : l'état lent est (x, I) , l'état rapide v , et la dynamique rapide est bien asymptotiquement stable. Le système lent

$$\frac{d}{dt}x = v, \quad \frac{d}{dt}I = K_i(\bar{x} - x)$$

est bien asymptotiquement stable. On conclut à la stabilité via le théorème 16. En l'absence de contraintes, la preuve de la stabilité repose uniquement sur le fait que ϵ est suffisamment petit.

Pour garantir la stabilité avec la prise en compte des effets non linéaires dus aux fonctions de saturation S_u^{at} et S_v^{at} , il convient de ne pas choisir K_p et K_i n'importe comment. Intuitivement, si notre raisonnement par échelles de temps est correct, il faut que le système rapide soit en mesure de suivre sa consigne $\bar{v} = S_v^{\text{at}}[K_p(\bar{x} - x) + I]$ même si cette dernière varie au cours du temps : cela entraîne qualitativement que $\frac{d}{dt}\bar{v}$ doit respecter les contraintes sur u , tout au moins sur la partie de u qui reste disponible après avoir dominé la dynamique $f(x, v)$. Cela veut dire approximativement que le module de $\frac{d}{dt}\bar{v}$ ne doit pas dépasser $(1 - \alpha)u^{\max}$. Il nous faut maintenant avoir une idée de l'ordre de grandeur de $\frac{d}{dt}\bar{v}$: pour cela on dérive dans le cas non saturé \bar{v} et on néglige la variation de I . Cela nous donne comme estimation de $\frac{d}{dt}\bar{v}$: $-K_p v$. Ainsi, il est naturel d'imposer $K_p \leq \frac{v^{\max}}{(1-\alpha)u^{\max}}$.

On constate, avec des simulations numériques qu'avec

$$K_p = \frac{v^{\max}}{(1 - \alpha)u^{\max}}, \quad K_i = \frac{K_p^2}{5}, \quad K_s = \frac{2K_p}{5}, \quad \epsilon = 3,$$

le contrôleur (1.35) est bien stabilisant tout en respectant les contraintes sur u et approximativement les contraintes sur v .

Ce qui précède n'est pas une preuve formelle mais plutôt une illustration de la démarche fondée sur la théorie des perturbations pour concevoir un régulateur simple pour un système du second ordre dont le modèle comporte d'importantes incertitudes que l'on peut dominer par un contrôle borné. De plus ce régulateur (1.35) prend en compte la contrainte d'état $|v| \leq v^{\max}$.

Chapitre 2

Fonctions de transfert

Les résultats du Chapitre 1 sont essentiellement de nature qualitative. La stabilité d'un système est garantie sous l'hypothèse qu'un paramètre ϵ soit suffisamment petit. Dans ces conditions, on peut alors raisonner sur un système simplifié (nominal). La question de la *robustesse* a été traitée de manière perturbative : si le *système nominal* en boucle fermée ou en boucle ouverte admet un *équilibre asymptotiquement stable*, alors le *système perturbé* (régulièrement et/ou singulièrement)¹ admet un équilibre voisin de celui du système nominal, équilibre qui est lui aussi asymptotiquement stable.

La question de savoir à partir de quelle valeur critique $\bar{\epsilon}$ de ϵ il y a perte de stabilité est importante, notamment d'un point de vue pratique. Comme nous allons le voir dans ce chapitre, une réponse est liée aux notions de *marges de robustesse* que nous allons définir. On notera que, d'un point de vue formel, l'étude des changements qualitatifs de comportement d'un système dynamique en fonction de paramètres intervenant dans les équations est l'objet de la *théorie des bifurcations et des catastrophes* [21]. Cette théorie dépasse largement le cadre de ce cours.

Les *marges de robustesse* sont des notions difficiles à aborder de façon générale pour un système non linéaire. Cependant, il est possible, une fois de plus, d'obtenir des informations précieuses en considérant le *système linéarisé tangent* autour de l'équilibre.

On utilise sur ce système linéaire le calcul symbolique², i.e. on remplace l'opérateur $\frac{d}{dt}$ par la variable de Laplace $s \in \mathbb{C}$ (en utilisant la *transformée de Laplace* voir [55]). Grâce à ce formalisme³, on peut faire des calculs algébriques pour aboutir à une équation reliant entrée et sortie. Le système est alors représenté par une fonction de la variable complexe s (notée aussi p et correspondant enfin à ω en théorie des circuits). Cette fonction est appelée *fonction de transfert*. L'analyse complexe fournit une réponse directe (par le Théorème de Cauchy) à la définition des marges de robustesse.

1. On parle de perturbations régulières lorsque ϵ apparaît dans les membres de droite des équations différentielles, comme par exemple, $\frac{d}{dt}x = f(x, z, \epsilon)$ et $\frac{d}{dt}z = g(x, z, \epsilon)$. On parle de perturbations singulières lorsque ϵ apparaît à gauche devant une dérivée en temps, comme par exemple $\frac{d}{dt}x = f(x, z)$ et $\epsilon \frac{d}{dt}z = g(x, z)$. Ainsi le système $\frac{d}{dt}x = f(x, z, \epsilon)$ et $\epsilon \frac{d}{dt}z = g(x, z, \epsilon)$ est à la fois régulièrement et singulièrement perturbé. Ses solutions sont alors proches de celles du système nominal (obtenu en faisant $\epsilon = 0$, $\frac{d}{dt}x = f(x, z, 0)$ et $0 = g(x, z, 0)$) sous certaines hypothèses comme celles du théorème 15.

2. Le calcul symbolique est aussi dénommé calcul opérationnel [78, 22] en référence à la théorie des opérateurs linéaires [45].

3. Nous laissons de côté les problèmes de convergence des transformées de Laplace qui sont relativement hors-sujet ici et qui sont abordés dans les cours [54, 55]. On pourra aussi se reporter à [15] pour un exposé mettant en lumière le rôle des bandes de convergence dans le plan complexe.

2.1 Passage à la fonction de transfert

Nous allons considérer l'exemple (1.30) de la fin du chapitre précédent

$$\frac{d}{dt}y(t) = -ay(t) + bu(t) + dw(t) \quad (2.1)$$

avec a, b et d positifs et constants. Ce système est linéaire. Le contrôle est u , la perturbation w et la mesure y . On a vu, toujours dans le chapitre précédent, qu'un simple⁴ régulateur PI

$$u = K_p(y_r - y) + I, \quad \frac{d}{dt}I = K_i(y_r - y) \quad (2.2)$$

avec $K_p, K_i > 0$ assure la convergence asymptotique de $y(t)$ vers la consigne (référence) y_r supposée constante, pour peu que la perturbation w soit constante. Rappelons qu'on n'a pas besoin de connaître cette constante pour prouver que de manière asymptotique $y = y_r$.

2.1.1 Questions de robustesse

Étant donné le système en boucle fermée (2.1)-(2.2), nous souhaitons étudier les trois points suivants

1. Que se passe-t-il si la mesure de y n'est pas instantanée mais est effectuée avec un certain temps de retard Δ ? Dans une telle configuration, à l'instant t , on connaît la valeur de $y(t - \Delta)$. La valeur $y(t)$ n'est disponible, pour calculer u , qu'à l'instant $t + \Delta$. Ce retard est par exemple imputable à la dynamique du capteur, à la chaîne informatique temps-réel, à des délais d'échantillonnage de certains capteurs, ou à des délais de communication de l'information.
2. Que se passe-t-il si la perturbation non mesurée w n'est plus constante mais sinusoïdale de la forme $w(t) = \cos(\omega t)$ avec ω constant. Ne peut-il pas y avoir propagation de ces oscillations? Est-ce gênant?
3. Que se passe-t-il si l'on superpose les deux points précédents, un retard de mesure et une perturbation sinusoïdale?

2.1.2 Principe des calculs

De manière préliminaire, nous montrons ici comment obtenir la *fonction de transfert* d'un système. Le principe des calculs est très simple : il suffit de remplacer l'opérateur $\frac{d}{dt}$ par s . Par abus de notation et soucis de simplicité, on note ici $y(s)$ la transformée de Laplace de $t \mapsto y(t)$: ainsi $y(s) = \int_0^{+\infty} e^{-st} y(t) dt$. Sauf indication contraire, on travaille toujours en Laplace unilatéral par référence à des problèmes de Cauchy avec des conditions initiales nulles en $t = 0$. Dans ce cas, l'opérateur retard devient $e^{-\Delta s}$ où Δ est le retard⁵

-
4. Nous ne prenons pas en compte les contraintes sur u ici.
 5. On peut d'abord utiliser le calcul

$$\int_0^{+\infty} e^{-st} y(t - \Delta) dt = \int_0^{+\infty} e^{-s(\tau + \Delta)} y(\tau) d\tau = e^{-s\Delta} y(s)$$

où y doit être nulle pour les temps t négatifs. On a aussi formellement (sans chercher à le justifier) le développement en série $y(t - \Delta) = \sum_{k=0}^{+\infty} \frac{(-\Delta)^k}{k!} y^{(k)}(t)$. Comme $y^{(k)} = \frac{d^k}{dt^k} y = s^k y$, on voit naturellement apparaître la série $\sum_{k=0}^{+\infty} \frac{(-\Delta)^k}{k!} s^k$ qui n'est autre que $e^{-\Delta s}$. Ainsi, en calcul symbolique, l'opérateur qui à la fonction y associe la fonction y retardée de Δ n'est autre qu'une multiplication par $e^{-\Delta s}$. On trouvera toutes les formules nécessaires à des calculs opérationnels plus compliqués dans [22].

Considérons le système bouclé avec y_r et w variables et un retard Δ sur la mesure de y

$$\begin{cases} \frac{d}{dt}y(t) = -ay(t) + bu(t) + dw(t) \\ u(t) = K_p(y_r(t) - y(t - \Delta)) + I(t) \\ \frac{d}{dt}I(t) = K_i(y_r(t) - y(t - \Delta)) \end{cases}$$

On obtient alors (dans le domaine des transformées de Laplace)

$$\begin{cases} sy = -ay + bu + dw \\ u = K_p(y_r - e^{-\Delta s}y) + I \\ sI = K_i(y_r - e^{-\Delta s}y) \end{cases}$$

On peut alors calculer algébriquement y , u , et I en fonction de y_r et w . Il suffit de résoudre ces équations par rapport aux variables recherchées. La sortie y est une combinaison linéaire de y_r et w , les coefficients étant des fractions rationnelles en s et $e^{-\Delta s}$. Tout calcul fait, on obtient

$$y = \frac{b(K_i + sK_p)}{s(s + a) + b(K_i + sK_p)e^{-\Delta s}} y_r + \frac{sd}{s(s + a) + b(K_i + sK_p)e^{-\Delta s}} w$$

La stabilité se déduit des zéros des dénominateurs (nommés *pôles*), i.e., les valeurs $s \in \mathbb{C}$ pour lesquelles la formule ci-dessus donnant y n'est pas définie. Nous reviendrons sur ce point dans la Définition 7 et la Proposition 18.

S'il existe $\alpha > 0$, tel que les valeurs de s pour lesquelles la fonction de la variable complexe (c'est une fonction dite entière car elle est définie pour toute valeur de $s \in \mathbb{C}$)

$$D(s) = s(s + a) + b(K_i + sK_p)e^{-\Delta s} \quad (2.3)$$

s'annule, sont toutes dans le demi plan $\Re(s) \leq -\alpha$, alors le système est asymptotiquement stable en boucle fermée. En revanche, si un des zéros est à partie réelle strictement positive, le système est instable et la sortie $y(t)$ n'est pas bornée lorsque $t \rightarrow +\infty$.

Cette caractérisation est une généralisation naturelle de celle que nous avons déjà vue pour les systèmes linéaires (voir Théorème 4) sous forme d'état $\frac{d}{dt}x = A(x - \bar{x})$: pour ce dernier système, le point d'équilibre \bar{x} est asymptotiquement stable en boucle ouverte lorsque les zéros de $\det(sI - A)$ sont tous à partie réelle strictement négative. La formule donnant la transformée de Laplace de x explicitement à partir de celle de \bar{x} (à conditions initiales nulles)

$$x = -(sI - A)^{-1}A\bar{x}$$

n'est pas définie pour les valeurs de s appartenant au spectre de A . Comme $\det(sI - A)$ est un polynôme, il ne possède qu'un nombre fini de zéros d'où l'existence de $\alpha > 0$.

Pour des fonctions de s plus compliquées comme les polynômes en s et $e^{-\Delta s}$ (appelés quasi-polynômes [31]), on a un nombre infini de zéros en général. La localisation des zéros de la fonction entière $D(s)$, définie par (2.3), est un problème ardu et non complètement résolu à l'heure actuelle (voir [29] pour un exposé complet et la Section 2.4.1 pour un exemple de résultats qu'on peut obtenir).

Pour $\Delta = 0$, $D(s) = s^2 + (bK_p + a)s + bK_i$ n'a que deux zéros et ils sont à partie réelle strictement négative (sous l'hypothèse $K_p, K_i, b > 0$ et $a \geq 0$). Nous verrons comment il est possible de calculer le retard critique Δ^* à partir duquel $D(s)$ admet nécessairement des zéros instables rendant le système en boucle fermée instable.

2.1.3 Régime asymptotique forcé

Lorsqu'un système est asymptotiquement stable en boucle fermée, on peut étudier comment il réagit à des perturbations (on dit aussi excitations) sinusoïdales w . Comme le précise la Proposition 19, un système asymptotiquement stable soumis à une telle excitation, converge, après un transitoire, vers un régime périodique de même période. Le système oublie sa condition initiale (de manière asymptotique), comme discuté à la Section 1.1.3.

Pour une excitation périodique $w = \cos(\omega t)$, on passe en complexes en notant $w = \exp(i\omega t)$ et on cherche la solution (y, u, I) , pour $y_r = 0$, des équations

$$\begin{cases} \frac{d}{dt}y(t) = -ay(t) + bu(t) + d \exp(i\omega t) \\ u(t) = K_p(y_r(t) - y(t - \Delta)) + I(t) \\ \frac{d}{dt}I(t) = K_i(y_r(t) - y(t - \Delta)) \end{cases}$$

sous la forme $y(t) = \mathcal{Y} \exp(i\omega t)$, $u(t) = \mathcal{U} \exp(i\omega t)$ et $I(t) = \mathcal{I} \exp(i\omega t)$. Cela revient à faire exactement les mêmes manipulations que celles faites à la Section 2.1.2 en remplaçant $\frac{d}{dt}$ par s et le retard par $e^{-\Delta s}$, mais avec $s = i\omega$. On obtient

$$\mathcal{Y} = \left. \frac{sd}{s(s+a) + b(K_i + sK_p)e^{-\Delta s}} \right|_{s=i\omega}$$

Ce nombre complexe peut s'écrire sous la forme

$$\left. \frac{sd}{s(s+a) + b(K_i + sK_p)e^{-\Delta s}} \right|_{s=i\omega} = G(\omega) \exp(i\phi(\omega))$$

où $G(\omega) \geq 0$ est le *gain* et $\phi(\omega) \in [-\pi, \pi[$, la *phase*. Il suffit alors de prendre la partie réelle de la formule

$$y(t) = G(\omega) \exp(i(\omega t + \phi(\omega)))$$

pour avoir le régime forcé y . La solution périodique du système en boucle fermée, avec $y_r = 0$ et $\bar{w} = \cos(\omega t)$, est

$$y(t) = G(\omega) \cos(\omega t + \phi(\omega))$$

On représente sur deux graphiques, appelés *diagrammes de Bode*⁶, le gain $G(\omega)$ en fonction de ω et le déphasage $\phi(\omega)$ en fonction de ω . Une grande variation locale de G en fonction de ω est la signature de *résonances*, dont les fréquences caractéristiques sont données approximativement par les maxima locaux de G (les *pics de résonance*). Le diagramme de Bode peut s'obtenir expérimentalement (jusqu'à une certaine fréquence au delà de laquelle les instruments sont inutilisables) avec un analyseur de fréquence si le système est effectivement asymptotiquement stable.

2.1.4 Simplifications pôles-zéros

Il y a une distinction entre la stabilité d'un système représenté sous forme d'état $\frac{d}{dt}x = Ax + Bu$, $y = Cx$, et celle de la relation entrée-sortie $y(s) = H(s)u(s)$.

6. En général, on trace, sur le premier graphique $20 \log_{10}(G(\omega))$ (décibels comme échelle des ordonnées) en fonction de $\log_{10}(\omega)$ (\log_{10} comme échelle en abscisse), sur le second $\phi(\omega)$ en fonction de $\log_{10}(\omega)$.

Il peut y avoir des *simplifications de pôles avec des zéros*. En d'autres termes, il se peut qu'on ait une racine commune entre le numérateur et le dénominateur de $H(s)$. Ainsi, $H(s)$ peut ne posséder que des pôles stables alors que A peut avoir des valeurs propres instables. On verra que la Définition 7 et la Proposition 18 tiennent compte de ce phénomène.

Prenons un exemple très simple qui fera bien comprendre le problème. Supposons que nous ayons le système suivant

$$\frac{d^2}{dt^2}y - y = \frac{d}{dt}u - u \quad (2.4)$$

Le calcul symbolique déjà utilisé précédemment donne $(s^2 - 1)y = (s - 1)u$, soit

$$y = \frac{s - 1}{s^2 - 1}u = \frac{1}{s + 1}u$$

On pourrait en conclure que le système est asymptotiquement stable car la relation entrée-sortie ne fait apparaître qu'un seul pôle $s = -1$ à partie réelle négative. Ceci est faux : le système est du second ordre, il admet une autre valeur propre. Cette dernière est instable et vaut 1. Pour comprendre l'origine du problème engendré par la simplification par le facteur commun $s - 1$ entre le numérateur et le dénominateur de la fonction de transfert, il faut reprendre la transformation de Laplace (voir [54]) et bien noter que, si on ne néglige plus la condition initiale, la transformation de Laplace de $\frac{d}{dt}y$ donne en fait $sy - y(0)$ au lieu de sy comme nous l'avons fait, pour simplifier les calculs. Ainsi, tous nos calculs sont corrects si à $t = 0$ tout est à l'équilibre et tout est nul, c.-à-d. $y_0 = 0$, $\dot{y}_0 = 0$ et $u_0 = 0$. Si tel n'est pas le cas, il faut utiliser les formules où les valeurs en $t = 0$ de (y, \dot{y}) et de u apparaissent

$$\ddot{y}(s) = s^2y(s) - sy_0 - \dot{y}_0, \quad \dot{u}(s) = su(s) - u_0.$$

Elles se démontrent avec deux intégrations par partie : une pour $\int_0^{+\infty} e^{-st}\ddot{y}(t)dt$ et une autre pour $\int_0^{+\infty} e^{-st}\dot{u}(t)dt$. On suppose aussi que $\ddot{y}(t)$ et $\dot{u}(t)$ sont continus par rapport à t et que les intégrales sont absolument convergentes. Ainsi on a

$$(s^2 - 1)y - sy_0 - \dot{y}_0 = (s - 1)u - u_0$$

où y_0 , \dot{y}_0 et u_0 sont les valeurs à $t = 0$ de y , $\frac{d}{dt}y$ et u . En effet on a

Dans ce cas, on obtient le calcul complet

$$y(s) = \frac{1}{s + 1}u + \frac{sy_0 + \dot{y}_0 - u_0}{s^2 - 1}$$

Cette équation révèle le transfert de u à y et celui des conditions initiales à y . La simplification initialement envisagée n'est plus possible, sauf pour des valeurs très particulières de y_0 , \dot{y}_0 , et u_0 . Le second terme dans la fonction de transfert correspond à une fonction qui diverge lorsque $t \rightarrow +\infty$. Il est nul si et seulement si on a $\dot{y}(0) = y(0) - u(0) = 0$. En général, cette condition n'est pas satisfaite, le système est instable à cause du pôle en $s = 1$ qui ne disparaît plus.

Dans la suite du chapitre, nous supposerons qu'il n'y a pas de canular de ce type. En pratique, ce genre de simplification se voit très vite en inspectant les équations. Il suffit de modifier un tout petit peu les coefficients, par exemple prendre $\dot{u} - 0.999u$ pour casser cette simplification et revoir apparaître le pôle instable qui avait disparu. Ce type de simplification est surtout gênant quand des pôles instables sont en jeu. Pour des pôles stables, c'est moins embêtant et ce d'autant plus s'ils ont des parties réelles fortement négatives. Les termes correspondants convergent de toutes façons très rapidement vers zéro lorsque $t \rightarrow +\infty$.

2.1.5 Formalisme

Dans ce qui suit, on s'attache à donner des éléments précis (définitions et propriétés) concernant les fonctions de transfert que nous avons introduites à partir d'exemples simples dans les Sections 2.1.2, 2.1.3 et 2.1.4. On considère un système avec $m \geq 1$ entrées u et $p \geq 1$ sorties y sous forme d'état

$$\frac{d}{dt}x = Ax + Bu, \quad y = Cx \quad (2.5)$$

où A est une matrice $n \times n$ à coefficients constants, B est une matrice (dans le cas où on a plusieurs entrées, sinon c'est un vecteur colonne) de taille $n \times m$ à coefficients constants, et C est une matrice (dans le cas où on a plusieurs sorties, sinon c'est un vecteur ligne) de taille $p \times n$ à coefficients constants.

La notion de stabilité dont nous avons parlé pour les fonctions de transfert correspond en fait à la définition suivante.

Définition 7 (Stabilité Entrée-Bornée-Sortie-Bornée). *Un système de la forme (2.5) est stable Entrée-Bornée-Sortie-Bornée (on dit EBSB), si pour toute entrée bornée $t \mapsto u(t)$ la sortie $t \mapsto y(t)$ reste bornée.*

Définition 8 (Matrice et fonction de transfert). *On appelle matrice de transfert du système (2.5) la matrice dépendant de la variable complexe*

$$\mathbb{C} \ni s \mapsto H(s) = C(sI - A)^{-1}B$$

. Les éléments de cette matrice sont des fractions rationnelles en s strictement propres. Lorsque $p = m = 1$ (système à une entrée et une sortie) $H(s)$ est une simple fraction rationnelle en s et on parle alors de fonction de transfert.

Une fraction rationnelle $\frac{P(s)}{Q(s)}$ est dite *causale* ou *propre* quand le degré du numérateur $P(s)$ est plus petit que (ou égal à) celui du dénominateur $Q(s)$. Lorsque l'inégalité est stricte on dit que la fraction rationnelle est *strictement causale* ou *strictement propre*.

Définition 9 (Pôles et zéros d'une fonction de transfert). *On appelle zéro d'une fonction de transfert (scalaire) $H(s)$ les racines de son numérateur. On appelle pôles les racines de son dénominateur.*

Il est facile de montrer le résultat suivant à partir de la formule de variation des constantes (1.14), page 23, donnant les solutions de (2.5) (prendre $b(t) = Bu(t)$).

Théorème 18 (Stabilité EBSB d'un système linéaire)

Si les valeurs propres de A sont toutes à partie réelle strictement négative, alors le système (2.5) est stable EBSB.

Ainsi la *stabilité asymptotique* de $\frac{d}{dt}x = Ax$ implique la stabilité EBSB mais l'inverse est faux comme on peut s'en convaincre simplement avec le système où $n = 2$ et $m = p = 1$:

$$\frac{d}{dt}x_1 = \lambda x_1 + u, \quad \frac{d}{dt}x_2 = -x_2 + u, \quad y = x_2.$$

Ce système est bien EBSB mais pour $\lambda > 0$, la matrice A est exponentiellement instable. Un simple calcul de $H(s) = C(sI - A)^{-1}B$ montre que le facteur $(s - \lambda)$ apparaissant dans les dénominateurs de $(sI - A)^{-1}$ disparaît du transfert entre y et u . Plus généralement, la réciproque de la proposition 18 est fausse à cause de *simplifications pôles-zéros* dans le calcul de la fonction de transfert $C(sI - A)^{-1}B$. Ces cas pathologiques sont dus au fait que l'état x peut ne pas être observable à partir de la sortie y (cf théorème 28) et aussi au fait que l'état x peut ne pas être commandable avec l'entrée u (cf théorème 22).

On a, de manière générale, la propriété suivante qui permet de caractériser la réponse asymptotique d'un système stable EBSB à une entrée sinusoïdale.

Théorème 19 (Réponse asymptotique d'un système EBSB à une entrée sinusoïdale)

Considérons le système (2.5) avec $m = p = 1$ et supposons le stable EBSB. Notons $H(s)$ sa fonction de transfert. Si on excite ce système par un signal d'entrée $u(t) = \cos(\omega t + \varphi)$ de pulsation ω et de phase φ constantes, alors, de manière asymptotique lorsque $t \rightarrow +\infty$, la sortie $y(t)$ converge vers le signal $\Re(H(i\omega)e^{i(\omega t+\varphi)}) = |H(i\omega)| \cos(\omega t + \varphi + \arg H(i\omega))$.

On notera ainsi que les pulsations correspondant aux zéros de $\mathbb{R} \ni \omega \mapsto H(\omega)$ sont asymptotiquement rejetées. Si $H(s)$ a un zéro en zéro (c'est souvent le cas lorsqu'on recourt à un *contrôleur PI*), alors les entrées constantes sont rejetées. On a *rejet de l'erreur statique*.

2.1.6 Du transfert vers l'état : réalisation

La donnée d'un transfert rationnel $H(s)$ est équivalente à la donnée d'une *forme d'état canonique minimale* que nous allons définir. C'est ce qu'on appelle la *réalisation sous forme d'état d'un transfert rationnel*.

De manière préliminaire, nous allons traiter un cas particulier qui permet de comprendre l'algorithme général de construction du système d'état.

Exemple 12 (Réalisation d'un système d'ordre 2). *Supposons que nous disposions du transfert suivant entre y et u*

$$y(s) = \frac{b_1 s + b_0}{s^2 + a_1 s + a_0} u(s)$$

où (a_0, a_1, b_0, b_1) sont des paramètres. Ce transfert correspond à l'équation différentielle du second ordre suivante

$$\frac{d^2}{dt^2}y + a_1 \frac{d}{dt}y + a_0 y = b_1 \frac{d}{dt}u + b_0 u$$

On pose $z = (z_1, z_2) = (y, \frac{d}{dt}y)$ pour obtenir

$$\frac{d}{dt}z_1 = z_2, \quad \frac{d}{dt}z_2 = -a_0 z_1 - a_1 z_2 + b_1 \frac{d}{dt}u + b_0 u$$

On regroupe les dérivées dans la seconde équation en une seule variable $x_2 = z_2 - b_1 u$ au lieu de z_2 . Avec les variables $(x_1 = z_1, x_2)$ au lieu des variables (z_1, z_2) , on a la réalisation d'état souhaitée

$$\frac{d}{dt}x_1 = x_2 + b_1 u, \quad \frac{d}{dt}x_2 = -a_0 x_1 - a_1 x_2 + (b_0 - a_1 b_1)u, \quad y = x_1$$

On notera les matrices

$$A = \begin{pmatrix} 0 & 1 \\ -a_0 & -a_1 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_0 - a_1 b_1 \end{pmatrix}, \quad C = (1 \quad 0) \quad (2.6)$$

On voit que l'astuce est de faire disparaître par des changements astucieux sur l'état, les dérivées de u .

Supposons maintenant que nous ayons eu un degré de plus au numérateur

$$y(s) = \frac{b_2 s^2 + b_1 s + b_0}{s^2 + a_1 s + a_0} u(s)$$

Alors, en reprenant les calculs précédents, on a avec $z = (y, \frac{d}{dt}y)$

$$\frac{d}{dt}z_1 = z_2, \quad \frac{d}{dt}z_2 = -a_0 z_1 - a_1 z_2 + b_2 \frac{d^2}{dt^2}u + b_1 \frac{d}{dt}u + b_0 u$$

On pose $\tilde{z}_2 = z_2 - b_2 \frac{d}{dt}u$. Avec les variables (z_1, \tilde{z}_2) , la plus haute dérivée de u , $\frac{d^2}{dt^2}u$, disparaît

$$\frac{d}{dt}z_1 = \tilde{z}_2 + b_2 \frac{d}{dt}u, \quad \frac{d}{dt}\tilde{z}_2 = -a_0 z_1 - a_1 \tilde{z}_2 + (b_1 - a_1 b_2) \frac{d}{dt}u + b_0 u$$

On regroupe les dérivées dans les deux variables suivantes :

$$x_1 = z_1 - b_2 u, \quad x_2 = \tilde{z}_2 - (b_1 - a_1 b_2) u$$

On obtient alors la forme d'état souhaitée (avec une loi de sortie qui dépend directement de u)

$$\frac{d}{dt}x_1 = x_2 + (b_1 - a_1 b_2) u, \quad \frac{d}{dt}x_2 = -a_0 x_1 - a_1 x_2 + (b_0 - a_0 b_2 - a_1(b_1 - a_1 b_2)) u$$

car $y = x_1 + b_2 u$.

D'après l'exemple précédent, on comprend ainsi que, par éliminations successives des plus hautes dérivées de u en rajoutant des dérivées de u dans les composantes de l'état, il est toujours possible, lorsque la fraction rationnelle $y(s)/u(s) = G(s)$ est propre, d'obtenir la forme d'état suivante

$$\frac{d}{dt}x = Ax + Bu, \quad y = Cx + Du$$

où $\dim(x)$ est égal au degré du dénominateur de $G(s)$

$$G(s) = (C(sI - A)^{-1}B + D)u$$

Cette méthode de réalisation se généralise aux systèmes multi-variables avec $m > 1$ entrées u et $p > 1$ sorties y , le transfert entre u et y étant donné (en accord avec la Définition 8), par une matrice dont les éléments sont des fractions rationnelles en s . Un lecteur intéressé par un exposé plus formel de la théorie de la réalisation pourra consulter [39].

Plusieurs réalisations différentes peuvent conduire au même transfert. Une réalisation différente de celle proposée par (2.6) peut être obtenue avec le schéma général de la Figure 2.1 qui fournit une réalisation d'une fonction de transfert strictement propre

$$\frac{b_{n-1}s^{n-1} + \dots + b_0}{s^n + a_{n-1}s^{n-1} + \dots + a_0}$$

ainsi que les *formes canoniques*

$$\dot{x} = Ax + Bu, \quad y = Cx$$

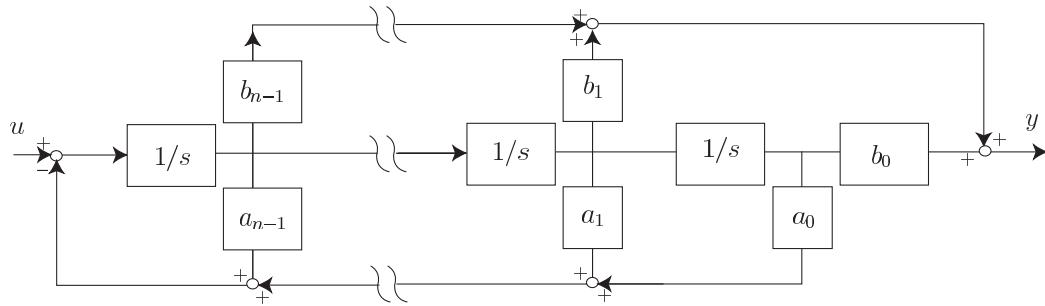


FIGURE 2.1 – Réalisation d'une fonction de transfert strictement propre.

1. (A_1, B_1, C_1) avec

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 1 \\ -a_0 & -a_1 & \dots & \dots & -a_{n-1} \end{pmatrix}, B_1 = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$$

$$C_1 = (b_0 \ \dots \ b_{n-1})$$

2. (A_2, B_2, C_2) avec

$$A_2 = \begin{pmatrix} 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & -a_{n-1} \end{pmatrix}, B_2 = \begin{pmatrix} b_0 \\ \dots \\ b_{n-1} \end{pmatrix}$$

$$C_2 = (0 \ \dots \ 0 \ 1)$$

2.2 Schémas blocs et fonctions de transfert

Les schémas blocs sont un langage relativement universel. Ils permettent d'exprimer l'action sur un système des commandes, des perturbations, et d'expliquer des schémas de régulation. On présente ici, pour une structure très générale de contrôle (voir Figure 2.5), les fonctions de transfert en boucle fermée.

2.2.1 De la forme d'état vers le transfert

Reprendons les notations d'un système non linéaire telles que nous les avons vues à la Section 1.4.3. On considère un système tout à fait général décrit par

$$\frac{d}{dt}x = f(x, u, w, p), \quad \mathbb{R}^n \ni x \mapsto y = h(x) \in \mathbb{R}^{m < n}$$

où p représente des paramètres, u représente des commandes et w représente des perturbations. On dispose de capteurs qui fournissent à chaque instant la valeur y (en général $\dim(y) < \dim(x)$ c.-à-d. que l'état du système n'est pas complètement mesuré). Les dimensions de x , u , w et y sont ici arbitraires. Ce système est représenté par le schéma blocs de la Figure 2.2.

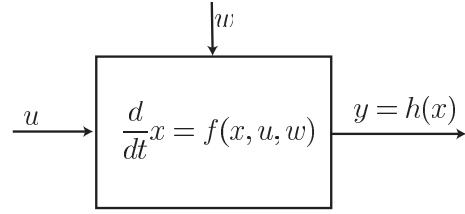


FIGURE 2.2 – Le schéma blocs de type fonctionnel pour le système en boucle ouverte avec comme contrôle u , comme perturbation w et comme sortie (mesure) y .

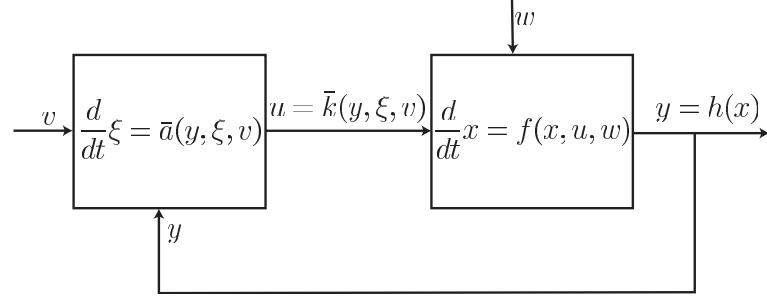


FIGURE 2.3 – Le schéma blocs de type fonctionnel pour le système en boucle fermée où le nouveau contrôle est v (typiquement la consigne que y doit suivre).

On élabore une loi de contrôle sous la forme d'un *retour dynamique de sortie*, c.-à-d. un système dynamique (comme l'est un régulateur PI)

$$\frac{d}{dt}\xi = \bar{a}(y, \xi, v), \quad u = \bar{k}(y, \xi, v)$$

où v est le nouveau contrôle. On complète alors le schéma blocs en boucle ouverte de la Figure 2.2 par une boucle de rétro-action en montrant bien que les informations issues de la sortie y , sont utilisées de manière *causale* pour ajuster le contrôle u de façon à atteindre un objectif précis. Typiquement, le nouveau contrôle v sera la consigne que devra suivre y (lorsque v et y ont la même dimension) et u sera calculé de façon à ce que y suive v . On obtient ainsi le schéma en boucle fermée de la Figure 2.3.

Supposons que l'on soit proche d'un point d'équilibre $(\bar{x}, \bar{y}, \bar{u}, \bar{w}, \bar{\xi}, \bar{v})$. On peut linéariser toutes les équations ci-dessus, celle du système et celle du contrôleur, et ainsi étudier de façon plus formelle et approfondie la dynamique du système en boucle ouverte et/ou en boucle fermée. On supposera ici pour simplifier que $\dim(u) = 1$, $\dim(y) = 1$ et que $v = y_r$ est la consigne que doit suivre y (problème de régulation mono-variable). On notera, par souci de simplicité, (x, y, u, w, ξ, v) les écarts à l'équilibre $(\bar{x}, y_r, \bar{u}, \bar{w}, \bar{\xi}, \bar{v})$.

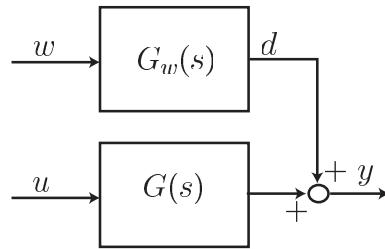


FIGURE 2.4 – Le schéma blocs en boucle ouverte.

2.2.2 Transfert avec perturbation et bouclage

Les équations du *système linéarisé tangent* s'écrivent

$$\begin{cases} \frac{d}{dt}x = Ax + Bu + Dw \\ y = Cx \\ u = Ly + M\xi + Nv \\ \frac{d}{dt}\xi = F\xi + Gy + Hv \end{cases}$$

où les matrices $A, B, D, C, L, M, N, F, G$, et H s'obtiennent à partir des dérivées partielles de f, h, \bar{k} et \bar{a} . Par le calcul symbolique, on remplace $\frac{d}{dt}$ par s pour obtenir le système d'équations algébriques suivant

$$\begin{cases} sx = Ax + Bu + Dw \\ y = Cx \\ u = Ly + M\xi + Nv \\ s\xi = F\xi + Gy + Hv \end{cases}$$

On résoud simplement $x = (sI - A)^{-1}(Bu + Dw)$ et on forme alors

$$y = G(s)u + G_w(s)w$$

avec $G(s)$ et $G_w(s)$ les fractions rationnelles suivantes

$$G(s) = C(sI - A)^{-1}B, \quad G_w = C(sI - A)^{-1}D$$

Ceci nous donne le schéma blocs du transfert en boucle ouverte de la Figure 2.4. Comme $\xi = (sI - F)^{-1}(Gy + Hv)$, on a

$$u = [L + M(sI - F)^{-1}G]y + [N + M(sI - F)^{-1}H]v$$

Pour un problème de régulation on note

$$v = y_r$$

et on choisit $L = -N$ et $G = -H$. On a donc

$$u = K(s)(y_r - y) = K(s)e$$

avec $K(s)$ la fraction rationnelle

$$K(s) = N + M(sI - F)^{-1}H$$

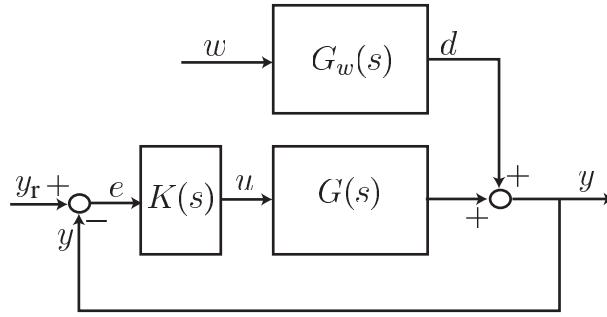


FIGURE 2.5 – Le schéma blocs en boucle fermée ; configuration normale d'une boucle d'asservissement.

On en déduit le schéma blocs de la boucle fermée de la Figure 2.5. On notera les indications de signe \pm . L'intérêt de ces schémas blocs réside surtout dans leur côté fonctionnel et visuel. Ils permettent de mieux comprendre les liens de cause à effet.

Pour résumer, on a obtenu de façon directe les relations entre les entrées w et y_r et certaines variables comme y et $e = y_r - y$

$$y = \frac{GK}{1 + GK} y_r + \frac{G_w}{1 + GK} w, \quad (2.7)$$

$$e = \frac{1}{1 + GK} y_r - \frac{G_w}{1 + GK} w \quad (2.8)$$

Il est alors usuel de fixer le cahier des charges d'un asservissement $K(s)$ en donnant les gabarits pour les diagrammes de Bode (surtout celui relatif au gain en fonction de la pulsation ω) des transferts en boucle fermée entre y et y_r (performances du suivi de consigne) et entre e et w (performances en rejet de perturbation). Nous renvoyons aux ouvrages classiques [74] et au cours de SI des classes préparatoires.

2.3 Marge de robustesse

2.3.1 Critère de Nyquist

Revenons au problème que nous avons évoqué en introduction dans la Section 2.1.1. Étant donnés le système (2.1) et le régulateur PI (2.2), on cherche à calculer le *retard critique* Δ^* au delà duquel le système en boucle fermée devient instable. Le schéma blocs en boucle fermée correspond à celui de la Figure 2.5 avec ici les transferts suivants

$$G(s) = \frac{b}{s+a}, \quad G_w(s) = \frac{d}{s+a}, \quad K(s) = \left(K_p + \frac{K_i}{s} \right) e^{-\Delta s}$$

Comme précisé dans l'équation (2.7), on a

$$y = \frac{GK}{1 + GK} y_r + \frac{G_w}{1 + GK} w$$

Ici, $G(s)$ et $K(s)$ sont les quotients de fonctions de s globalement définies sur le plan complexe.

Les objets que nous allons devoir manipuler sont des *fonctions méromorphes* sur le plan complexe \mathbb{C} : une fonction méromorphe $F(s)$ de la variable complexe est en fait (voir par exemple [15]) le

quotient $P(s)/Q(s)$ de deux fonctions entières $P(s)$ et $Q(s)$, la fonction Q étant non identiquement nulle. Une fonction entière est la fonction associée à une série dont le rayon de convergence est infini. Ainsi, F est méromorphe sur \mathbb{C} si et seulement si

$$F(s) = \frac{\sum_{n=0}^{+\infty} a_n s^n}{\sum_{n=0}^{+\infty} b_n s^n}$$

avec a_n et b_n deux suites de nombres complexes⁷ telles que les séries associées ont un rayon de convergence infini, c.-à-d. pour tout $r > 0$, $\lim_{n \rightarrow +\infty} a_n r^n = \lim_{n \rightarrow +\infty} b_n r^n = 0$. Ainsi e^s , $\cosh(\sqrt{s})$, $\sin(s)/s$ sont des fonctions entières et $\frac{\cos(s)}{\sin(s)}$, $1/s$, $\frac{\sin(s^2)}{\cos(s^3)}$ sont des fonctions méromorphes sur \mathbb{C} . En revanche, $\exp(1/s)$ n'est ni une fonction entière, ni une fonction méromorphe sur \mathbb{C} . En effet, elle admet en 0 une singularité "de degré infini". Pour une fonction méromorphe⁸ $F = \frac{P}{Q}$, les zéros de P sont les zéros de F et les zéros de Q sont les pôles de F . De manière informelle, les fonctions entières peuvent être vues comme des polynômes de degré infini et les fonctions méromorphes comme l'analogue des fractions rationnelles en remplaçant les polynômes par des fonctions entières. Autour d'un pôle z_0 , fonctions méromorphes et fractions rationnelles ont des développements limités généralisés très similaires de la forme

$$\sum_{n=-n_0}^{+\infty} a_n (z - z_0)^n$$

avec $n_0 > 0$ ($a_{n_0} \neq 0$) étant la multiplicité du pôle.

Dans notre cas, on a les fonctions méromorphes G et K qui suivent

$$G(s) = \frac{N_G(s)}{D_G(s)}, \quad K(s) = \frac{N_K(s)}{D_K(s)}$$

avec⁹

$$N_G = b, \quad D_G = s + a, \quad N_K = (sK_p + K_i)e^{-\Delta s}, \quad D_K = s$$

et on doit considérer dans le transfert en boucle fermée, la fonction méromorphe T

$$T = \frac{GK}{1 + GK} = \frac{N(s)}{D(s)} = \frac{N_G N_K}{D_G D_K + N_G N_K}$$

avec

$$N(s) = b(sK_p + K_i)e^{-\Delta s}, \quad D(s) = s(s + a) + b(sK_p + K_i)e^{-\Delta s}$$

La valeur critique Δ^* est la valeur de Δ à partir de laquelle $D(s)$ admet au moins un zéro dans le demi plan $\Re(s) \geq 0$, i.e., un zéro instable. Ce zéro sera un *pôle instable* de T (il sera également un pôle instable du transfert depuis la perturbation $\frac{G_w}{1+GK}$).

Comme $a \geq 0$, $b, K_p, K_i > 0$, il est clair que les zéros instables de $D(s)$, correspondent aux zéros instables de $1 + GK = 1 + b \frac{sK_p + K_i}{s(s+a)} e^{-\Delta s}$. La fonction $s \mapsto 1 + GK(s)$ est ici définie pour s différent de 0 et $-a < 0$.

On va maintenant utiliser un résultat important d'analyse complexe qui relie le nombre de zéros et de pôles d'une fonction méromorphe $F(s)$ définie sur un domaine du plan complexe, à la variation de son argument lorsque l'on parcourt le bord du domaine dans le sens direct. On suppose, pour que

7. Il faut qu'au moins l'un de b_n soit différent de 0.

8. On peut toujours choisir les fonctions entières P et Q pour qu'elles n'aient aucun zéro en commun.

9. On remarque que N_G et D_G ne s'annulent pas en même temps, ni N_K et D_K . C'est important : il faut prendre ici les fractions irréductibles pour G et K .

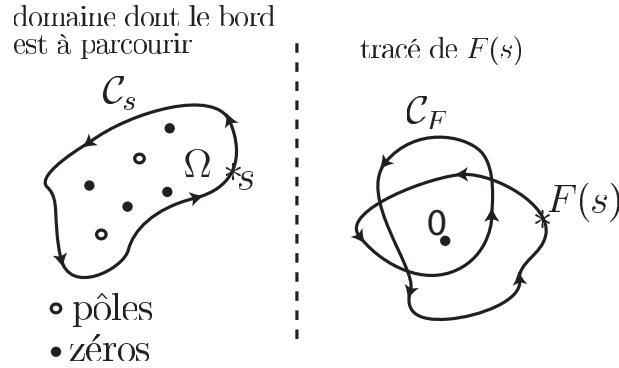


FIGURE 2.6 – Le théorème de Cauchy pour une fonction méromorphe $F(s)$ de $s \in \mathbb{C}$ et un domaine $\Omega \subset \mathbb{C}$ délimité par la courbe fermée notée C_s ; $F(s)$ décrit une courbe fermée notée C_F lorsque s parcourt C_s dans le sens direct. Alors, le nombre de tours N que fait C_F autour de 0 est relié aux nombres de pôles P et de zéros Z de $F(s)$ dans Ω par $N = Z - P$. Ici $Z = 4$ et $P = 2$, donc $F(s)$ fait deux fois le tour de 0 dans le sens direct, i.e., $N = 2$.

l'argument de $F(s)$ soit défini sur le bord du domaine, que F est bien définie sur le bord (pas de pôle sur le bord) et aussi qu'elle ne s'annule pas (pas de zéro sur le bord). Comme on effectue une boucle, il est clair que la variation totale de l'argument est nécessairement un multiple de 2π , disons, $2\pi N$ où N est alors le nombre de tours autour de 0 que fait le point d'affixe complexe $F(s)$ lorsque s parcourt le bord dans le sens direct une seule fois. *Le Théorème 20 de Cauchy* illustré sur la Figure 2.6, dit alors que $N = Z - P$, où Z est le nombre de zéros et P le nombre de pôles à l'intérieur du domaine. Les zéros et les pôles sont comptés avec leur multiplicité. On en donne ici un énoncé plus formel dont la preuve complète s'appuie sur le théorème des résidus appliqué à la fonction F'/F (voir [55]).

Théorème 20 (Cauchy)

Soient C_s une courbe fermée simple dans le plan complexe \mathbb{C} orientée dans le sens direct, et $F(s)$ une fonction méromorphe n'ayant ni pôles ni zéros sur C_s . Soit C_F la courbe fermée du plan complexe décrite par $F(s)$ lorsque s parcourt C_s dans le sens direct. Notons N le nombre de tours (comptés dans le sens direct) que fait C_F autour de 0, et P et Z respectivement le nombre de pôles et de zéros contenus à l'intérieur de C_s . Alors on a l'égalité suivante

$$N = Z - P$$

Démonstration. Sur le domaine délimité par C_s on peut toujours écrire $F(z) = \frac{R(z)}{Q(z)}$ avec

$$R(z) = \prod_{k=1}^Z (z - r_k) \bar{R}(z), \quad Q(z) = \prod_{k=1}^P (z - q_k) \bar{Q}(z)$$

où les r_k sont les Z zéros et les q_k les P pôles (apparaissant plusieurs fois s'ils sont multiples) et où les fonctions analytiques \bar{R} et \bar{Q} ne s'annulent ni sur la courbe fermée C_s ni sur le domaine qu'elle entoure. L'argument de F est la partie imaginaire de $\log F$:

$$\log F(z) = \sum_{k=1}^Z \log(z - r_k) - \sum_{k=1}^P \log(z - q_k) + \log \bar{R}(z) - \log \bar{Q}(z)$$

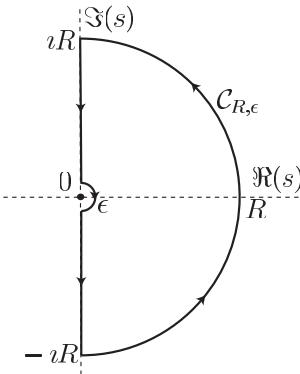


FIGURE 2.7 – La boucle que doit faire s pour englober tout le demi plan à partie réelle positive tout en évitant le pôle de $GK(s) = b \frac{sK_p+K_i}{s(s+a)} e^{-\Delta s}$ en $s = 0$; $R > 0$ est grand et $\epsilon > 0$ est petit.

Ainsi la variation de l'argument de F lorsque z décrit \mathcal{C}_s est la somme de trois termes : les variations des arguments des $\log(z - r_k)$, les opposés des variations des arguments des $\log(z - q_k)$ et la variation de l'argument de $\log(\bar{R}(z)/\bar{Q}(z))$. Or la variation de l'argument de $\log(z - r_k)$ ou de $\log(z - q_k)$ est 2π car r_k et q_k sont entourés par \mathcal{C}_s . Comme $\bar{R}(z)/\bar{Q}(z)$ ne s'annule pas, la variation de son argument vaut 0 lorsque z parcours une fois \mathcal{C}_s ¹⁰. Ainsi la variation de l'argument de F est bien $Z - P$. \square

Dans notre cas, $F = 1 + GK = 1 + b \frac{sK_p+K_i}{s(s+a)} e^{-\Delta s}$ admet un pôle simple en 0. Prenons Ω de la forme d'un demi-disque comme illustré sur la Figure 2.7 avec $R > 0$ grand et $\epsilon > 0$ petit. On note $\mathcal{C}_{R,\epsilon}$ le bord de Ω . On voit que si $1 + GK$ admet un zéro dans le demi-plan $\Re(s) \geq 0$, il doit faire partie nécessairement d'un tel Ω pour un R assez grand et un ϵ assez petit. Maintenant, il faut compter le nombre de tour que fait $F = 1 + GK$ autour de 0 lorsqu'on parcourt $\mathcal{C}_{R,\epsilon}$: c'est nécessairement dans le sens positif car ici $1 + GK$, n'a pas de pôle dans Ω , c.-à-d. $P = 0$.

Maintenant, pourquoi considérer $F = 1 + GK$ au lieu de $D = N_G N_K + D_G D_K$ à la place de F . On éviterait le pôle en $s = 0$ qu'on est obligé de contourner avec le petit demi-cercle de rayon ϵ . Cela vient du fait que sur la partie de $\mathcal{C}_{R,\epsilon}$ loin de l'origine (bref sur le grand demi cercle de rayon R) GK est très petit et reste donc très proche de 0. Sur cette partie, F reste ainsi très proche de +1 et ne tourne donc pas autour de 0. Le fait que GK soit petit pour les s de $\mathcal{C}_{R,\epsilon}$ loin de l'origine vient du fait que les transferts G et K sont des transferts causaux : $GK(s) = b \frac{sK_p+K_i}{s(s+a)} e^{-\Delta s}$ tend vers 0 lorsque s tend vers l'infini tout en restant à partie réelle positive. On voit que $\Delta > 0$ est capital ici pour avoir cette propriété et que ce n'est plus vrai pour s tendant vers l'infini avec une partie réelle négative.

Compter le nombre de tours de $1 + GK$ autour de 0 revient à compter le nombre de tours de GK autour de -1 . Ainsi, si la courbe fermée du plan complexe décrite par $GK(s)$ avec s décrivant le contour de la Figure 2.7 n'entoure pas -1 , alors, $F = 1 + GK$ n'entoure pas 0. On a donc $N = 0$, or $P = 0$ donc, d'après le Théorème de Cauchy 20, on déduit $Z = 0$, F n'admet pas de zéro dans le demi plan $\Re(s) > 0$. On vient de retrouver ici le *critère du revers* (connu en classes préparatoires).

Exemple 13 (Illustration du Théorème de Cauchy pour différents contours). Soit le système de fonction de transfert $\frac{1}{s+2}$ bouclé par un contrôleur PI de fonction de transfert $K_P + \frac{K_I}{s} = 3 + \frac{40}{s}$. Pour

10. Pour s'en convaincre, il suffit de déformer la courbe fermée \mathcal{C}_s en la réduisant continûment en un point, intérieur au domaine initialement entouré par \mathcal{C}_s . Au cours de cette déformation (homotopie), $\bar{R}(z)/\bar{Q}(z)$ ne s'annule jamais, son argument est donc bien défini et sa variation sur un tour de la courbe déformée est un multiple de 2π . À la fin de la déformation, la courbe fermée suivie par z est réduite à un point et donc la variation de l'argument de $\bar{R}(z)/\bar{Q}(z)$ est nulle. Au cours de la déformation, la variation ne peut pas sauter brusquement d'un multiple non nul de 2π à 0 pour cause de continuité. Ainsi cette variation doit être nulle dès le départ.

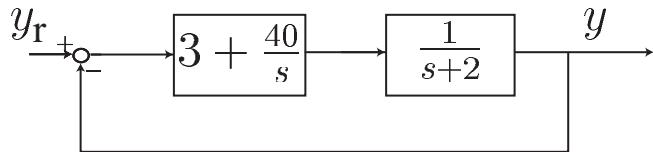
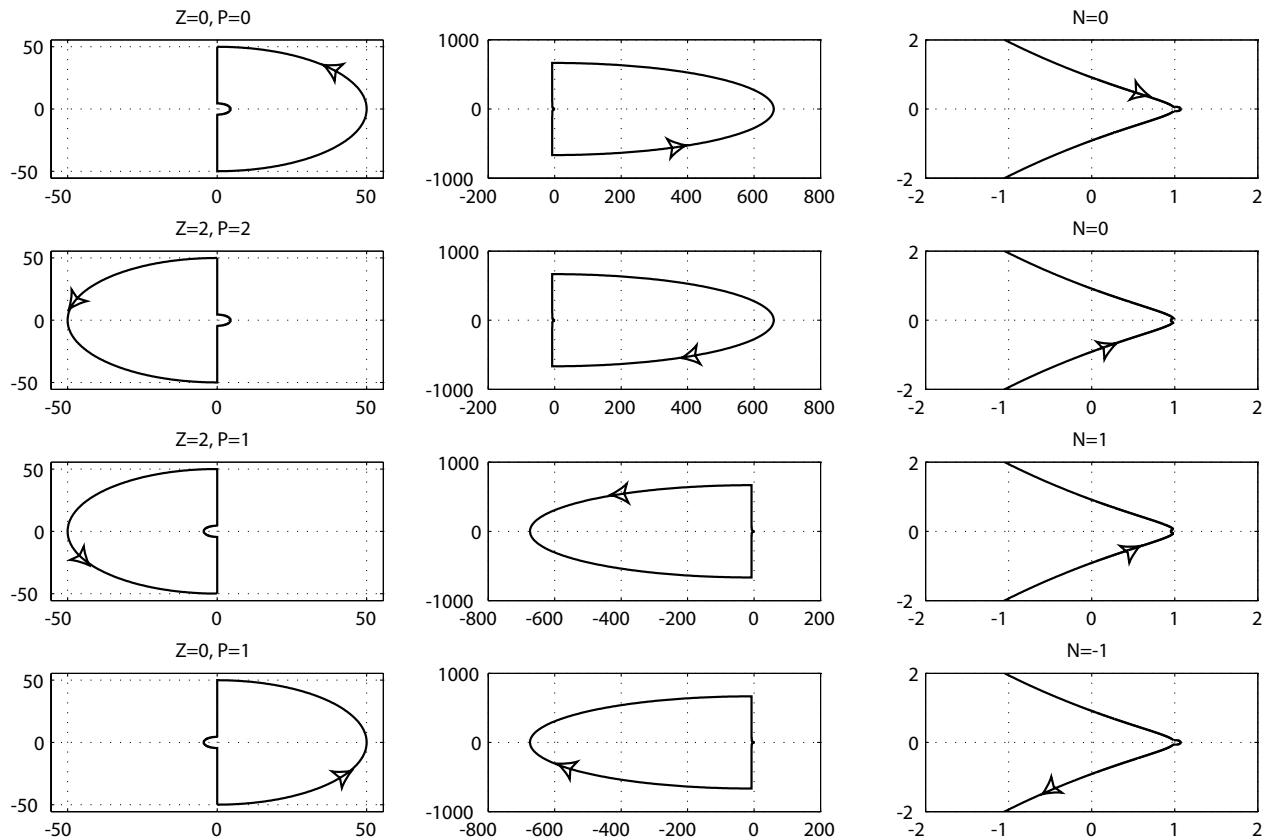


FIGURE 2.8 – Exemple de système pour l'étude du Théorème de Cauchy.

l'étude du système en boucle fermée représenté sur la Figure 2.8, on doit étudier l'annulation de la fonction de transfert

$$1 + \frac{sK_p + K_I}{s(s+2)} = \frac{s^2 + 5s + 40}{s(s+2)}$$

Ce transfert possède deux pôles (0 et -2) et deux zéros complexes conjugués $\frac{-5 \pm 3i\sqrt{15}}{2}$. On peut, en utilisant différents contours comme on l'a fait sur la Figure 2.9, vérifier la formule $N = Z - P$ du Théorème 20 de Cauchy.

FIGURE 2.9 – Illustrations du Théorème 20 de Cauchy. Gauche : courbe fermée parcourue par s . Milieu : lieu de $1 + \frac{sK_p + K_I}{s(s+2)}$. Droite : agrandissement du lieu autour de 0.

Des considérations ci-dessus on déduit sans peine le *critère de Nyquist* pour les systèmes bouclés de la Figure 2.5.

Définition 10 (Lieu de Nyquist). Étant donnée une fonction de transfert en boucle ouverte $G(s)$ strictement causale et un transfert de contrôleur $K(s)$ également strictement causal¹¹. On suppose que, s'il y a des simplifications par des facteurs du type $(s - s_0)$ dans le produit $G(s)K(s)$, elles ne portent pas sur des s_0 à partie réelle strictement positive. On appelle lieu de Nyquist la courbe décrite dans le plan complexe par $G(\omega)K(\omega)$ lorsque ω parcourt $+\infty$ vers $-\infty$.

Théorème 21 (Critère de Nyquist)

Avec les notations de la Définition 10, le système en boucle fermée de transfert

$$\frac{GK}{1 + GK}$$

est asymptotiquement stable si et seulement si le *lieu de Nyquist* ne passe pas par -1 et s'il entoure le point -1 *dans le sens indirect* autant de fois que GK admet de pôles à partie réelle strictement positive (les pôles sont comptés avec leur multiplicité).

Exemple 14 (Stabilisation avec retard). Considérons le système de fonction de transfert avec retard

$$G(s) = \exp\left(-\frac{s}{2}\right) \frac{s^2 + 4s + 2}{s^2 - 1} \quad (2.9)$$

On boucle ce système avec un contrôleur P proportionnel de gain k , pour obtenir le schéma en boucle fermée représenté sur la Figure 2.10. On se demande sous quelle condition sur k , le paramètre de

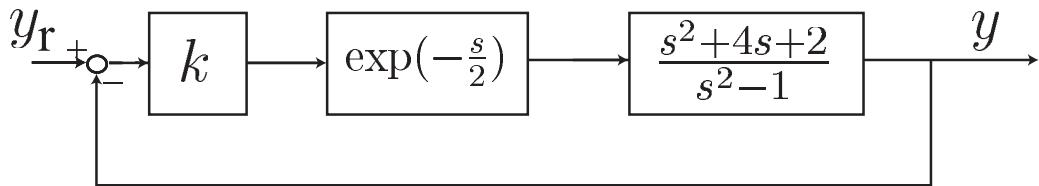


FIGURE 2.10 – Bouclage d'un système à retard.

réglage du contrôleur, on obtient la stabilité en boucle fermée. Pour répondre à cette question, il suffit de tracer le lieu de Nyquist de $kG(s)$, qu'on a représenté sur la Figure 2.11 (pour $k = 1$). Le transfert $G(s)$ comporte un unique pôle à partie réelle positive. D'après le Théorème 21, pour que le système en boucle fermée représenté sur la Figure 2.10 soit asymptotiquement stable, il faut et il suffit que le lieu de Nyquist entoure une seule fois (dans le sens indirect) le point -1 . Ce n'est pas le cas avec $k = 1$, comme on peut le constater sur la Figure 2.11. Le paramètre k agit comme une homothétie. Pour $k \in]\frac{1}{2}, \frac{1}{1.1845}[$, le système est asymptotiquement stable, il est instable sinon.

Si, GK admet des pôles sur l'axe imaginaire, alors le lieu de Nyquist admet des branches infinies. Il convient alors pour compter correctement le nombre tours de modifier légèrement le contour le long de l'axe imaginaire descendant comme on l'a fait auparavant pour le pôle en $s = 0$ (voir Figure 2.7) :

11. Par strictement causal, on signifie ici que $G(s)$ et $K(s)$ sont des fonctions méromorphes qui tendent vers zéro quand s tend vers l'infini dans le demi plan complexe à partie réelle positive. C'est le cas pour des fractions rationnelles dont le degré du numérateur est strictement inférieur à celui du dénominateur.

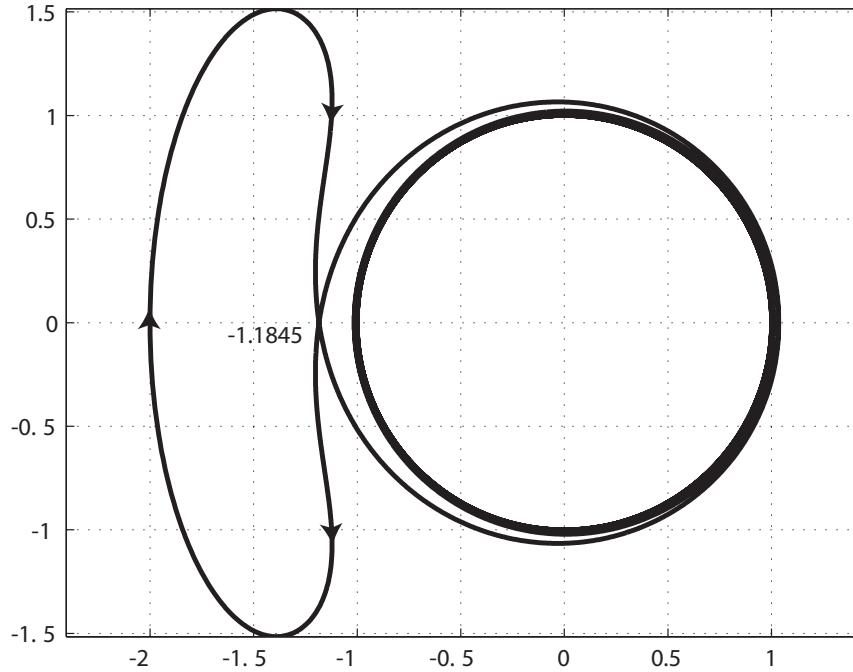


FIGURE 2.11 – Lieu de Nyquist de $kG(s)$ donné par (2.9) avec $k = 1$.

on contourne chaque pôle imaginaire avec un petit demi-cercle, centré sur ce pôle imaginaire et qui passe du côté de $\Re(s) > 0$.

Noter enfin que, en toute généralité, pour que le critère de Nyquist implique la stabilité asymptotique, il faut qu'il existe $\eta > 0$ tel que,

$$\lim_{\begin{array}{l} |s| \mapsto +\infty \\ \Re(s) \geq -\eta \end{array}} G(s)K(s) = 0$$

ce qui est le cas pour les fractions rationnelles et les fractions de quasi-polynômes tels que ceux que nous avons considérés. En effet si GK vérifie les conditions du critère de Nyquist, alors pour un parcours en s légèrement décalé vers la gauche, $s = -\alpha + i\omega$, ω allant de $+\infty$ vers $-\infty$ et $0 < \alpha \leq \eta$ très petit a priori, on aura un lieu de Nyquist, $GK(-\alpha + i\omega)$ légèrement déplacé et qui entourera le point -1 avec le même nombre de tours que celui pour lequel $\alpha = 0$. Ainsi, on est sûr que les zéros de $1+GK$ et donc de $D = N_GN_K + D_GD_K$ sont tous à partie réelle inférieure à $-\alpha$. Nous détaillons ces conditions ici car nous souhaitons prendre en compte des retards. Donc on ne peut pas se contenter de prendre G et K dans la classe des fractions rationnelles en s uniquement. On souhaite aussi traiter des fractions en s et $e^{-\Delta s}$.

2.3.2 Marge de phase et retard critique

Revenons à l'exemple $G(s)K(s) = b \frac{sK_p + K_i}{s(s+a)} e^{-\Delta s}$. Nous allons calculer explicitement le *retard critique* $\Delta^* > 0$ à partir duquel, le lieu de Nyquist commence à entourer -1 . Il est donné par ce que l'on appelle la *marge de phase*.

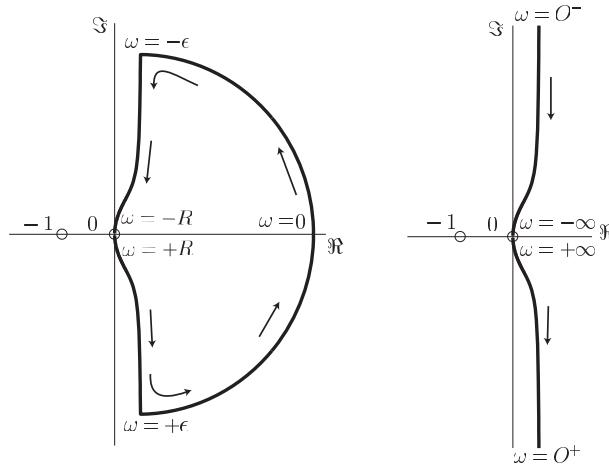


FIGURE 2.12 – Le lieu de Nyquist de $G(s)K(s) = ke^{-\delta s} \frac{s+\eta}{s(s+1)}$ avec $k = 1$, $\delta = 0$ et $\eta = 0.5$; la courbe fermée du plan complexe décrite par $G(s)K(s)$ lorsque s décrit la courbe $\mathcal{C}_{R,\epsilon}$ de la Figure 2.7 pour $R \gg 1$ et $0 < \epsilon \ll 1$.

On commence par une normalisation en posant

$$\tilde{s} = \frac{s}{a}, \quad k = abK_p, \quad \eta = \frac{K_i}{K_p a}, \quad \delta = a\Delta$$

Alors

$$G(s)K(s) = G(\tilde{s})K(\tilde{s}) = ke^{-\delta \tilde{s}} \frac{\tilde{s} + \eta}{\tilde{s}(\tilde{s} + 1)}$$

Ainsi, quitte à changer l'échelle de temps, on peut toujours supposer $a = 1$, on confond dans la suite \tilde{s} et s et on écrit

$$G(s)K(s) = ke^{-\delta s} \frac{s + \eta}{s(s + 1)}$$

Sur la Figure 2.12, on a tracé le lieu de $G(s)K(s)$ pour s suivant le parcours de la Figure 2.7. Le point -1 est à l'extérieur du domaine entouré par le lieu de Nyquist. En général, on ne représente pas le grand demi-cercle du demi plan droit. On ne trace que la figure simplifiée qui est à droite. On sait alors qu'il faut refermer cette courbe en partant de la branche infinie du bas $\Im = -\infty$ lorsque $\omega = 0^+$ vers celle du haut $\Im = +\infty$ lorsque $\omega = 0^-$ en passant à l'infini par le demi plan droit $\Re > 0$.

On trouvera sur la Figure 2.13, les déformations successives qu'on observe lorsque le paramètre η grandit, c.-à-d. lorsqu'on change les réglage du contrôleur. On se convaincra sans difficulté que lorsque $\delta = 0$, le lieu de Nyquist de $G(s)K(s)$ n'entoure jamais -1 quels que soient k et $\eta > 0$. Noter que lorsque $\delta = 0$, la stabilité est donnée par les racines d'un polynôme de degré deux et on a déjà vu, au premier chapitre, qu'un *régulateur PI* sur un premier ordre est toujours asymptotiquement stable. Mais ce n'est pas pour retrouver ce fait trivial que l'on a fait tout cela. C'est pour voir ce qui se passe lorsque le retard δ est non nul.

Pour les mêmes k et η , quelle relation y-a-t-il entre le *lieu de Nyquist* de $G(s)K(s)$ lorsque $\delta = 0$ et $\delta > 0$? Pour $s = i\omega$, $G(i\omega)K(i\omega)$ subit une rotation de $-\delta\omega$ (multiplication par $e^{-i\delta\omega}$). Comme l'illustre la Figure 2.14, la valeur critique Δ^* à partir de laquelle -1 rentre à l'intérieur du domaine entouré par le lieu de Nyquist, est directement liée aux points d'intersection du lieu de Nyquist pour $\delta = 0$ avec le cercle unité. Pour chaque k et η , on note $\pm\vartheta$ les deux valeurs de pulsation qui correspondent aux points d'intersection avec le cercle unité. Un simple calcul faisant intervenir une équation

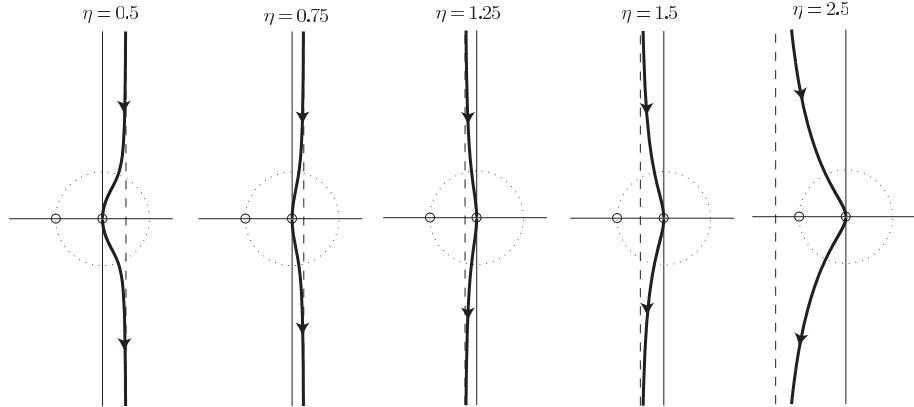


FIGURE 2.13 – Le lieu de Nyquist de $G(s)K(s) = ke^{-\delta s} \frac{s+\eta}{s(s+1)}$ avec $k = 1$, $\delta = 0$ et diverses valeurs de η .

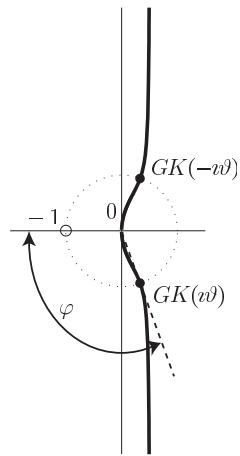


FIGURE 2.14 – La *marge de phase* φ pour $G(s)K(s) = k \frac{s+\eta}{s(s+1)}$ avec $k = 1$, $\eta = 0.5$.

bi-carrée traduisant le fait que le complexe $k \frac{(\vartheta + \eta)}{\vartheta(\vartheta + 1)}$ est de module 1, donne

$$\vartheta = \sqrt{\frac{-(1 - k^2) + \sqrt{(1 - k^2)^2 + 4k^2\eta^2}}{2}}$$

On note $\varphi \in [0, \pi[$ l'argument suivant (noter le signe – au second membre)

$$e^{i\varphi} = -k \frac{(\vartheta + \eta)}{\vartheta(\vartheta + 1)}$$

c'est la *marge de phase*. Alors,

$$\delta^* = \frac{\varphi}{\vartheta}$$

est le *retard critique* au delà duquel des instabilités apparaissent. La marge de phase n'a pas vraiment de sens intrinsèque. Seul compte en fait le retard critique δ^* qui s'en déduit : si, par exemple, la pulsation critique ϑ est très petit alors le retard critique δ^* est grand dès que φ n'est pas proche de 0.

Pour $k = 1$ et $\eta = 0.5$, la Figure 2.15 montre la perte progressive de stabilité lorsque δ grandit, la valeur critique étant autour de 2.7, d'après les formules ci-dessus. De ce graphique on tire aussi l'information suivante : lorsque δ franchit le seuil δ^* , deux pôles complexes¹² conjugués traversent en même temps l'axe imaginaire en venant de la partie stable. On sait cela car juste après δ^* , -1 est entouré deux fois donc on a exactement deux racines à partie réelle strictement positive : un peu avant δ^* , le système est très oscillant mais les oscillations sont lentement amorties ; un peu après δ^* , le système est toujours très oscillant mais les oscillations sont lentement amplifiées.

2.3.3 Marge de gain

La *marge de gain* permet de quantifier la préservation de la stabilité, quand sur le modèle

$$\frac{d}{dt}y = -ay + bu + dw$$

on connaît mal le coefficient b devant u (on connaît son signe cependant), c.-à-d. qu'on connaît mal le gain de la commande. Cette incertitude revient à multiplier le bloc contrôle $K(s)$ par un coefficient $k > 0$, en théorie proche de 1. $\frac{1}{1+G(s)K(s)}$ étant stable, on cherche à savoir à partir de quelle valeur de k (en partant de 1), le transfert $\frac{1}{1+kG(s)K(s)}$ devient instable. Les valeurs maximales et minimales admissibles de k sont appelées *marges de gain*.

Reprendons l'exemple $G(s)K(s) = e^{-\delta s} \frac{s+\eta}{s(s+1)}$. On considère tout d'abord $\delta = 0$. On voit d'après les divers lieux de Nyquist tracés sur la Figure 2.13 que, quelle que soit l'homothétie de rapport k positif, ce dernier ne passe jamais par -1 . Ainsi, on a une *marge de gain* infinie dans ce cas. C'est une façon indirecte de retrouver le résultat du Chapitre 1 qui dit qu'un régulateur PI sur un 1er ordre stable est toujours asymptotiquement stable pourvu qu'on connaisse le signe du gain sur le contrôle.

Cependant, supposons que nous ayons un petit retard, par exemple $\delta = 0.25$ pour $\eta = 0.5$. Alors on voit sur la Figure 2.15 que ce petit retard ne modifie sensiblement le lieu de Nyquist qu'autour de 0, là où les pulsations ω sont grandes en valeurs absolues. Plus ω est grand, plus la multiplication par $\exp(i\delta\omega)$ aura tendance à enrouler le lieu de Nyquist autour de 0. Aussi, il est évident que, quelle que soit la taille du retard $\delta > 0$, même très petit, on trouvera toujours une dilatation du lieu de Nyquist,

12. Ce sont les pôles en boucle fermée, pas ceux de GK mais ceux de $1/(1 + GK)$.

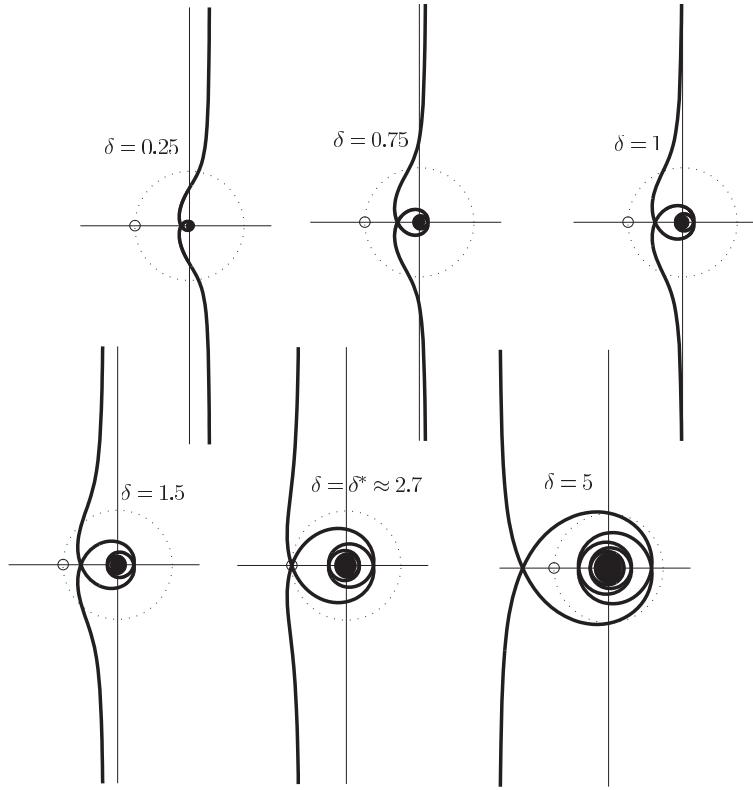


FIGURE 2.15 – Lieu de Nyquist pour $G(s)K(s) = ke^{-\delta s} \frac{s+\eta}{s(s+1)}$ avec $k = 1$, $\eta = 0.5$ et diverses valeurs du retard δ .

qui l'amène à passer par -1 . Bien-sûr, plus le retard sera petit, plus la dilatation k devra être grande pour introduire une instabilité.

On retrouve ainsi ce que nous avons vu déjà au chapitre précédent sur les dynamiques négligées : les gains K_p et K_i ne peuvent pas être pris arbitrairement grands. Ici, un petit retard $\delta = \epsilon > 0$ correspond à une dynamique rapide négligée. Il est intéressant de remarquer que le Théorème de Tikhonov 15 porte sur des systèmes singulièrement perturbés, c'est à dire avec un petit paramètre ϵ positif devant d/dt . C'est formellement et aussi fondamentalement la même chose pour un retard d'ordre ϵ : dans l'opérateur $e^{-\epsilon s}$ on retrouve bien le même petit paramètre devant $s = d/dt$.

2.3.4 Lecture des marges sur le diagramme de Bode

Il suffit de prendre un exemple pour comprendre comment on fait. Sur la figure 2.16 on a tracé le diagramme de Bode pour $G(s)K(s) = e^{-s} \frac{s+\frac{1}{2}}{s(s+1)}$ qui n'admet aucun pôle à partie réelle strictement positive. Ainsi on voit que

- la marge de gain k^* est égale à l'inverse du module de $GK(i\omega)$ pour la pulsation $\omega = \omega_{k^*}$ où l'argument de GK passe la première fois par π (sur le diagramme cela correspond au premier saut de -180 à $+180$ degrés).
- La marge de phase ϕ^* correspond à l'écart entre la phase de $GK(i\omega)$ et π pour la pulsation $\omega = \omega_{\phi^*}$ correspondant au premier passage du module de GK par 1 (sur le diagramme cela correspond donc à 0 en décibel ($20 \log_{10}$))).

Alors la marge de retard, qui seule a une signification physique, est donnée par $\Delta^* = \frac{\phi^*}{\omega_{\phi^*}}$. Dans cette formule l'unité d'angle pour ϕ^* et ω_{ϕ^*} doit être la même (il est conseillé prendre des radians pour ne

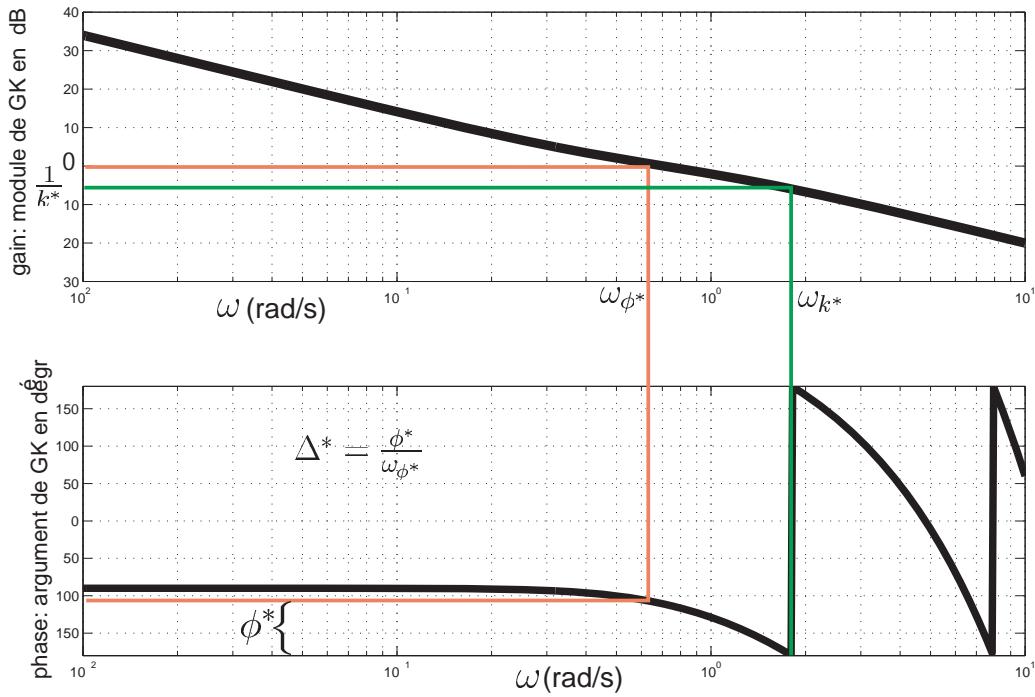


FIGURE 2.16 – Visualisation sur le diagramme de Bode des marges de gains k^* et de phase ϕ^* lorsque $G(s)K(s) = e^{-s} \frac{s + \frac{1}{2}}{s(s+1)}$.

pas se tromper).

2.3.5 Pôles dominants

Supposons, comme nous l'avons fait dans le chapitre précédent, que le capteur qui fournit y n'est pas instantané mais a une petite dynamique que l'on peut représenter comme un filtre rapide de transfert $\frac{1}{1+\epsilon s}$. Ainsi $G(s)$ est remplacé par $G_\epsilon = G(s)R_\epsilon(s)$. Supposons $1/(1 + G(s)K(s))$ stable. Alors pour $\epsilon > 0$, suffisamment petit, $1/(1 + G_\epsilon(s)K(s))$ reste stable.

Pour ϵ assez petit, l'effet sur le *lieu de Nyquist* de $G(s)K(s)$ de la multiplication par $\frac{1}{1+\epsilon s}$, n'est notable que pour des $s = i\omega$ assez grands en module. Or, pour de tels ω de l'ordre de $1/\epsilon$, $G(i\omega)K(i\omega)$ est déjà très petit et donc seule la partie du lieu qui est proche de l'origine est modifiée. Donc le nombre de tour autour du point -1 ne change pas, et donc le transfert $1/(1 + G_\epsilon(s)K(s))$ est lui aussi stable. Aussi, il est légitime de faire brutalement $\epsilon = 0$ dans le transfert G_ϵ pour ne garder que G . On dit alors que $G(s)$ ne conserve que les *pôles dominants* : il est ainsi légitime d'éliminer les pôles très rapides et stables.

En résumé, ne conserver que les pôles dominants dans le transfert G est identique au fait de prendre comme modèle de contrôle la partie lente du système, la partie rapide étant stable : l'approximation, décrite dans la Section 1.4.1 du système lent/rapide (Σ^ϵ) par le système lent (Σ^0) est de même nature que celle de G_ϵ par $G = G_0$.

2.4 Compléments

2.4.1 Calcul de tous les contrôleurs PID stabilisant un système du premier ordre à retard

Comme on l'a vu, le *contrôleur PI* et plus généralement le *contrôleur PID* (D pour dérivée) est extrêmement efficace et utilisable dans un nombre important de situations. On a vu que le transfert K d'un PI est $K(s) = k_P + \frac{k_I}{s}$ avec k_P et k_I les gains proportionnel et intégral. Le *régulateur PID* a simplement le transfert

$$K(s) = k_P + \frac{k_I}{s} + k_D s$$

où k_D est le gain dérivé. Ainsi, régler un PID revient à trouver les bons gains, quand c'est possible, k_P , k_I et k_D , pour avoir un transfert en boucle fermée $GK/(1 + GK)$ au minimum stable.

On a vu comment tester sa *robustesse*, notamment au retard. On propose ici quelques résultats théoriques importants qui donnent des conditions nécessaires et suffisantes pour obtenir la stabilité en boucle fermée en utilisant un tel contrôleur. Ces résultats se limitent aux systèmes du premier ordre (on se référera à [68], pour différentes extensions aboutissant à des énoncés beaucoup plus indirects), leur démonstration fait appel à des résultats d'analyse complexe permettant de qualifier la négativité des racines d'un polynôme sans les calculer. De tels critères (critère de Routh, ou de Hermite-Biehler énoncés dans les Théorèmes 7 et 6) sont souvent utilisés car numériquement très simples.

On considère un système du premier ordre à retard dont la fonction de transfert est

$$G(s) = \frac{k}{1 + \tau s} e^{-\Delta s} \quad (2.10)$$

où $k \in \mathbb{R}^{+*}$, $\tau \in \mathbb{R}^*$, $\Delta \in \mathbb{R}^{+*}$. Dans un premier temps, on considère qu'on utilise seulement un contrôleur proportionnel

$$K(s) = k_P \quad (2.11)$$

et qu'on l'utilise tel que représenté sur la Figure 2.17.

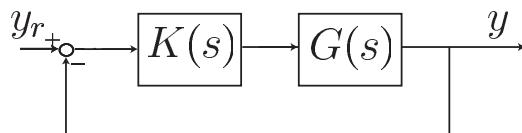


FIGURE 2.17 – Bouclage d'un système du premier ordre à retard.

Deux cas sont alors à considérer. Soit le système est stable (cas du Théorème 4) en boucle ouverte, et on cherche juste à améliorer ses performances par un contrôleur en boucle fermée, soit le système est instable (cas du Théorème 5) et c'est à la boucle fermée de le stabiliser.

Proposition 4 ([68]). *On suppose $\tau > 0$, alors l'ensemble des contrôleurs proportionnels préservant la stabilité du système (2.10) est défini par les valeurs admissibles suivantes*

$$-\frac{1}{k} < k_P < \frac{-1}{k \cos z_1}$$

où z_1 est l'unique racine sur $\pi/2, \pi]$ de l'équation

$$\tan(z) = -\frac{\tau}{\Delta} z$$

On notera de plus que la borne supérieure sur k_P dans cet énoncé est une fonction décroissante du retard Δ . Dans le cas instable, on souhaite en général que le contrôleur stabilise également le système en l'absence de retard. On aboutit alors à l'énoncé suivant.

Proposition 5 ([68]). *On suppose $\tau < 0$. Une condition nécessaire pour que le contrôleur proportionnel (2.11) stabilise à la fois le système (2.10) et ce même système en l'absence de retard est que $\Delta < |\tau|$. On suppose cette condition vérifiée, alors l'ensemble des contrôleurs proportionnels assurant la stabilité du système (2.10) est défini par les valeurs admissibles suivantes*

$$\frac{\tau}{k\Delta} \sqrt{z_1^2 + \left(\frac{\Delta}{\tau}\right)^2} < k_P < -\frac{1}{k}$$

où z_1 est l'unique racine sur $]0, \pi/2[$ de l'équation

$$\tan(z) = -\frac{\tau}{\Delta} z$$

La preuve de ces théorèmes utilise l'extension du Théorème 6 de Hermite-Biehler pour les *quasi-polynômes* (i.e. des polynômes en s avec coefficients en $e^{-\Delta s}$ comme on en rencontre dans les fonctions de transferts des systèmes à retard bouclés).

Le cas de la stabilisation par un PID

$$K(s) = k_P + \frac{k_I}{s} + k_D s \quad (2.12)$$

est traité dans l'énoncé suivant (un peu compliqué à première vue mais simple à mettre en œuvre). Seul le cas des systèmes stables en boucle ouverte est traité ici. On se reportera à [68] pour un exposé concernant le cas des systèmes instables en boucle ouverte.

Proposition 6 ([68]). *On suppose $\tau > 0$, alors l'ensemble des contrôleurs PID (2.12) préservant la stabilité du système (2.10) est défini par les valeurs admissibles suivantes*

$$-\frac{1}{k} < k_P < \frac{1}{k} \left(\frac{\tau}{\Delta} \alpha_1 \sin(\alpha_1) - \cos(\alpha_1) \right)$$

où α_1 est l'unique racine sur $]0, \pi[$ de l'équation

$$\tan(\alpha) = -\frac{\tau}{\tau + \Delta} \alpha$$

Pour k_p en dehors de ces bornes, il n'existe pas de contrôleur PID (2.12) stabilisant (2.10). Une fois choisi k_p dans ces bornes, il faut et il suffit pour stabiliser (2.10) de choisir (k_I, k_D) comme suit.

Soient z_1, z_2 les deux racines réelles (ordonnées) sur $]0, 2\pi[$ de l'équation

$$kk_P + \cos z - \frac{\tau}{\Delta} z \sin z = 0$$

On note $m(z) = \frac{\Delta^2}{z^2}$, $b(z) = -\frac{\Delta}{kz} (\sin z + \frac{\tau}{\Delta} z \cos z)$, et $m_i = m(z_i)$, $b_i = b(z_i)$, $i = 1, 2$.

1. si $k_P \in]-\frac{1}{k}, \frac{1}{k}[$, alors (k_I, k_D) doit être choisi de telle sorte que $-\frac{\tau}{k} < k_d < \frac{\tau}{k}$, $k_i > 0$ et $k_d > m_1 k_I + b_1$
2. si $k_P = \frac{1}{k}$, alors (k_I, k_D) doit être choisi de telle sorte que $k_d < \frac{\tau}{k}$, $k_i > 0$ et $k_d > m_1 k_I + b_1$
3. si $k_P \in]\frac{1}{k}, \frac{1}{k} [\left(\frac{\tau}{\Delta} \alpha_1 \sin(\alpha_1) - \cos(\alpha_1) \right)[$, alors (k_I, k_D) doit être choisi de telle sorte que $m_1 k_I + b_1 < k_d < \min(m_2 k_I + b_2, \frac{\tau}{k})$, et $k_i > 0$

On pourra remarquer la similitude apparente entre les énoncés des Propositions 6 et 4. Les bornes supérieures sur le gain proportionnels sont différentes. En effet, l'énoncé 6 nécessite bien souvent l'usage effectif du terme intégral et du terme dérivé car le point $(k_I, k_D) = (0, 0)$ peut être exclu de l'adhérence des contraintes (dans le cas 3, on peut avoir $b_2 > b_1 > 0$).

2.4.2 Méthodes de réglage de Ziegler-Nichols

Il existe plusieurs méthodes systématiques de réglage des régulateurs PID. Ces méthodes sont souvent utilisées en pratique comme heuristiques, sans justification. On propose de les re-situer dans le cadre d'analyse de stabilité que nous avons présenté, notamment à la Section 2.4.1. Ces règles sont très nombreuses et diffèrent par les performances qu'on peut en atteindre en terme de temps de convergence, dépassement prévu, robustesse, etc. Historiquement, ce sont les règles de Ziegler-Nichols [79, 80] qui sont apparues les premières, elles sont toujours parmi les plus utilisées. On trouve aussi les règles de Cohen-Coon [18], Chien-Hrones-Reswick [17], ou Lee-Park-Lee-Brosilow [47].

Première méthode de Ziegler-Nichols

Pour concevoir un régulateur PID pour un process donné $G(s)$, on réalise l'expérience boucle-ouverte suivante. On enregistre la réponse du système à un échelon d'entrée et on relève les paramètres a et Δ construits à partir de la ligne de plus grande pente de la réponse. On note également τ et k tels que représentés sur la Figure 2.18, construits à partir de la valeur asymptotique estimée et du temps de réponse à $1 - e^{-1} = 0.63$ ¹³.

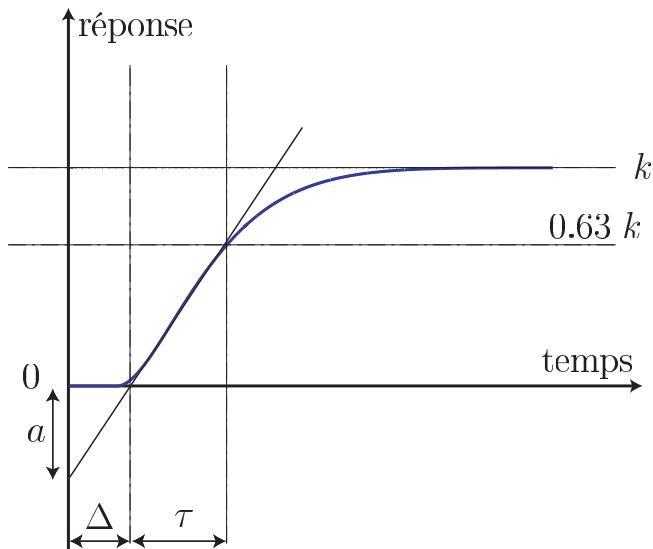


FIGURE 2.18 – Détermination d'un modèle du premier ordre à retard à partir de la *réponse à un échelon* donné par $y(t) = k(1 - e^{\frac{t-\Delta}{\tau}})\mathbb{I}_{t \geq \Delta}$.

Les réglages heuristiques de Ziegler-Nichols sont reportés dans le tableau 2.1. Ils permettent de régler au choix un contrôleur P , un PI ou un PID .

Seconde méthode de Ziegler-Nichols

La seconde méthode de Ziegler-Nichols nécessite une expérimentation en boucle fermée avec un contrôleur déjà installé dont il suffit de modifier les gains. On boucle le système avec un régulateur P (en mettant le gain intégral et le gain dérivé à 0) dont on fait progressivement augmenter le gain jusqu'à atteindre un régime oscillatoire entretenu. Autrement dit, on atteint la limite de stabilité. Une

13. Un modèle du premier ordre atteint environ 63% de sa valeur finale en un temps égal à 1 fois sa constante de temps τ

Contrôleur	k_P	k_I	k_D
P	$1/a$		
PI	$0.9/a$	$0.3/(a\Delta)$	
PID	$1.2/a$	$0.6/(a\Delta)$	$0.6\Delta/a$

TABLE 2.1 – Réglages de Ziegler-Nichols (première méthode).

fois ce régime obtenu, on note k_u le gain proportionnel “ultime” et la période des oscillations T_u lui correspondant et on règle alors le contrôleur comme expliqué sur le tableau 2.2.

Contrôleur	k_P	k_I	k_D
P	$0.5k_u$		
PI	$0.4k_u$	$0.5k_u/T_u$	
PID	$0.6k_u$	$1.2k_u/T_u$	$0.075k_uT_u$

TABLE 2.2 – Réglages de Ziegler-Nichols (seconde méthode).

Bien que très simple à utiliser, ce mode opératoire de réglage de contrôleur a le désavantage de nécessiter la *presque* déstabilisation de l’installation qu’on souhaite contrôler. Cet inconvénient peut être évité. On pourra se reporter à [74] pour une présentation d’une méthode de réglage aboutissant aux coefficients de la seconde méthode de Ziegler-Nichols, sans risque de déstabilisation. Cette méthode, qui utilise un bloc relai en boucle fermée, consiste à faire apparaître un cycle limite dont on analyse les caractéristiques pour retrouver les paramètres du système à contrôler.

Liens avec les résultats de stabilisation

Il est assez simple de faire un lien entre la seconde méthode de Ziegler-Nichols et le Théorème 4 si on suppose que le système qu’on cherche à contrôler est effectivement un système du premier ordre à retard stable de la forme (2.10). Ce théorème montre qu’il existe une valeur déstabilisante du gain proportionnel. C’est cette valeur k_u que l’expérience en boucle fermée permet d’estimer sans connaissance des paramètres du modèle. Ensuite, la méthode de Ziegler-Nichols du tableau 2.2 propose de réduire ce gain et donc de se retrouver avec un contrôleur proportionnel stabilisant.

Les coefficients proposés par la première méthode de Ziegler-Nichols sont en pratique très proches de ceux de la deuxième méthode. Néanmoins, c’est de ces derniers que nous allons prouver une propriété très intéressante. Dans le cas où le système à contrôler est effectivement un système du premier ordre à retard stable de la forme (2.10), un calcul direct montre que le paramètre a tel que défini sur la Figure 2.18 vaut

$$a = k \frac{\Delta}{\tau}$$

Notons $\mu = \Delta/\tau$. Avec cette valeur, les gains proposés dans le tableau 2.1 sont

$$k_P = \frac{1.2}{k\mu}, \quad k_I = \frac{0.6}{k\mu^2\tau}, \quad k_D = \frac{0.6\tau}{k} \quad (2.13)$$

Considérons en premier lieu le gain proportionnel suggéré. Nous allons montrer qu’il est toujours acceptable, en d’autres termes, qu’il satisfait toujours les hypothèses du Théorème 6. Pour notre système

stable, le gain proposé satisfait $k_P > 0 > \frac{-1}{k}$. En ce qui concerne la borne supérieure indiquée dans le Théorème 6, on peut la réécrire

$$k_{max} = \frac{1}{k} \left(\frac{1}{\mu} \alpha_1 \sin \alpha_1 - \cos \alpha_1 \right)$$

où α_1 est l'unique solution sur $]0, \pi[$ de $\tan \alpha_1 = \frac{-1}{1+\mu} \alpha_1$. En éliminant μ en fonction de α_1 dans cette dernière équation, on peut établir que

$$\begin{aligned} k_{max} - k_P &= \frac{1}{k\mu} (\alpha_1 \sin \alpha_1 - \mu \cos \alpha_1 - 1.2) \\ &= \frac{1}{k\mu} \left(\cos \alpha_1 + \frac{\alpha_1}{\sin \alpha_1} - 1.2 \right) \end{aligned}$$

Par construction, $\alpha_1 \in]0, \pi[$. On vérifie alors $k_{max} - k_P > 0$. Par conséquent, le gain proportionnel suggéré par la première méthode de Ziegler-Nichols satisfait toujours les hypothèses du Théorème 6 de stabilisation par un PID pour un système du premier ordre à retard. C'est une importante justification de cette méthode heuristique. En outre, on peut montrer par une étude détaillée (on se reportera à [68]) que les gains k_I et k_D défini dans (2.13) par cette même méthode satisfont également les hypothèses les concernant de ce même théorème. Plus précisément, on peut établir qu'ils les satisfont mais de manière possiblement non robuste lorsque τ devient grand (lorsque le retard Δ devient grand devant la constante de temps τ), i.e. ils sont près des frontières des domaines décrits dans le Théorème 6. C'est également un phénomène connu et observé en pratique qui engage souvent à modifier un peu les gains au prix de performances moindres.

2.4.3 Prédicteur de Smith

Le prédicteur de Smith est un outil de régulation pour les systèmes asymptotiquement stables en boucle ouverte retardés mono-variables (proposé dans [70]). Des généralisations aux cas multi-variables et pour les systèmes instables sont possibles mais alors il n'est plus possible de réaliser le bouclage avec un transfert rationnel. On utilise alors des bouclages à retards répartis faisant apparaître des fractions rationnelles en s et $e^{-\Delta s}$. Nous ne présentons ici que la version historique publiée par Smith.

Dans ce contexte, un retard sur l'entrée ou la sortie est équivalent. En effet les formes d'état suivantes possèdent la même fonction de transfert

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t - \Delta), y(t) = Cx(t) \\ \dot{x}(t) &= Ax(t) + Bu(t), y(t) = Cx(t - \Delta) \end{aligned}$$

où $x \in \mathbb{R}^n$, $u \in \mathbb{R}$ et $y \in \mathbb{R}$. Cette fonction de transfert est un produit d'une fraction rationnelle propre et d'un opérateur à retard

$$G(s) \triangleq G_0(s)e^{-\Delta s} = C(sI - A)^{-1}B e^{-\Delta s}$$

L'intérêt du prédicteur de Smith est de permettre d'utiliser un contrôleur K_0 conçu pour le système sans retard $G_0(s)$ en l'insérant dans le schéma général de la Figure 2.19. Les blocs additionnels \hat{G}_0 et $\hat{G}(s)$ représentent la connaissance du système. Idéalement, on a $\hat{G}_0 = G_0$ et $\hat{G} = G$. On suppose donc le transfert G_0 stable.

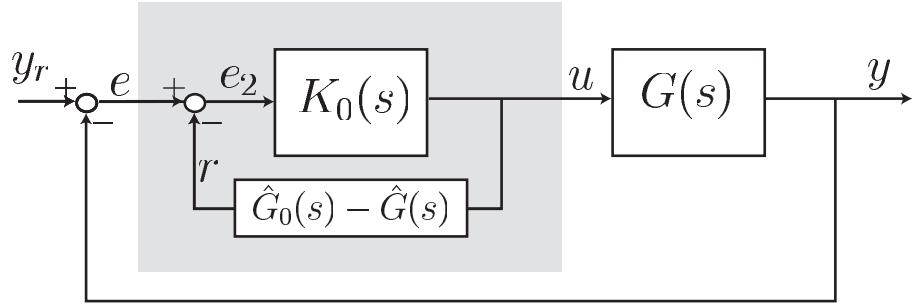


FIGURE 2.19 – Le prédicteur de Smith (zone grisée).

C'est sous cette hypothèse qu'on comprend le fonctionnement du prédicteur de Smith. Grâce à la boucle interne au prédicteur (dans la zone grisée de la Figure 2.19), le signal entrant dans le contrôleur K_0 est

$$e_2 = y_r - G(s)u - (G_0(s) - G(s))u = y_r - G_0(s)u$$

Ce terme est la différence entre la référence et la prédiction de la valeur de sortie du système à l'horizon Δ .

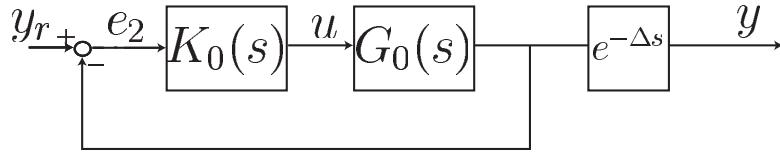


FIGURE 2.20 – Forme équivalente du système bouclé avec prédicteur de Smith.

Le comportement entrée-sortie du système bouclé par le prédicteur de Smith est équivalent à celui du système donné sur la Figure 2.20. Artificiellement, on a réussi à intercepter le signal de mesure avant le bloc retard et donc à nous ramener à un problème de régulation d'un système sans retard. On notera toutefois que le retard est toujours présent sur la sortie. Il n'a pas disparu du problème mais il n'interfère pas avec la régulation.

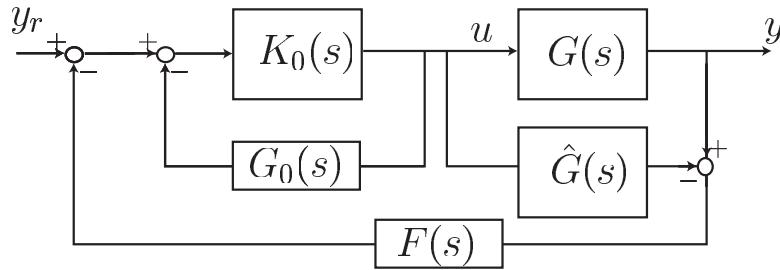


FIGURE 2.21 – Implémentation du prédicteur de Smith avec filtre additionnel de robustesse.

En pratique, le *prédicteur de Smith* souffre d'un manque de robustesse par rapport à une incertitude sur le retard, alors que la *robustesse* par rapport aux paramètres est moins problématique. En ce cas, \hat{G} et G sont différents, de même que \hat{G}_0 et G_0 . Il ne suffit pas toujours de calculer un contrôleur K_0 procurant de bonnes marges de stabilité au système G_0 bouclé (on pourra se reporter à [49, Chap 10.8] pour de nombreux contre-exemples). Certaines conditions suffisantes de robustesse sont

exposées dans [62] ; elles portent sur le choix du contrôleur K_0 mais sont difficiles à mettre en œuvre. Il est indispensable de l'utiliser dans la configuration proposée sur la Figure 2.21. On a rajouté dans la boucle externe un filtre “de robustesse” qu'on peut ajuster pour dégrader les performances mais assurer une stabilité en boucle fermée en dépit d'une incertitude sur le retard. Lorsque ce filtre $F(s)$ vaut 1 on retrouve le prédicteur de Smith des Figures 2.19 et 2.20.

2.4.4 Systèmes à non minimum de phase

Le transfert rationnel $G(s) = \frac{N(s)}{D(s)}$ où $N(s)$ et $D(s)$ sont deux polynômes premiers entre eux (i.e., pas de racine commune dans \mathbb{C}) est dit à *non minimum de phase* si l'un de ses zéros (i.e. l'une des racines de $N(s)$) est à partie réelle strictement positive. On parle alors de zéro instable. Ce type de systèmes est réputé difficile à contrôler. Ces systèmes sont souvent associés à une *réponse inverse* lors d'une variation en échelon de l'entrée¹⁴.

Pour comprendre la difficulté, rien ne vaut un petit exemple. Considérons le second ordre suivant

$$G(s) = \frac{1-s}{s^2}.$$

Ce transfert correspond, sous forme d'état, à un double intégrateur, typiquement un système mécanique régi par la loi de Newton,

$$\frac{d}{dt}x_1 = x_2, \quad \frac{d}{dt}x_2 = u$$

où la mesure y est une combinaison "contradictoire" de la position x_1 et la vitesse x_2 :

$$y = x_1 - x_2.$$

En partant de l'équilibre $x_1 = x_2 = y = u = 0$ pour les $t \leq 0$, considérons la réponse à un échelon unitaire $u = 1$ pour $t > 0$. Un calcul simple donne $y(t) = \frac{t^2}{2} - t$ pour $t > 0$. Ainsi au tout début y commence par être négatif pour à terme devenir positif. Si au lieu de $1 - s$ on avait eu $1 + s$, nous n'aurions pas eu cette *réponse inverse*. Ainsi, pour réguler y , un contrôleur proportionnel ne sait pas dans quelle direction partir. Le bouclage $u = -K_p y$ conduit toujours à une boucle fermée instable quelque soit le gain K_p choisi. En effet le système bouclé,

$$\frac{d}{dt}x_1 = x_2, \quad \frac{d}{dt}x_2 = -K_p x_1 + K_p x_2,$$

est instable dès que $K_p \neq 0$: la matrice $\begin{pmatrix} 0 & 1 \\ -K_p & K_p \end{pmatrix}$ est toujours instable avec une trace et un déterminant de même signe (voir le critère de Routh, condition 1.12). En transfert, on retrouve ce résultat en regardant les zéros de $1 + GK$ avec $K(s) = K_p$, soit $1 + K_p \frac{1-s}{s^2} = 0$ qui donne $s^2 - K_p s + K_p = 0$.

Avec un correcteur PI de transfert $K(s) = K_p + \frac{K_i}{s}$ (K_p gain proportionnel et K_i gain intégral) la situation est identique : $1 + GK = 0$ s'écrit $1 + \frac{1-s}{s^2} \left(K_p + \frac{K_i}{s} \right) = 0$ soit

$$s^3 - K_p s^2 + (K_p - K_i)s + K_i = 0.$$

Le critère de Routh (conditions 1.13) donne ici les conditions de stabilité

$$-K_p > 0, \quad K_p - K_i > 0, \quad K_i > 0, \quad -K_p(K_p - K_i) > K_i$$

14. Cette réponse inverse n'est pas cependant une caractérisation mathématique du fait qu'un zéro soit instable.

qui sont impossibles à satisfaire quelques soient les valeurs de K_p et K_d .

Il est possible, pour stabiliser, de remplacer le terme "intégral" par un terme "dérivé"

$$K(s) = K_p + K_d s$$

de gain K_d . Maintenant, $1 + GK = 0$ donne l'équation d'ordre 2 suivante :

$$(1 - K_d)s^2 + (K_d - K_p)s + K_p = 0$$

Ces racines sont stables si, et seulement si,

$$\frac{K_d - K_p}{1 - K_d} > 0, \quad \frac{K_p}{1 - K_d} > 0.$$

Un raisonnement simple montre que ces conditions de stabilité sont équivalentes à

$$0 < K_d < 1, \quad 0 < K_p < K_d,$$

Ces deux inégalités définissent un triangle. Le fait que ce domaine de stabilité soit si peu étendu dans le plan (K_p, K_d) est une indication de petites marges de robustesse. Pour s'en convaincre, il suffit de tracer le lieu de Nyquist avec $K_d = 1/2$ et $K_p = 1/4$, par exemple. De plus un terme "dérivé" amplifie notablement les bruits de mesure et donc il vaut mieux prendre K_d petit.

Pour les systèmes d'ordre plus élevé, une telle analyse avec Routh n'est pas très commode. Il est alors utile de tracer le lieu de Nyquist de GK pour certaines valeurs des paramètres du correcteur K et d'analyser les marges de gain et de retard.

Chapitre 3

Commandabilité, stabilisation, feedback

Un système commandé $\frac{d}{dt}x = f(x, u)$ est un système sous-déterminé. La différence entre le nombre d'équations (indépendantes) et le nombre de variables donne le nombre de commandes indépendantes $m = \dim u$. Il faut noter que le degré de sous-détermination est ici infini car on ne manipule pas des scalaires mais des *fonctions* du temps. L'étude des systèmes sous-déterminés d'équations différentielles ordinaires est d'une nature différente de celle des systèmes déterminés qui ont été l'objet du Chapitre 1. Néanmoins, toutes les notions que nous y avons vues vont nous être utiles.

La particularité de ces systèmes sous-déterminés est qu'on peut utiliser la commande pour agir sur eux. En pratique, deux problèmes nous intéressent : la *planification* et le *suivi de trajectoires*.

La notion clef de ces problèmes est la *commandabilité*. Dans ce chapitre, après une rapide définition de cette propriété dans le cas non linéaire général (voir Définition 12), nous étudions en détails les systèmes linéaires $\frac{d}{dt}x = Ax + Bu$. Leur commandabilité est caractérisée par le critère de Kalman (voir Théorème 22). Nous présentons alors deux méthodes pour résoudre les problèmes qui nous préoccupent.

La première méthode utilise la *forme normale* dite de Brunovsky. L'écriture sous cette forme met en lumière un paramétrage explicite de toutes les trajectoires en fonctions de m fonctions scalaires arbitraires $t \mapsto y(t)$ et d'un nombre fini de leurs dérivées. Ces quantités y , dites *sorties de Brunovsky*, sont des combinaisons linéaires de x . Elles permettent de calculer très simplement des commandes u faisant transiter le système d'un état vers un autre. Le *problème de planification de trajectoire* est ainsi résolu. Ce point est détaillé à la Section 3.3.4. Les sorties de Brunovsky permettent également de construire le bouclage ("feedback") qui assure le *suivi asymptotique d'une trajectoire de référence* arbitraire. La convergence est obtenue par la technique de *placement de pôles*. On se reportera au Théorème 24.

La seconde méthode s'attache à optimiser les trajectoires. On caractérise le contrôle en boucle ouverte permettant d'aller d'un état vers un autre tout en minimisant un critère quadratique. On calcule également la commande minimisant l'écart quadratique entre la trajectoire de référence et la trajectoire réelle. On montre que la solution optimale est donnée par un feedback dont les gains sont calculés à partir d'une équation matricielle de Riccati. C'est la *commande linéaire quadratique*. Cette technique est détaillée à la Section 3.4.

3.1 Un exemple de planification et de suivi de trajectoires

Nous considérons un système de deux oscillateurs de fréquences différentes soumis en parallèle à un même contrôle. Comme nous le verrons dans l'Exemple 15, ce modèle est d'une portée assez

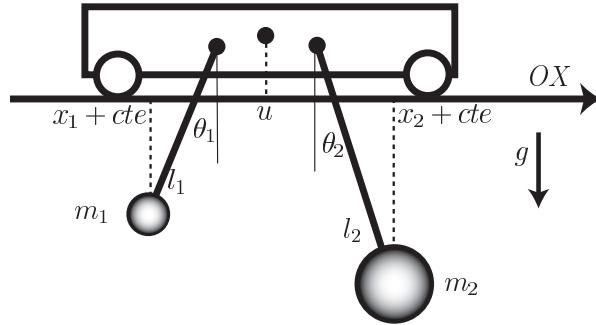


FIGURE 3.1 – Deux pendules accrochés au même chariot d'abscisse u , le contrôle.

générale. Il correspond, à une réécriture près, à la dynamique linéarisée d'un système quantique à trois niveaux dont les transitions entre le niveau fondamental et les deux niveaux excités (d'énergies différentes) sont contrôlées par la lumière d'un laser.

3.1.1 Modélisation de deux oscillateurs en parallèle

On considère deux pendules ponctuels accrochés à un même chariot de position u (la commande), tels que représentés sur la Figure 3.1. On note θ_i et x_i , l'inclinaison et l'abscisse¹ de la masse du pendule numéro i , $i = 1, 2$. Autour de l'équilibre stable, les équations de Newton linéarisées sont

$$\frac{d^2}{dt^2}x_1 = -a_1(x_1 - u), \quad \frac{d^2}{dt^2}x_2 = -a_2(x_2 - u) \quad (3.1)$$

où $a_1 = g/l_1$ et $a_2 = g/l_2$ sont deux paramètres positifs correspondant aux carrés des pulsations (l_1 et l_2 correspondant aux longueurs supposées différentes des pendules $l_1 \neq l_2$).

3.1.2 Planification de trajectoires

Soit $T > 0$ et un déplacement désiré D . On souhaite trouver un contrôle en boucle ouverte qui amène le système de l'équilibre en $x_1 = x_2 = u = 0$ à $t = 0$ vers l'équilibre $x_1 = x_2 = u = D$ en $t = T$.

Pour trouver de telles trajectoires, le plus simple est d'introduire la nouvelle variable

$$z = \frac{x_1}{a_1} - \frac{x_2}{a_2}$$

Des calculs élémentaires donnent alors les relations suivantes

$$\begin{cases} z = \frac{x_1}{a_1} - \frac{x_2}{a_2} \\ z^{(2)} = x_2 - x_1 \\ z^{(4)} = (a_1 - a_2)u + a_1x_1 - a_2x_2 \end{cases} \quad (3.2)$$

où x_1 , x_2 et u s'expriment explicitement en fonction de z , $z^{(2)}$ et $z^{(4)}$ dès que $a_1 \neq a_2$. Les deux équations (3.1), qui ne font intervenir que les trois variables (x_1, x_2, u) , s'obtiennent en éliminant z

1. En fait les vraies abscisses sont $x_1 + c_1$ et $x_2 + c_2$, c_1 et c_2 constantes, au lieu de x_1 et x_2 . $u + c_1$ et $u + c_2$ sont les abscisses des points auxquels sont suspendus les deux pendules. Cependant les constants c_i disparaissent dans les équations de Newton.

des trois équations précédentes. Ainsi, elles sont bien contenues dans les trois équations ci-dessus. En fait (3.2) est une description équivalente de la dynamique du système décrit initialement par (3.1) mais avec une équation de plus et une variable de plus z .

L'intérêt de rajouter z vient du fait que si l'on calcule x_1 , x_2 et u avec les formules (3.2), alors les fonctions du temps $x_1(t)$, $x_2(t)$ et $u(t)$ vérifient automatiquement les équations différentielles (3.1). Réciproquement, si les fonctions du temps $x_1(t)$, $x_2(t)$ et $u(t)$ vérifient (3.1) alors en posant $z = a_2x_1 - a_1x_2$, on obtient encore x_1 , x_2 et u via (3.2). Ainsi, on paramétrise via z qui est une fonction arbitraire du temps, dérivable au moins 4 fois, toutes les solutions de (3.1). Le raisonnement précédent montre clairement que nous n'en manquons aucune en les décrivant comme des combinaisons linéaires de z , $z^{(2)}$ et $z^{(4)}$.

Les trois conditions d'équilibre en $t = 0$, $x_1 = x_2 = u = 0$, imposent la valeur 0 à z et à ses dérivées jusqu'à l'ordre 3 (ordre 4 si on ne tolère pas de discontinuités sur u) : $z^{(i)}(0) = 0$ pour $i = 0, 1, 2, 3, 4$. De même en $t = T$, on a $z(T) = \left(\frac{1}{a_1} - \frac{1}{a_2}\right)D$ et $z^{(i)}(T) = 0$ pour $i = 1, 2, 3, 4$. Pour $t \neq 0, T$, $z(t)$ est libre. Il faut cependant respecter la continuité pour z et ses dérivées jusqu'à l'ordre 4 (ou 3 si tolère des sauts sur le contrôle u). Une des fonctions les plus simples $z(t)$ qui respectent ces conditions est la suivante

$$z(t) = \left(\frac{1}{a_1} - \frac{1}{a_2}\right) D \phi\left(\frac{t}{T}\right) \quad (3.3)$$

où ϕ est la fonction C^4 croissante de \mathbb{R} dans $[0, 1]$ définie par

$$\phi(\sigma) = \begin{cases} 0, & \text{si } \sigma \leq 0; \\ \frac{\sigma^5}{\sigma^5 + (1-\sigma)^5}, & \text{si } 0 \leq \sigma \leq 1; \\ 1, & \text{si } 1 \leq \sigma \end{cases}$$

Nous avons obtenu explicitement une trajectoire allant de l'équilibre d'abscisse 0 à l'équilibre d'abscisse D en temps fini T arbitraire. Si on utilise le *contrôle en boucle ouverte* ($z(t)$ est donnée par (3.3))

$$u(t) = \frac{z(t) + \left(\frac{1}{a_1} - \frac{1}{a_2}\right) z^{(2)}(t) + \frac{1}{a_1 a_2} z^{(4)}(t)}{\frac{1}{a_1} - \frac{1}{a_2}}$$

alors la solution de (3.1) qui passe en $t = 0$ par l'équilibre $x_1 = x_2 = 0$, $\frac{d}{dt}x_1 = \frac{d}{dt}x_2 = 0$, passe en $t = T$ par l'équilibre $x_1 = x_2 = D$, $\frac{d}{dt}x_1 = \frac{d}{dt}x_2 = 0$. On connaît même explicitement cette solution avec

$$x_1(t) = \frac{z(t) + \frac{1}{a_2} z^{(2)}(t)}{\frac{1}{a_1} - \frac{1}{a_2}}$$

$$x_2(t) = \frac{z(t) + \frac{1}{a_1} z^{(2)}(t)}{\frac{1}{a_1} - \frac{1}{a_2}}$$

Nous avons résolu le problème de *planification de trajectoires*.

3.1.3 Stabilisation et suivi de trajectoires

Supposons donnée une trajectoire de référence (par exemple issue de la construction précédente) notée $t \mapsto (x_{1,r}(t), x_{2,r}(t), u_r(t))$ solution de (3.1). On cherche un *feedback* qui assure le *suivi asymptotique* de cette trajectoire. En d'autres termes, nous cherchons des corrections sur le contrôle en

boucle ouverte $u_r(t)$ pour compenser les déviations entre la trajectoire réelle et la trajectoire de référence. Ici, l'état du système est formé par les positions (x_1, x_2) et les vitesses (v_1, v_2) ($v_i = \frac{dx_i}{dt}$, $i = 1, 2$). On note $\Delta x_i = x_i - x_{i,r}$ et $\Delta v_i = v_i - v_{i,r}$ les déviations en position et en vitesse par rapport à la référence ($i = 1, 2$). On cherche Δu en fonction des Δx_i et Δv_i . Le *contrôle en boucle fermée* u qu'on appliquera sera $u_r + \Delta u$. Dans le cas particulier où la trajectoire de référence correspond à un point d'équilibre, on parle alors de *stabilisation* plutôt que de suivi.

Comme les équations du système sont linéaires, un calcul direct montre que Δx_i et Δu vérifient les mêmes équations que x_i et u

$$\frac{d^2}{dt^2} \Delta x_1 = -a_1(\Delta x_1 - \Delta u), \quad \frac{d^2}{dt^2} \Delta x_2 = -a_2(\Delta x_2 - \Delta u)$$

Aussi, il suffit de résoudre le problème de la *stabilisation* pour obtenir une solution au *suivi*. Supposons donc que la trajectoire de référence est nulle $x_{1,r} = x_{2,r} = u_r = 0$. La façon la plus simple de concevoir le feedback stabilisant est alors d'utiliser une méthode de Lyapounov. Utilisons comme fonction candidate à être de Lyapounov l'énergie du système. Pour $u = 0$, l'énergie du système,

$$V = \frac{1}{2}(v_1)^2 + \frac{a_1}{2}(x_1)^2 + \frac{1}{2}(v_2)^2 + \frac{a_2}{2}(x_2)^2,$$

reste constante le long des trajectoires. Donc, comme les équations sont linéaires par rapport à u , la dérivée de V le long des trajectoires est le produit de deux termes : le premier facteur est u et le second est une fonction des positions et vitesses uniquement. Il suffit alors de prendre u du signe opposé au second facteur (ce qui est facile comme nous allons le voir) pour faire décroître cette énergie. On conclut alors par le Théorème 11 de LaSalle.

Faisons ici les calculs. On a

$$\frac{d}{dt} V = (a_1 v_1 + a_2 v_2)u$$

Avec k paramètre strictement positif, on considère le feedback suivant

$$u = K(v_1, v_2) = \begin{cases} u^{\min}, & \text{si } -k(a_1 v_1 + a_2 v_2) \leq u^{\min}; \\ -k(a_1 v_1 + a_2 v_2), & \text{si } u^{\min} \leq -k(a_1 v_1 + a_2 v_2) \leq u^{\max}; \\ u^{\max}, & \text{si } u^{\max} \leq -k(a_1 v_1 + a_2 v_2) \end{cases} \quad (3.4)$$

On suppose bien sûr que les contraintes sur u sont telles que $u^{\min} < 0 < u^{\max}$.

Avec le feedback précédent, la fonction V est toujours positive, infinie à l'infini et maintenant elle décroît le long des trajectoires. On peut donc appliquer le Théorème 11 de LaSalle, qui caractérise l'ensemble vers lequel tendent les trajectoires : les solutions du système en boucle fermée avec le contrôle 3.4 qui vérifient en plus $\frac{d}{dt} V = 0$ c.-à-d. $a_1 v_1 + a_2 v_2 = 0$. Dans ce cas $u = 0$ et donc

$$a_1 \frac{d}{dt} v_1 + a_2 \frac{d}{dt} v_2 = -(a_1)^2 x_1 - (a_2)^2 x_2 = 0$$

En dérivant encore deux fois, on a

$$-(a_1)^4 x_1 - (a_2)^4 x_2 = 0$$

Comme $a_1 \neq a_2$ et $a_1, a_2 > 0$, on déduit que, nécessairement, $x_1 = x_2 = 0$ et donc tout est nul. Les trajectoires convergent toutes vers 0. Notre feedback stabilise globalement asymptotiquement le système en 0.

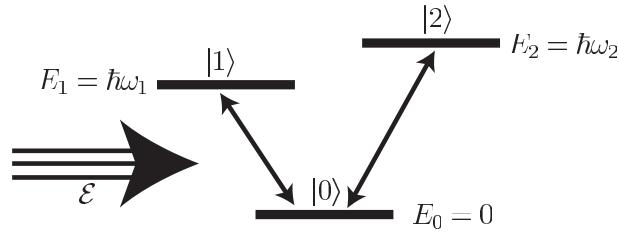


FIGURE 3.2 – Un atome à trois niveaux en interaction avec une lumière laser polarisée linéairement et associée au champ électrique $\mathcal{E} \in \mathbb{R}$, le contrôle.

D'après ce qui précède, le contrôle complet avec suivi de trajectoire s'écrit très simplement. On peut gérer très simplement des contraintes sur u : il suffit de prendre une trajectoire de référence donnant un contrôle u_r qui ne passe par trop près des bornes. Cette hypothèse est vérifiée s'il existe $\epsilon > 0$ tel que pour tout temps t , $u_r(t) \in [u^{\min} + \epsilon, u^{\max} - \epsilon]$. Dans ces conditions, le contrôle

$$u = u_r + \begin{cases} -\epsilon, & \text{si } -k(a_1\Delta v_1 + a_2\Delta v_2) \leq -\epsilon; \\ -k(a_1\Delta v_1 + a_2\Delta v_2), & \text{si } -\epsilon \leq -k(a_1\Delta v_1 + a_2\Delta v_2) \leq \epsilon; \\ \epsilon, & \text{si } \epsilon \leq -k(a_1\Delta v_1 + a_2\Delta v_2). \end{cases}$$

est à chaque instant dans $[u^{\min}, u^{\max}]$ et de plus assure la convergence asymptotique vers la trajectoire de référence.

3.1.4 Autres exemples

Exemple 15 (Un atome à trois niveaux). *La fonction d'onde ψ de l'atome à trois niveaux schématisé sur la Figure 3.2 obéit, sous l'approximation dipolaire électrique et dans un champ électrique $\mathcal{E}(t)$ polarisé linéairement, à l'équation de Schrödinger*

$$\imath\hbar\frac{d}{dt}\psi = (H_0 + \mathcal{E}(t)H_1)\psi$$

où $\psi \in \mathbb{C}^3$ est la fonction d'onde appartenant à l'espace de Hilbert \mathbb{C}^3 et où H_0 et H_1 sont les opérateurs auto-adjoints décrits par les matrices

$$H_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \hbar\omega_1 & 0 \\ 0 & 0 & \hbar\omega_2 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 0 & \mu & \mu \\ \mu & 0 & 0 \\ \mu & 0 & 0 \end{pmatrix}$$

où ω_1 , ω_2 et μ sont des paramètres réels et strictement positifs. Avec les notations usuelles en mécanique quantique

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad |2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

on voit que $|0\rangle$ est l'état d'énergie $E_0 = 0$, $|1\rangle$ celui d'énergie $E_1 = \hbar\omega_1$ et $|2\rangle$ celui d'énergie $E_2 = \hbar\omega_2$, $\hbar = h/(2\pi)$ étant la constante de Plank. Les pulsations de transitions (les raies atomiques donc) entre le niveau fondamental $|0\rangle$ et les deux niveaux excités $|1\rangle$ et $|2\rangle$ ont pour pulsations ω_1 et ω_2 .

On a donc $\psi = \psi_0 |0\rangle + \psi_1 |1\rangle + \psi_2 |2\rangle$ avec ψ_0, ψ_1 et ψ_2 trois nombres complexes dont le carré du module représente la probabilité pour que l'atome soit dans le niveau $|0\rangle, |1\rangle$ ou $|2\rangle$, respectivement. L'équation de Schrödinger pour ce système s'écrit

$$\begin{aligned}\imath \frac{d}{dt} \psi_0 &= \frac{\mu}{\hbar} \mathcal{E} (\psi_1 + \psi_2) \\ \imath \frac{d}{dt} \psi_1 &= \omega_1 \psi_1 + \frac{\mu}{\hbar} \mathcal{E} \psi_0 \\ \imath \frac{d}{dt} \psi_2 &= \omega_2 \psi_2 + \frac{\mu}{\hbar} \mathcal{E} \psi_0\end{aligned}$$

On constate que $\psi_0 = 1, \psi_1 = \psi_2 = 0$ et $\mathcal{E} = 0$ est un régime d'équilibre (état fondamental, laser éteint). Linéarisons les équations autour de cet équilibre. En notant $\delta\psi_i, i = 0, 1, 2$, et $\delta\mathcal{E}$ les petits écarts, on obtient les équations linéarisées

$$\imath \frac{d}{dt} \delta\psi_0 = 0, \quad \imath \frac{d}{dt} \delta\psi_1 = \omega_1 \delta\psi_1 + \frac{\mu}{\hbar} \delta\mathcal{E}, \quad \imath \frac{d}{dt} \delta\psi_2 = \omega_2 \delta\psi_2 + \frac{\mu}{\hbar} \delta\mathcal{E}.$$

Il ne faut pas oublier que les $\delta\psi_i$ sont des complexes. Au premier ordre $\delta\psi_0$ est constant. C'est une partie de l'état qui n'est pas influencée par le contrôle (au moins au premier ordre). Cette partie est non commandable (au premier ordre). Elle ne nous gène pas, elle correspond au fait que le module de ψ vaut 1 (conservation de la probabilité). Nous allons l'oublier ici. On a donc le système dans \mathbb{C}^2 , c'est à dire dans \mathbb{R}^4 suivant

$$\imath \frac{d}{dt} \delta\psi_1 = \omega_1 \delta\psi_1 + \frac{\mu}{\hbar} \delta\mathcal{E}, \quad \imath \frac{d}{dt} \delta\psi_2 = \omega_2 \delta\psi_2 + \frac{\mu}{\hbar} \delta\mathcal{E}$$

En posant

$$\delta\psi_1 = \frac{\mu (\omega_1 x_1 + \imath \frac{d}{dt} x_1)}{\hbar(\omega_1)^2}, \quad \delta\psi_2 = \frac{\mu (\omega_2 x_2 + \imath \frac{d}{dt} x_2)}{\hbar(\omega_2)^2}, \quad \delta\mathcal{E} = -u$$

on obtient exactement les équations des deux pendules (3.1) pour x_1, x_2 et u avec $a_1 = (\omega_1)^2$ et $a_2 = (\omega_2)^2$. Il s'agit en fait du même système. Le carré du module de $\delta\psi_i, |\delta\psi_i|^2$ correspond alors, à une constante près, à l'énergie mécanique du pendule numéro $i : \frac{a_i}{2}(x_i)^2 + (\frac{d}{dt} x_i)^2$.

Il est possible de reprendre les calculs de transitoires faits pour les deux pendules et de les adapter pour traiter le problème suivant. On part de l'état fondamental $|0\rangle$ en $t = 0$. On cherche la forme de l'impulsion laser $\mathcal{E}(t)$ qui assure le transfert en $t = T$ d'une petite fraction $0 < p_1 \ll 1$ de la probabilité sur l'état $|1\rangle$ en gardant, au final, nulle celle de l'état $|2\rangle$ ². Ce problème se traduit de la façon suivante : en $t = 0$ on part de $\delta\psi_1 = \delta\psi_2 = 0$ avec $\mathcal{E} = 0$. En $t = T$, on souhaite arriver en $\delta\psi_1 = \sqrt{p_1}, \delta\psi_2 = 0$ avec $\mathcal{E} = 0$. Cela se traduit dans les variables x_1 et x_2 par des conditions suivantes. En $t = 0$, tout est nul. En $t = T$, on doit avoir

$$x_1 = \bar{x}_1 = \sqrt{p_1} \frac{\hbar \omega_1}{\mu}, \quad v_1 = x_2 = v_2 = u = 0$$

Sur la variable z cela donne comme condition en $t = 0$ (voir (3.3))

$$z = z^{(1)} = z^{(2)} = z^{(3)} = z^{(4)} = 0$$

2. L'intérêt éventuel d'un transfert dans ces conditions vient du fait que nous ne supposons pas le contrôle résonnant comme il est usuel de le faire. Nous n'utilisons pas ici l'approximation du champ tournant et les oscillations de Rabi.

et en $t = T$,

$$z = \frac{\bar{x}_1}{a_1}, \quad z^{(1)} = z^{(3)} = 0, \quad z^{(2)} = -\bar{x}_1, \quad z^{(4)} = a_1 \bar{x}_1$$

Il suffit alors de prendre

$$z(t) = \bar{x}_1 \left(\frac{1}{a_1} - \frac{(t-T)^2}{2} + \frac{a_1(t-T)^4}{4!} \right) \phi \left(\frac{t}{T} \right)$$

et de calculer le contrôle en boucle ouverte par la formule

$$\mathcal{E}(t) = - \frac{z(t) + \left(\frac{1}{a_1} - \frac{1}{a_2} \right) z^{(2)}(t) + \frac{1}{a_1 a_2} z^{(4)}(t)}{\frac{1}{a_1} - \frac{1}{a_2}}$$

Comme $p_1 \ll 1$, il est facile de garantir en prenant un temps T assez grand que $|\delta\psi_1(t)|$ et $|\delta\psi_2(t)|$ restent petits devant 1 tout au long du transfert. C'est important pour que notre modèle linéarisé reste valable et que le calcul précédent ait un sens.

Nous ne traiterons pas le suivi de trajectoire car la notion de feedback pour un système quantique est en cours d'élaboration. La rétro-action quantique se heurte au problème de la mesure. En effet, toute mesure perturbe notamment le système et donc il n'est plus possible de dissocier l'appareil de mesure du système lui-même et de supposer que la perturbation due à la mesure est arbitrairement petite. Pour prendre en compte cette spécificité, on distingue actuellement deux types de rétro-actions quantiques [77] :

1. la rétro-action fondée sur la mesure (*measurement-based feedback*) : le système est quantique ; le processus de mesure associée à la sortie y est quantique, perturbe le système avec un effet intrinsèque de rétro-action et fournit des informations classiques au contrôleur ; le contrôleur est classique et ajuste en temps réel le contrôle u qui est une grandeur classique ; la première réalisation expérimentale d'une telle boucle de feedback d'état (*observateur-contrôleur*) a été faite par le groupe de Serge Haroche et Jean-Michel Raimond [66] pour réguler le nombre (≤ 10) de photons piégés entre deux miroirs.
2. la rétro-action cohérente (*coherent feedback*) : le système et le contrôleur sont des systèmes quantiques ; les effets dissipatifs et irréversibles qui assurent la stabilité asymptotique sont liés à des processus de décohérence et de mesure intervenant surtout sur le contrôleur ; cette notion de feedback s'inspire notamment de l'ingénierie quantique de réservoir (*quantum reservoir engineering*) proposée dans [64] ; pour des travaux récents à la fois en physique et en automatique voir par exemple [38, 41].

Un lecteur, désirant mieux comprendre les principes physiques sur lesquels reposent les systèmes quantiques ouverts soumis aux effets irréversibles dus à la mesure et à la décohérence, pourra consulter utilement [32] ainsi que les cours de Serge Haroche au Collège de France.

Exemple 16 (Le “hockey puck”). Étant donné un corps rigide (de masse unitaire) dans le plan, dont le centre de gravité est repéré par les coordonnées (x, y) , dont l'orientation est donnée par l'angle θ , et dont l'inertie est notée J . Ce système est soumis exclusivement³ à une force (commande) u s'exerçant en un point bien déterminé (situé à une distance r du centre de gravité) de ce corps. Il est représenté

3. Il s'agit d'un “hockey-puck”, c.-à-d. un palet sur une table à coussin d'air

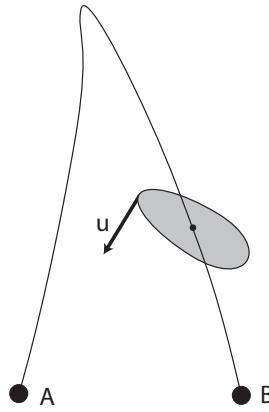


FIGURE 3.3 – Le “hockey puck” : un solide soumis à une force u (commande). Pour effectuer des déplacements (entre A et B), il faut le faire tourner en le faisant avancer puis le retourner brutalement avant de le faire avancer à nouveau pour enfin le freiner en prenant en compte la rotation induite par la décélération.

sur la Figure 3.3. Les équations de la dynamique sont

$$\begin{aligned}\frac{d^2}{dt^2}x &= -u \sin \theta \\ \frac{d^2}{dt^2}y &= u \cos \theta \\ J \frac{d^2}{dt^2}\theta &= ru\end{aligned}$$

Ce système mécanique est sous-actionné. Il possède 3 degrés de liberté et une seule commande. On souhaite pourtant le déplacer d'une configuration à une autre (c.-à-d. d'un point de départ et d'une orientation à un point d'arrivée et à une orientation possiblement différente). On ne dispose que d'une seule commande. Pour réaliser ces manœuvres, on est obligé d'utiliser des stratégies compliquées. Par exemple, on ne peut pas déplacer latéralement le système sans engendrer de rotation. Pour arrêter le solide lorsqu'il a acquis de la vitesse, il faut le faire tourner et adapter la force suivant l'angle de rotation. On se reportera à [59] pour un exposé détaillé.

3.2 Commandabilité non linéaire

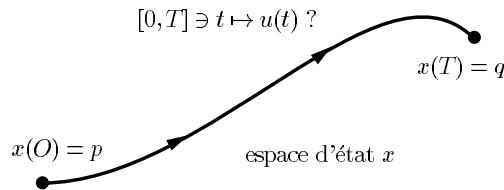


FIGURE 3.4 – Planification de trajectoire entre deux états p et q .

On considère le système (f étant une fonction régulière)

$$\frac{d}{dt}x(t) = f(x(t), u(t)), \quad x(t) \in \mathbb{R}^n, \quad u(t) \in \mathbb{R}^m \quad (3.5)$$

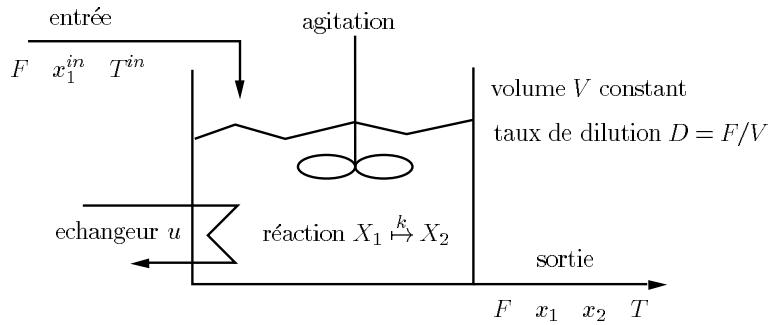


FIGURE 3.5 – Un réacteur chimique exothermique où u correspond aux échanges thermiques avec l’extérieur.

Définition 11 (Trajectoire d’un système commandé). *On appelle trajectoire du système (3.5) toute fonction régulière $I \ni t \mapsto (x(t), u(t)) \in \mathbb{R}^n \times \mathbb{R}^m$ qui satisfait identiquement sur un intervalle d’intérieur non vide I de \mathbb{R} les équations (3.5).*

Définition 12 (Commandabilité). *Le système (3.5) est dit commandable en temps $T > 0$ si et seulement si pour tout $p, q \in \mathbb{R}^n$, il existe une loi horaire $[0, T] \ni t \mapsto u(t) \in \mathbb{R}^m$, appelée commande en boucle ouverte, qui amène le système de l’état $x(0) = p$ à l’état $x(T) = q$, i.e., telle que la solution du problème de Cauchy*

$$\begin{aligned}\frac{d}{dt}x(t) &= f(x(t), u(t)) \quad \text{pour } t \in [0, T] \\ x(0) &= p\end{aligned}$$

vérifie $x(T) = q$. Le système est dit simplement commandable lorsqu’il est commandable pour au moins un temps $T > 0$.

D’autres définitions sont possibles : elles correspondent toutes à des variantes⁴ plus ou moins subtiles de la Définition 12. Comme l’illustre la Figure 3.4, la commandabilité exprime une propriété topologique très naturelle. En général, la *commande en boucle ouverte* $[0, T] \ni t \mapsto u(t)$ n’est pas unique, il en existe une infinité. Cette étape s’appelle *planification de trajectoire*. Calculer $t \mapsto u(t)$ à partir de la connaissance de f , p et q constitue l’une des questions majeures de l’Automatique. Cette question est loin d’être complètement résolue actuellement.

Très souvent, l’absence de commandabilité est due à l’existence d’*intégrales premières* non triviales. Ce sont des quantités qui restent constantes le long de toute trajectoire. Nous allons les étudier.

3.2.1 Exemple de non commandabilité par la présence d’intégrale première

Considérons le réacteur exothermique de la Figure 3.5. Les équations de bilan matière et d’énergie donnent les équations différentielles suivantes

$$\begin{aligned}\frac{d}{dt}x_1 &= D(x_1^{in} - x_1) - k_0 \exp(-E/RT)x_1 \\ \frac{d}{dt}x_2 &= -Dx_2 + k_0 \exp(-E/RT)x_1 \\ \frac{d}{dt}T &= D(T^{in} - T) + \alpha \Delta H \exp(-E/RT)x_1 + u\end{aligned}\tag{3.6}$$

4. À titre d’exemple, on pourra considérer la notion de commandabilité en temps petits, qui demande que toutes les directions soient accessibles en temps court. Un contre exemple est constitué par une voiture n’ayant pas de marche arrière, qui bien que pouvant, au prix d’un certain nombre de manœuvres, atteindre tous les points de son espace d’état, ne peut se déplacer latéralement.

On reconnaît l'effet non linéaire essentiel de la *loi d'Arrhenius*

$$k = k_0 \exp(-E/RT)$$

qui relie la constante de vitesse k à la température T . Comme nous l'avons écrit, la cinétique est supposée du premier ordre. Les *constantes* physiques usuelles (D , x_1^{in} , k_0 , E , T^{in} , α et ΔH) sont toutes positives. La commande u est proportionnelle à la puissance thermique échangée avec l'extérieur. On note x_i la concentration de l'espèce chimique X_i , $i = 1, 2$. Il est assez facile de voir que ce système n'est pas commandable. En effet, le bilan global sur $X_1 + X_2$, élimine le terme non linéaire pour donner

$$\frac{d}{dt}(x_1 + x_2) = D(x_1^{in} - x_1 - x_2).$$

La quantité $\xi = x_1 + x_2$ vérifie une équation différentielle autonome $\frac{d}{dt}\xi = D(x_1^{in} - \xi)$. Donc

$$\xi = x_1^{in} + (\xi_0 - x_1^{in}) \exp(-Dt) \quad (3.7)$$

où ξ_0 est la valeur initiale de ξ . Si, dans la Définition 12, on prend l'état initial p tel que $\xi = x_1 + x_2 = x_1^{in}$ et q tel que $\xi = x_1 + x_2 = 0$, il n'existe pas de commande qui amène le système de p vers q . En effet, pour toute trajectoire démarrant en un tel p , la quantité $x_1 + x_2$ reste constante et égale à x_1^{in} . Cette partie non commandable du système représentée par la variable ξ admet ici un sens physique précis. Elle est bien connue des chimistes qui la nomment *invariant chimique*.

L'exemple ci-dessus nous indique que l'absence de commandabilité peut-être liée à l'existence d'invariants, i.e., à des combinaisons des variables du système et éventuellement du temps, qui sont conservées le long de toute trajectoire. Pour (3.6), il s'agit de $(x_1 + x_2 - x_1^{in}) \exp(Dt)$ comme le montre (3.7). Nous définissons la notion d'intégrale première pour les systèmes commandés.

Définition 13 (Intégrale première). Une fonction régulière $\mathbb{R} \times \mathbb{R}^n \ni (t, x) \mapsto h(t, x) \in \mathbb{R}$ est appelée *intégrale première du système* (3.5), si elle est constante le long de toute trajectoire du système. Une intégrale première est dite triviale si c'est une fonction constante sur $\mathbb{R} \times \mathbb{R}^n$.

Si h est une intégrale première, sa dérivée le long d'une trajectoire arbitraire est nulle pour toute trajectoire $t \mapsto (x(t), u(t))$ du système, c.-à-d.

$$\frac{d}{dt}h = \frac{\partial h}{\partial t} + \frac{\partial h}{\partial x} \frac{dx}{dt} \equiv 0$$

Si (3.5) admet une intégrale première non triviale $t \mapsto h(t, x)$ alors (3.5) n'est pas commandable. Sinon, il existe $T > 0$, tel que pour tout $p, q \in \mathbb{R}^n$ et tout instant initial t , $h(t, p) = h(t + T, q)$ (il existe une trajectoire reliant p à q sur $[t, t + T]$). Donc h est une fonction périodique du temps et indépendante de x . Mais alors la dérivée de h le long des trajectoires du système correspond à $\frac{\partial}{\partial t}h$. Comme elle est nulle, h est une constante, ce qui contredit l'hypothèse. Nous avons montré la proposition suivante

Proposition 7. Si le système (3.5) est commandable, alors ses intégrales premières sont triviales.

Il est possible de caractériser, à partir de f et d'un nombre fini de ses dérivées partielles, l'existence d'intégrales premières non triviales. Nous allons nous restreindre dans ce cours au cas linéaire. La démarche est transposable au cas non linéaire mais elle conduit alors à des calculs plus lourds qui, pour être présentés de façon compacte, nécessitent le langage de la géométrie différentielle et des crochets de Lie (voir par exemple [42]).

3.3 Commandabilité linéaire

Nous considérons ici les systèmes linéaires stationnaires du type

$$\frac{d}{dt}x = Ax + Bu \quad (3.8)$$

où l'état $x \in \mathbb{R}^n$, la commande $u \in \mathbb{R}^m$ et les matrices A et B sont constantes et de tailles $n \times n$ et $n \times m$, respectivement.

3.3.1 Matrice de commandabilité et intégrales premières

Supposons que (3.8) admette une intégrale première $h : \mathbb{R} \times \mathbb{R}^n \ni (t, x) \mapsto h(t, x) \in \mathbb{R}$. Soit le changement de variables sur x défini par $x = \exp(tA)z$. Avec les variables (z, u) , (3.8) devient

$$\frac{d}{dt}z = \exp(-tA)Bu \quad (3.9)$$

car d'après la Proposition 1, $\frac{d}{dt}(\exp(tA)) = \exp(tA)A = A\exp(tA)$. Dans l'équation (3.9), il est notable que le second membre ne dépend que de u (et de t mais pas de z). L'intégrale première devient $h(t, \exp(tA)z) = l(t, z)$. Comme la valeur de l est constante le long de toute trajectoire $t \mapsto (z(t), u(t))$, nous avons

$$\frac{d}{dt}l = \frac{\partial l}{\partial t} + \frac{\partial l}{\partial z} \frac{d}{dt}z = 0$$

Comme $\frac{d}{dt}z = \exp(-tA)Bu$, pour toute valeur de z et u , on a l'identité suivante

$$\frac{\partial l}{\partial t}(t, z) + \frac{\partial l}{\partial z}(t, z)\exp(-tA)Bu \equiv 0$$

En prenant, $u = 0$, z et t arbitraires, on en déduit

$$\frac{\partial l}{\partial t}(t, z) \equiv 0$$

Donc, nécessairement, l est uniquement fonction de z . Ainsi

$$\frac{\partial l}{\partial z}(z)\exp(-tA)B \equiv 0$$

En dérivant cette relation par rapport à t , on a,

$$\frac{\partial l}{\partial z}(z)\exp(-tA)AB \equiv 0$$

Plus généralement, une dérivation à n'importe quel ordre $k \geq 0$ donne

$$\frac{\partial l}{\partial z}(z)\exp(-tA)A^kB \equiv 0$$

En spécifiant $t = 0$, on obtient

$$\frac{\partial l}{\partial z}(z)A^kB = 0, \quad \forall k \geq 0.$$

Ainsi le vecteur $\frac{\partial l}{\partial z}(z)$ appartient à l'intersection des noyaux à gauche de la famille infinie de matrice $(A^k B)_{k \geq 0}$. Le noyau à gauche de $A^k B$ n'est autre que $\text{Im}(A^k B)^\perp$, l'orthogonal de l'image de $A^k B$. Donc

$$\frac{\partial l}{\partial z}(z) \in \bigcap_{k \geq 0} \text{Im}(A^k B)^\perp$$

Mais

$$\bigcap_{k \geq 0} \text{Im}(A^k B)^\perp = (\text{Im}(B) + \dots + \text{Im}(A^k B) + \dots)^\perp$$

La suite d'espaces vectoriels $E_k = \text{Im}(B) + \dots + \text{Im}(A^k B)$ est une suite croissante pour l'inclusion, $E_k \subset E_{k+1}$. Si pour un certain k , $E_k = E_{k+1}$, cela signifie que $\text{Im}(A^{k+1} B) \subset E_k$, donc $A(E^k) \subset E_k$. Mais $\text{Im}(A^{k+2} B) = \text{Im}(AA^{k+1} B) \subset A(E^{k+1})$. Ainsi $\text{Im}(A^{k+2} B) \subset E_k$. On voit donc que pour tout $r > 0$, $\text{Im}(A^{k+r} B) \subset E_k$, d'où $E_{k+r} = E_k$. La suite des E_k est une suite de sous-espaces vectoriels de \mathbb{R}^n emboîtés les uns dans les autres. Cette suite stationne dès qu'elle n'est plus, pour un certain k , strictement croissante. Il suffit donc de ne considérer que ses n premiers termes, c.-à-d. E_0, \dots, E_{n-1} , car automatiquement $E_{n-1} = E_{n+r}$ pour tout $r > 0$.

En revenant à la suite des noyaux à gauche de $A^k B$, nous voyons que le fait que $\frac{\partial l}{\partial z}(z)$ est dans le noyau à gauche de la suite infinie de matrices $(A^k B)_{k \geq 0}$, est équivalent au fait que $\frac{\partial l}{\partial z}(z)$ est dans le noyau à gauche de la *suite finie* de matrices $(A^k B)_{0 \leq k \leq n-1}$.⁵

Ainsi, pour tout z , $\frac{\partial l}{\partial z}(z)$ appartient au noyau à gauche de la matrice $n \times (nm)$,

$$\mathcal{C} = (B, AB, A^2 B, \dots, A^{n-1} B) \quad (3.10)$$

dite *matrice de commandabilité* de Kalman. Si \mathcal{C} est de rang n , son noyau à gauche est nul, donc l ne dépend pas de z : l est alors une fonction constante et h également.

Réiproquement, si la matrice de commandabilité \mathcal{C} n'est pas de rang maximal, alors il existe un vecteur $w \in \mathbb{R}^n \setminus \{0\}$, dans le noyau à gauche de (3.10). En remontant les calculs avec $l(z, t) = w^T z$, on voit que $\frac{d}{dt} l = 0$ le long des trajectoires. En passant aux variables (x, u) , on obtient une *intégrale première non triviale* $h(t, x) = w^T \exp(-tA)x$. Toute trajectoire du système se situe dans un hyperplan orthogonal à w .

En résumé, nous avons démontré la proposition suivante.

Proposition 8. *La matrice de commandabilité*

$$\mathcal{C} = (B, AB, A^2 B, \dots, A^{n-1} B)$$

est de rang n , si, et seulement si, les seules intégrales premières du système (3.8) sont triviales.

Des Propositions 7 et 8, il vient que, si le système (3.8) est commandable, sa matrice de commandabilité est de rang n . Nous allons voir que la réciproque est vraie. Pour cela, nous avons besoin de certaines propriétés d'invariance.

5. On pourrait aussi utiliser le théorème de Cayley-Hamilton qui donne un résultat plus précis : toute matrice carrée est racine de son polynôme caractéristique. Cela veut dire, A étant de taille n , que A^n est une combinaison linéaire des $(A^k)_{0 \leq k \leq n-1}$

$$A^n = \sum_{k=0}^{n-1} p_k A^k$$

où les p_k sont définis par $\det(\lambda I_n - A) = \lambda^n - \sum_{k=0}^{n-1} p_k \lambda^k$. Nous avons préféré un argument plus simple avec la suite des E_k car il a l'avantage de s'appliquer au cas non linéaire. Il correspond au calcul de l'algèbre de Lie de commandabilité.

3.3.2 Invariance

Définition 14 (Changement d'état, bouclage statique régulier). Un changement linéaire de coordonnées $x \mapsto \tilde{x}$ est défini par une matrice M inversible d'ordre n : $x = M\tilde{x}$. Un bouclage statique régulier $u \mapsto \tilde{u}$ est défini par une matrice N inversible d'ordre m et une autre matrice K , $m \times n$: $u = K\tilde{x} + N\tilde{u}$. C'est un changement de variables sur les commandes paramétré par l'état.

L'ensemble des transformations

$$\begin{pmatrix} \tilde{x} \\ \tilde{u} \end{pmatrix} \mapsto \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} M & 0 \\ K & N \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{u} \end{pmatrix} \quad (3.11)$$

forment un groupe lorsque les matrices M , N et K varient (M et N restant inversibles).

Si $\frac{d}{dt}x = Ax + Bu$ est commandable (resp. n'admet pas d'intégrale première) alors $\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\tilde{u}$ obtenu avec (3.11) est commandable (resp. n'admet pas d'intégrale première). Les notions de commandabilité et d'intégrale première sont intrinsèques, c.-à-d. indépendantes des coordonnées avec lesquelles les équations du système sont établies. Si la matrice de commandabilité dans les coordonnées (x, u) est de rang n , la matrice de commandabilité dans les coordonnées (\tilde{x}, \tilde{u}) sera aussi de rang n . Cette simple remarque conduit au résultat non évident suivant

$$\text{rang}(B, AB, \dots, A^{n-1}B) = n \quad \text{équivaut à} \quad \text{rang}(\tilde{B}, \tilde{A}\tilde{B}, \dots, \tilde{A}^{n-1}\tilde{B}) = n$$

où \tilde{A} et \tilde{B} s'obtiennent en écrivant $\frac{d}{dt}x = Ax + Bu$ dans les coordonnées (\tilde{x}, \tilde{u})

$$\dot{\tilde{x}} = M^{-1}(AM + BK)\tilde{x} + M^{-1}BN\tilde{u}.$$

soit

$$\tilde{A} = M^{-1}(AM + BK), \quad \tilde{B} = M^{-1}BN$$

En fait, il est possible d'aller beaucoup plus loin et de montrer que les indices de commandabilité définis ci-dessous sont aussi invariants.

Définition 15 (Indices de commandabilité). Pour tout entier k , on note σ_k le rang de la matrice $(B, AB, A^2B, \dots, A^kB)$. Les (σ_k) sont appelés indices de commandabilité du système linéaire (3.8),

La suite σ_k est croissante, majorée par n . L'absence d'intégrale première est équivalente à $\sigma_{n-1} = n$.

Proposition 9 (invariance). *Les indices de commandabilité de $\frac{d}{dt}x = Ax + Bu$ sont invariants par changement de variable sur x et bouclage statique régulier sur u .*

Nous laissons la preuve de ce résultat par récurrence sur n en exercice. Il est important de comprendre l'idée géométrique de cette invariance. Les transformations $(x, u) \mapsto (\tilde{x}, \tilde{u})$ du type (3.11) forment un groupe. Ce groupe définit une relation d'équivalence entre deux systèmes ayant même nombre d'états et même nombre de commandes. La proposition précédente signifie simplement que les indices de commandabilité sont les mêmes pour deux systèmes appartenant à la même classe d'équivalence, i.e le même objet géométrique vu dans deux repères différents. En fait, on peut montrer que les indices de commandabilité sont les seuls invariants : il y a autant de classes d'équivalence que d'indices de commandabilité possibles. Nous ne montrerons pas en détail ce résultat. Tous les éléments nécessaires à cette preuve se trouvent dans la construction de la forme de Brunovsky ci-dessous (voir aussi [39, 71]).

Nous allons voir, avec la forme normale de Brunovsky, qu'une telle correspondance entre y et les trajectoires du système est générale. Il suffit que (3.8) soit commandable. Tout revient donc à trouver la sortie de Brunovsky y de même dimension que la commande u .

3.3.3 Critère de Kalman et forme de Brunovsky

Théorème 22 (Critère de commandabilité de Kalman)

Le système $\frac{d}{dt}x = Ax + Bu$ est commandable si et seulement si la *matrice de commandabilité*

$$\mathcal{C} = (B, AB, \dots, A^{n-1}B)$$

est de rang $n = \dim(x)$.

Pour abréger, on dit souvent que *la paire* (A, B) est *commandable* pour dire que le rang de la matrice de commandabilité \mathcal{C} est maximum.

La preuve que nous allons donner de ce résultat n'est pas la plus courte possible. Cependant, elle permet de décrire explicitement, pour toute durée $T > 0$ et pour $p, q \in \mathbb{R}^n$, les trajectoires du système qui partent de p et arrivent en q . Cette preuve utilise la *forme* dite de *Brunovsky*, objet du théorème suivant. La forme de Brunovsky se construit en suivant une méthode d'élimination proche du pivot de Gauss.

Théorème 23 (Forme de Brunovsky)

Si $(B, AB, \dots, A^{n-1}B)$, la *matrice de commandabilité* de $\frac{d}{dt}x = Ax + Bu$, est de rang $n = \dim(x)$ et si B est de rang $m = \dim(u)$, alors il existe un *changement d'état* $z = Mx$ (M matrice inversible $n \times n$) et un *bouclage statique régulier* $u = Kz + Nv$ (N matrice inversible $m \times m$), tels que les équations du système dans les variables (z, v) admettent la *forme normale* suivante (écriture sous la forme de m équations différentielles d'ordre ≥ 1)

$$y_1^{(\alpha_1)} = v_1, \quad \dots, \quad y_m^{(\alpha_m)} = v_m \quad (3.12)$$

avec comme état $z = (y_1, y_1^{(1)}, \dots, y_1^{(\alpha_1-1)}, \dots, y_m, y_m^{(1)}, \dots, y_m^{(\alpha_m-1)})$, les α_i étant des entiers positifs.

Les m quantités y_i , qui sont des combinaisons linéaires de l'état x , sont appelées *sorties de Brunovsky*.

Les *indices de commandabilité* σ_k et les m entiers α_i de la forme de Brunovsky sont intimement liés (pour plus de détail voir [39]) : la connaissance des uns est équivalente à celle des autres.

Démonstration. La preuve de ce résultat repose sur

- une mise sous forme triangulaire des équations d'état et l'élimination de u ;
- l'invariance du rang de $(B, AB, \dots, A^{n-1}B)$ par rapport aux transformations (3.11) ;
- une récurrence sur la dimension de l'état.

Mise sous forme triangulaire

On suppose que B est de rang $m = \dim(u)$ (sinon, on peut faire un regroupement des commandes en un nombre plus petit que m de façon à se ramener à ce cas). Alors, il existe une partition de l'état $x = (x_r, x_u)$ avec $\dim(x_r) = n - m$ et $\dim(x_u) = m$ telle que les équations (3.8) admettent la

structure bloc suivante

$$\begin{aligned}\frac{d}{dt}x_r &= A_{rr}x_r + A_{ru}x_u + B_ru \\ \frac{d}{dt}x_u &= A_{ur}x_r + A_{uu}x_u + B_uu\end{aligned}$$

où B_u est une matrice carrée inversible. Cette partition n'est pas unique, bien sûr. En résolvant la seconde équation par rapport à u et en reportant cette expression dans la première équation, on obtient

$$\begin{aligned}\frac{d}{dt}x_r &= A_{rr}x_r + A_{ru}x_u + B_rB_u^{-1}(\frac{d}{dt}x_u - A_{ur}x_r - A_{uu}x_u) \\ \frac{d}{dt}x_u &= A_{ur}x_r + A_{uu}x_u + B_uu\end{aligned}$$

En regroupant les dérivées dans la première équation, on a

$$\begin{aligned}\frac{d}{dt}x_r - B_rB_u^{-1}\frac{d}{dt}x_u &= (A_{rr} - B_rB_u^{-1}A_{ur})x_r + (A_{ru} - B_rB_u^{-1}A_{uu})x_u \\ \frac{d}{dt}x_u &= A_{ur}x_r + A_{uu}x_u + B_uu.\end{aligned}$$

Avec la transformation du type (3.11) définie par

$$\tilde{x}_r = x_r - B_rB_u^{-1}x_u, \quad \tilde{x}_u = x_u, \quad \tilde{u} = A_{ur}x_r + A_{uu}x_u + B_uu,$$

les équations $\frac{d}{dt}x = Ax + Bu$ deviennent

$$\begin{aligned}\frac{d}{dt}\tilde{x}_r &= \tilde{A}_r\tilde{x}_r + \tilde{A}_u\tilde{x}_u \\ \frac{d}{dt}\tilde{x}_u &= \tilde{u}\end{aligned}$$

où $\tilde{A}_r = (A_{rr} - B_rB_u^{-1}A_{ur})$ et $\tilde{A}_u = (A_{rr} - B_rB_u^{-1}A_{ur})B_rB_u^{-1} + (A_{ru} - B_rB_u^{-1}A_{uu})$. Dans cette structure triangulaire, où la commande \tilde{u} n'intervient pas dans la seconde équation, on voit apparaître un système plus petit d'état \tilde{x}_r et de commande \tilde{x}_u . Cela nous permet de réduire la dimension de x et de raisonner par récurrence.

Invariance de rang

Un simple calcul par blocs nous montre que si $(B, AB, \dots, A^{n-1}B)$ est de rang n alors $(A_u, A_rA_u, \dots, A_r^{n-m-1}A_u)$ est de rang $n - m$. Du système de taille n on passe ainsi au système de taille réduite $n - m$, $\frac{d}{dt}\tilde{x}_r = \tilde{A}_r\tilde{x}_r + \tilde{A}_u\tilde{x}_u$ (\tilde{x}_r l'état, \tilde{x}_u la commande).

Récurrence sur le nombre d'états

Étant donné n , supposons le résultat vrai pour toutes les dimensions d'état inférieures ou égales à $n - 1$. Considérons un système $\frac{d}{dt}x = Ax + Bu$ avec $n = \dim(x)$, sa matrice de commandabilité de rang n , et B de rang $m = \dim(u) > 0$. L'élimination de u donne, après une transformation du type (3.11),

$$\begin{aligned}\frac{d}{dt}x_r &= A_rx_r + A_ux_u \\ \frac{d}{dt}x_u &= u\end{aligned}$$

où $\dim(u) = \dim(x_u) = m$ et $\dim(x_r) = n - m$ avec $(A_u, A_rA_u, \dots, A_r^{n-m-1}A_u)$ de rang $n - m < n$ (les \sim ont été enlevés pour alléger les notations). Notons \bar{m} le rang de A_u . Comme $\bar{m} \leq m$, un changement de variable sur x_u , $(\bar{x}_u, \tilde{x}_u) = Px_u$ avec P inversible, permet d'écrire le système sous la forme

$$\begin{aligned}\frac{d}{dt}x_r &= A_rx_r + \bar{A}_u\bar{x}_u \\ \frac{d}{dt}\bar{x}_u &= \bar{u} \\ \frac{d}{dt}\tilde{x}_u &= \tilde{u}\end{aligned} \tag{3.13}$$

avec $(\bar{u}, \tilde{u}) = Pu$, $\dim(\bar{x}_u) = \bar{m}$ et \bar{A}_u de rang \bar{m} . Le rang de la matrice de commandabilité de $\frac{d}{dt}x_r = A_rx_r + \bar{A}_u\bar{x}_u$ (x_r est l'état et \bar{x}_u la commande) est égal à $n - m = \dim(x_r)$, l'hypothèse

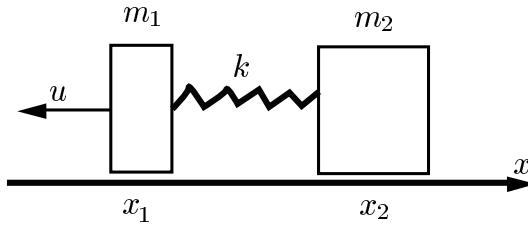


FIGURE 3.6 – Deux masses couplées par un ressort. L’ensemble est soumis à une seule force u .

de récurrence assure l’existence d’un changement de variable $x_r = Mz$ et d’un bouclage statique régulier $\bar{x}_u = Kz + N\bar{v}$ (\bar{v} est la nouvelle commande ici) mettant ce sous-système sous forme de Brunovsky. Le changement d’état $(x_r, \bar{x}_u, \tilde{x}_u)$ défini par

$$\begin{pmatrix} x_r \\ \bar{x}_u \\ \tilde{x}_u \end{pmatrix} = \begin{pmatrix} M & 0 & 0 \\ K & N & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z \\ \bar{v} \\ \tilde{x}_u \end{pmatrix}$$

et le bouclage statique régulier sur (\bar{u}, \tilde{u})

$$\bar{u} = KM^{-1}(A_r x_r + \bar{A}_u \bar{x}_u) + N\bar{v}, \quad \tilde{u} = \tilde{v}$$

transforment le système (3.13) sous forme de Brunovsky avec $v = (\bar{v}, \tilde{v})$ comme nouvelle commande. \square

Preuve du Théorème 22 La commandabilité est indépendante du choix des variables sur x et d’un bouclage statique régulier sur u . On peut donc supposer le système sous sa forme de Brunovsky. Dans ces coordonnées, aller d’un état à un autre est élémentaire. Ce problème se ramène à étudier la commandabilité du système scalaire $y^{(\alpha)} = v$. L’état initial $(y_0, \dots, y_0^{\alpha-1})$, l’état final $(y_T, \dots, y_T^{\alpha-1})$ ainsi que la durée T étant donnés, les lois horaires $t \mapsto v(t)$ assurant le passage entre ces deux états pendant la durée T correspondent alors à la dérivée α -ième de fonctions $[0, T] \ni t \mapsto \varphi(t) \in \mathbb{R}$, dont les dérivées jusqu’à l’ordre $\alpha - 1$ en 0 et T sont imposées par

$$\varphi^{(r)}(0) = y_0^r, \quad \varphi^{(r)}(T) = y_T^r, \quad r = 0, \dots, \alpha - 1$$

Il existe bien sûr une infinité de telles fonctions φ (on peut prendre pour φ un polynôme de degré $2\alpha - 1$, par exemple).

Exemple 17 (Un exemple d’utilisation de la forme de Brunovsky). Soit le système mécanique à deux degrés de liberté et une seule commande représenté sur la Figure 3.6. En négligeant les frottements et en supposant le ressort linéaire de raideur k , on obtient le modèle suivant

$$\begin{cases} m_1 \ddot{x}_1 = k(x_2 - x_1) + u \\ m_2 \ddot{x}_2 = k(x_1 - x_2) \end{cases} \quad (3.14)$$

Montrons que ce système est commandable. Il suffit pour cela de remarquer que la quantité x_2 , l’abscisse de la masse qui n’est pas directement soumise à la force u , joue un rôle très particulier (sortie de Brunovsky). Si au lieu de donner $t \mapsto u(t)$ et d’intégrer (3.14) à partir de positions et vitesses initiales, on fixe $t \mapsto x_2(t) = y(t)$, alors, on peut calculer $x_1 = \frac{m_2}{k} \dot{y} + y$ et donc, $u =$

$m_1\ddot{x}_1 + m_2\ddot{x}_2 = \frac{m_1m_2}{k}y^{(4)} + (m_1 + m_2)\ddot{y}$. Ainsi, on peut écrire le système en faisant jouer à x_2 un rôle privilégié

$$\begin{cases} x_1 = (m_2/k)\ddot{y} + y \\ x_2 = y \\ u = (m_1m_2/k)y^{(4)} + (m_1 + m_2)\ddot{y} \end{cases}$$

On obtient une paramétrisation explicite de toutes les trajectoires du système. Les relations précédentes établissent une correspondance bi-univoque et régulière entre les trajectoires de (3.14) et les fonctions régulières $t \mapsto y(t)$. Cette correspondance permet de calculer de façon élémentaire une commande $[0, T] \ni t \mapsto u(t)$ qui fait passer le système de l'état $p = (x_1^p, v_1^p, x_2^p, v_2^p)$ à l'état $q = (x_1^q, v_1^q, x_2^q, v_2^q)$ (v_i correspond à $\frac{d}{dt}x_i$). Étant données les équations

$$\begin{cases} x_1 = (m_2/k)\ddot{y} + y \\ v_1 = (m_2/k)y^{(3)} + \frac{d}{dt}y \\ x_2 = y \\ v_2 = \frac{d}{dt}y \end{cases} \quad (3.15)$$

il apparaît qu'imposer p en $t = 0$ revient à imposer y et ses dérivées jusqu'à l'ordre 3 en 0 : $(y_0, y_0^{(1)}, y_0^{(2)}, y_0^{(3)})$. Il en est de même en $t = T$ avec $(y_T, y_T^{(1)}, y_T^{(2)}, y_T^{(3)})$. Il suffit donc de trouver une fonction régulière $[0, T] \ni t \mapsto y(t)$ dont les dérivées jusqu'à l'ordre 3 sont données a priori en 0 et en T . Un polynôme de degré 7 en temps répond à la question mais il existe bien d'autres possibilités. Nous en détaillons une ci-dessous qui est complètement explicite.

Soit la fonction $C^4 \phi : [0, 1] \mapsto [0, 1]$ définie par

$$\phi(\sigma) = \frac{\sigma^4}{\sigma^4 + (1 - \sigma)^4}.$$

Il est facile de voir que ϕ est croissante avec $\phi(0) = 0$, $\phi(1) = 1$ et $\frac{d^k\phi}{d\sigma^k}(0) = \frac{d^k\phi}{d\sigma^k}(1) = 0$ pour $k = 1, 2, 3$. La fonction

$$\begin{aligned} y(t) &= \left(1 - \phi\left(\frac{t}{T}\right)\right) \left(y_0 + ty_0^{(1)} + \frac{t^2}{2}y_0^{(2)} + \frac{t^3}{6}y_0^{(3)}\right) \\ &\quad + \phi\left(\frac{t}{T}\right) \left(y_T + (t - T)y_T^{(1)} + \frac{(t - T)^2}{2}y_T^{(2)} + \frac{(t - T)^3}{6}y_T^{(3)}\right) \end{aligned}$$

vérifie $y(0) = y_0$, $y(T) = y_T$ avec $\frac{d^k y}{dt^k}(0, T) = y_{0,T}^{(k)}$, $k = 1, 2, 3$. On obtient ainsi une trajectoire en x correspondant à une telle transition en utilisant les formules (3.15) et le contrôle correspondant avec

$$u = (m_1m_2/k)y^{(4)} + (m_1 + m_2)\ddot{y}$$

En particulier, le transfert de la configuration stationnaire $x_1 = x_2 = 0$ à la configuration stationnaire $x_1 = x_2 = D > 0$ durant le temps T s'obtient avec le contrôle en boucle ouverte suivant (feedforward)

$$u(t) = \frac{Dm_1m_2}{kT^4} \frac{d^4\phi}{d\sigma^4}\Big|_{\frac{t}{T}} + \frac{D(m_1 + m_2)}{T^2} \frac{d^2\phi}{d\sigma^2}\Big|_{\frac{t}{T}}$$

car alors $y(t) = D\phi\left(\frac{t}{T}\right)$.

3.3.4 Planification et suivi de trajectoires

Des preuves des Théorèmes 22 et 23, il est important de retenir deux choses.

1. Dire que le système $\frac{d}{dt}x = Ax + Bu$ est *commandable*, est équivalent à l'existence d'un bouclage statique régulier $u = Kz + Nv$ et d'un changement d'état $x = Mz$ permettant de ramener le système à la *forme de Brunovsky* $y^{(\alpha)} = v$ et $z = (y, \dots, y^{(\alpha-1)})$ (par abus de notation $y = (y_1, \dots, y_m)$ et $y^{(\alpha)} = (y_1^{(\alpha_1)}, \dots, y_m^{(\alpha_m)})$). Ces changements donnent

$$x = M(y, \dots, y^{(\alpha-1)}), \quad u = L(y, \dots, y^{(\alpha)})$$

où la matrice L est construite avec K , N et M . Lorsqu'on considère une fonction régulière arbitraire du temps $t \mapsto \varphi(t) \in \mathbb{R}^m$ et qu'on calcule $x(t)$ et $u(t)$ par les relations

$$x(t) = M(\varphi(t), \dots, \varphi^{(\alpha-1)}(t)), \quad u(t) = L(\varphi(t), \dots, \varphi^{(\alpha)}(t))$$

alors $t \mapsto (x(t), u(t))$ est une *trajectoire* du système. L'équation $\frac{d}{dt}x(t) - Ax(t) - Bu(t) = 0$ est identiquement satisfaite. Réciproquement, toutes les trajectoires régulières du système se paramétrisent de cette façon, grâce à m fonctions scalaires arbitraires $\varphi_1(t), \dots, \varphi_m(t)$ et un nombre fini de leurs dérivées par les formules ci-dessus.

2. La commandabilité de $\frac{d}{dt}x = Ax + Bu$ implique la possibilité de *stabilisation* par retour d'état. La forme de Brunovsky fait apparaître cette possibilité. Sous cette forme normale, chacun des m sous-systèmes indépendants s'écrit $y_i^{(\alpha_i)} = v_i$. Soient α_i valeurs propres, $\lambda_1, \dots, \lambda_{\alpha_i}$, correspondant au spectre d'une matrice réelle de dimension α_i . Notons s_k les fonctions symétriques des λ_i (des quantités réelles donc) homogènes de degré k , obtenues par le développement suivant

$$\prod_{k=1}^{\alpha_i} (X - \lambda_k) = X^{\alpha_i} - s_1 X^{\alpha_i-1} + s_2 X^{\alpha_i-2} + \dots + (-1)^{\alpha_i} s_{\alpha_i}$$

Alors, dès que les λ_k sont à partie réelle strictement négative (c.-à-d. qu'ils correspondent à une *matrice Hurwitz*), le bouclage

$$v_i = s_1 y_i^{(\alpha_i-1)} - s_2 y_i^{(\alpha_i-2)} + \dots + (-1)^{\alpha_i-1} s_{\alpha_i} y_i$$

assure la stabilité de $y_i^{(\alpha_i)} = v_i$. En effet, les exposants caractéristiques (on dit aussi les *pôles*) du système bouclé sont les λ_k .

Comme nous venons de le voir, de la forme de Brunovsky, on déduit directement le résultat suivant

Théorème 24 (Placement de pôles)

Si la paire (A, B) est commandable (voir le Théorème 22) alors, pour toute matrice réelle F de taille $n \times n$, il existe une matrice $m \times n$, K (non nécessairement unique si $m > 1$), telle que le spectre de $A - BK$ coïncide avec celui de F .

De retour dans les coordonnées de modélisation, $\frac{d}{dt}x = Ax + Bu$, la *planification de trajectoire* nous donne une *trajectoire en boucle ouverte* du système (par exemple la trajectoire que doit suivre une fusée au décollage, la manœuvre d'atterrissement d'un avion, ...). Nous la notons $t \mapsto$

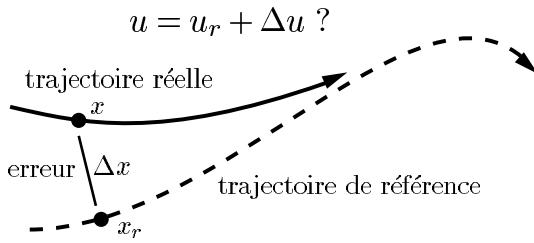


FIGURE 3.7 – Suivi de trajectoire.

$(x_r(t), u_r(t))$ avec l’indice r pour référence. En pratique, il convient, comme l’illustre la Figure 3.7, de corriger en boucle fermée en fonction de l’écart Δx , la commande de référence u_r . Le problème est donc de calculer la correction Δu à partir de Δx de façon à revenir sur la trajectoire de référence. C’est un problème de *suivi de trajectoire*. On peut également le résoudre en utilisant un bouclage stabilisant par placement de pôles sous la forme de Brunovsky. Nous allons détailler ce point.

Comme $\frac{d}{dt}x_r = Ax_r + Bu_r$, on obtient, par différence avec $\frac{d}{dt}x = Ax + Bu$ l’équation d’erreur suivante

$$\frac{d}{dt}(\Delta x) = A \Delta x + B \Delta u$$

où $\Delta x = x - x_r$ et $\Delta u = u - u_r$. Le système étant commandable, il existe K , matrice $m \times n$, telle que les valeurs propres de $A + BK$ sont toutes à partie réelle strictement négative (d’après le Théorème de placement de pôles 24). La correction

$$\Delta u = -K \Delta x$$

assure le *suivi asymptotique* de la *trajectoire de référence* $t \mapsto x_r(t)$.

Nous terminerons par une constatation d’ordre expérimental : lorsque le modèle dynamique $\frac{d}{dt}x = Ax + Bu$ est d’origine physique, il n’est pas rare que sa partie non commandable, i.e., ses intégrales premières, ait une signification physique immédiate, tout comme les grandeurs y , fonction de x et intervenant dans la *forme de Brunovsky* de sa partie commandable. Cet état de fait n’est vraisemblablement pas dû entièrement au hasard : en physique, les grandeurs qui admettent une signification intrinsèque, i.e., les grandeurs physiques, sont celles qui ne dépendent pas du repère de l’observateur. En Automatique, le passage d’un repère à un autre correspond, entre autres, à une transformation de type (3.11). Il est alors clair que le “sous-espace” engendré par les sorties de Brunovsky est un invariant. Il a donc toutes les chances d’avoir un sens physique immédiat. Les sorties de Brunovsky admettent un équivalent non linéaire pour de nombreux systèmes physiques (voir la Section 3.5.1).

3.4 Commande linéaire quadratique LQR

Nous avons étudié la propriété de *commandabilité* (objet de la Définition 12), qui requiert, pour être satisfaite par un système, l’existence pour tout p, q d’une trajectoire reliant un état p à un état q . Sous l’hypothèse de commandabilité, nous avons vu comment construire une telle trajectoire par la *forme de Brunovsky* (voir Théorème 23). Cette trajectoire est loin d’être unique. On va maintenant montrer comment choisir la trajectoire la meilleure au sens d’un critère quadratique. C’est l’objet de la *commande linéaire quadratique*.

Nous nous intéressons ici, en toute généralité, aux *systèmes linéaires instationnaires* du type

$$\frac{d}{dt}x(t) = A(t)x(t) + B(t)u \quad (3.16)$$

où l'état $x(t) \in \mathbb{R}^n$, la commande $u(t) \in \mathbb{R}^m$ et, pour tout $t \in [0, +\infty[$ les matrices $A(t)$ et $B(t)$ sont de tailles $n \times n$ et $n \times m$, respectivement. Ces systèmes peuvent provenir de bilans exacts ou, plus souvent, de la linéarisation autour de trajectoires de manœuvre entre des points stationnaires. Ainsi si $t \mapsto (x_r(t), u_r(t))$ est une trajectoire de référence de $\frac{d}{dt}x = f(x, u)$, on a $\frac{d}{dt}x_r = f(x_r, u_r)$ et on note

$$A(t) = \left. \frac{\partial f}{\partial x} \right|_{(x_r(t), u_r(t))}, \quad B(t) = \left. \frac{\partial f}{\partial u} \right|_{(x_r(t), u_r(t))}$$

En fait, x et u dans (3.16), correspondent aux écarts $x - x_r(t)$ et $u - u_r(t)$.

Considérons un intervalle de temps donné $[0, t_f]$ (sur lequel l'éventuelle trajectoire de référence est définie). On cherche à minimiser le critère quadratique suivant

$$J = \int_0^{t_f} (x^T(t)Rx + u^T(t)Qu(t)) dt \quad (3.17)$$

où R est une matrice symétrique positive et Q une matrice symétrique définie positive⁶.

Implicitement, on cherche une commande $[0, t_f] \ni t \mapsto u(t)$ "faible" (car pondérée par Q dans le calcul de l'intégrale) limitant les déviations de x elles-mêmes pondérées par R sur l'horizon temporel $[0, t_f]$. La commande minimisant (3.17) réalise le meilleur compromis (au sens précisé par les pondérations R et Q) entre déviation de l'état et effort sur les actionneurs.

Exemple 18 (Réentrée atmosphérique). *Un exemple de trajectoire définissant un système linéarisé tangent instationnaire du type (3.16) est donné sur la Figure 3.8. On y a représenté l'historique des valeurs de l'altitude pour une trajectoire de vaisseau spatial (type navette) dans la manœuvre de réentrée atmosphérique au cours de laquelle on cherche à maximiser le déport latéral (i.e. en faisant pivoter l'appareil pour rejoindre une plus haute latitude où est situé le point de chute désiré pour l'engin). Cet objectif et les contraintes aérodynamiques (notamment la thermique et la variation de densité de l'atmosphère) confèrent à l'optimum de nombreuses bosses (appelées rebonds atmosphériques). Afin d'assurer le suivi de cette trajectoire, on doit concevoir un contrôleur en boucle fermée. On linéarise alors les équations autour de cette trajectoire à rebonds et on obtient naturellement un modèle de la forme (3.16). On pourra se reporter à [11] pour une présentation détaillée de ce problème.*

Dans le cas multivariable en particulier, i.e. où on dispose de plusieurs commandes, on a en général "trop" de degrés de liberté pour résoudre les problèmes de planification et de suivi de trajectoire. En évoquant un critère à minimiser comme nous venons de le faire en (3.17), on définit implicitement l'utilisation des degrés de liberté supplémentaires : ils permettent d'optimiser cet indice de performance.

Exemple 19 (Réacteur nucléaire à eau pressurisée). *On présente ici un exemple de système multivariable pour lequel on peut réaliser des transitoires de très nombreuses façons. De manière générale, les réacteurs nucléaires de production d'électricité présentent d'intéressants problèmes d'Automatique. Leur schéma de principe est donné sur la Figure 3.9, on pourra se référer à [50] pour une présentation générale. Dans le circuit primaire, la neutronique est contrôlée par la contre-réactivité des barres et du bore dont on peut agir sur la concentration par un circuit de borication/dilution. Les échanges avec le circuit secondaire fournissent, à travers le générateur de vapeur, de la puissance*

6. L'hypothèse de symétrie n'enlève rien à la généralité. En effet pour tout $x \in \mathbb{R}^n$, on a $x^T G x = x^T ((G + G^T)/2)x$. Autrement dit, seule la partie symétrique de G compte dans le calcul de $x^T G x$.

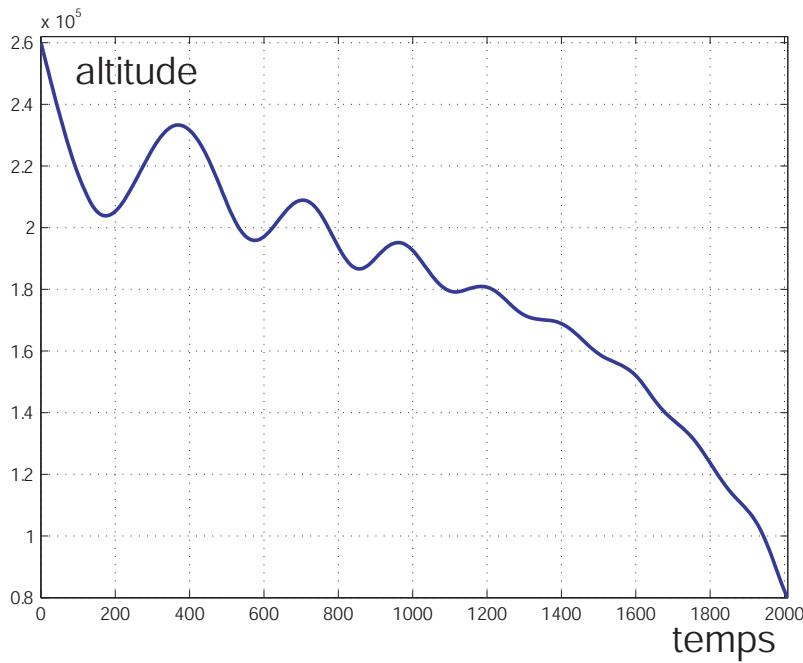


FIGURE 3.8 – Trajectoire de réentrée d'un véhicule spatial (départ en orbite, arrivée à 80.000 pieds). Les *rebonds atmosphériques* permettent de maximiser le déport latéral en utilisant les effets aérodynamiques.

à la turbine qui est accouplée à une génératrice. On peut utiliser une vanne de contournement pour fournir plus ou moins de puissance à la turbine. En dehors des phases de régulation, c.-à-d. de la stabilisation autour d'un point de fonctionnement (on parle de régulation de fréquence du réseau), on est souvent amené à réaliser des transitoires lorsque la demande du réseau électrique varie. Une manœuvre difficile est l'ilôtage, qui consiste à isoler le réseau. Il faut alors utiliser les différentes commandes de manière coordonnée pour réduire la puissance en minimisant la fatigue des installations. Une formulation type commande linéaire quadratique permet de spécifier ce genre de cahier des charges.

De manière encore plus générale, on peut considérer le coût avec pondération finale

$$J = \frac{1}{2}x^T(t_f)S_fx(t_f) + \frac{1}{2} \int_0^{t_f} (x^T(t)Rx(t) + u^T(t)Qu(t)) dt \quad (3.18)$$

qui permet de rajouter de l'importance au point final atteint (S_f étant une matrice symétrique positive).

L'intérêt des formulations quadratiques (3.18) (et (3.17)) est qu'elles sont en général bien posées au sens mathématique, et qu'elles admettent une solution unique, calculable numériquement⁷

3.4.1 Multiplicateurs de Lagrange en dimension infinie

Le but de cette section est de présenter de façon simple la méthode des *multiplicateurs de Lagrange* en dimension infinie. Cette méthode permet de caractériser l'optimum du critère (3.18) (ou (3.17)).

7. Calculable analytiquement pour un ou deux états mais pas plus en pratique. C'est très différent de l'approche utilisant la forme normale de Brunovsky, approche que l'on peut conduire analytiquement sans difficulté sur de nombreux exemples physiques à bien plus que trois états.

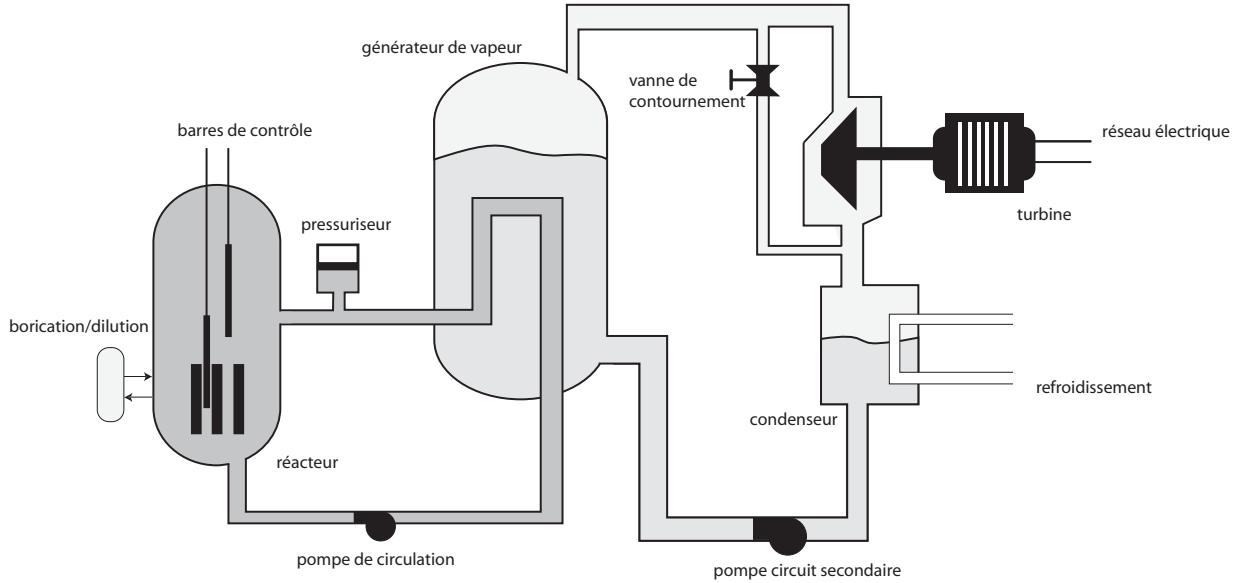


FIGURE 3.9 – Schéma de principe d'un réacteur nucléaire à eau pressurisée. Pour contrôler la production d'électricité, on peut utiliser 3 actionneurs (commandes) : barres de contrôle, système de borication/dilution, vanne de contournement de la turbine.

Un lecteur intéressé pourra consulter [4, 13] ou les notes de l'ES d'optimisation [63].

Commençons par la dimension finie. On part du problème suivant où le critère $J : \mathbb{R}^n \mapsto \mathbb{R}$ et les p contraintes scalaires $h_i : \mathbb{R}^n \mapsto \mathbb{R}$ sont des fonctions dérивables du temps

$$\begin{cases} \min_{x \in \mathbb{R}^n} & J(x) \\ h_1(x) = 0 \\ \vdots \\ h_p(x) = 0 \end{cases}$$

Pour résoudre ce problème on introduit

- p multiplicateurs de Lagrange (les variables *adiointes*) $\lambda = (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$;
- on calcule le *Lagrangien*

$$\mathcal{L}(x, \lambda) = J(x) + \sum_{i=1}^p \lambda_i h_i(x) = J(x) + \lambda^T h(x);$$

- on calcule les conditions de stationnarité du Lagrangien au 1er ordre en faisant comme si les variables x et λ pouvaient varier librement dans toutes les directions. Ces conditions traduisent le fait que $\delta\mathcal{L} = 0$ pour toutes variations infinitésimales δx et $\delta\lambda$ des variables x et λ .

La condition $\delta\mathcal{L} = 0$ pour tout δx donne n équations

$$\frac{\partial J}{\partial x_k} + \sum_{i=1}^p \lambda_i \frac{\partial h_i}{\partial x_k} = 0, \quad k = 1, \dots, n$$

La condition $\delta\mathcal{L} = 0$ pour tout $\delta\lambda$ redonne les p contraintes

$$h_i(x) = 0, \quad i = 1, \dots, p$$

On a ainsi un système de $n+p$ inconnues avec $n+p$ équations à résoudre pour trouver l'optimum. La partie la plus dure reste cependant à faire : résoudre ce système d'équations non-linéaires et vérifier que la solution trouvée est bien l'optimum⁸.

Passons maintenant à la dimension infinie avec le problème suivant où toutes les fonctions $L : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, $f : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$, $l : \mathbb{R}^n \mapsto \mathbb{R}$ sont dérivables et où le temps final t_f et la condition initiale x^0 sont des constantes données

$$\left\{ \begin{array}{l} [0, t_f] \ni t \mapsto (x(t), u(t)) \in \mathbb{R}^n \times \mathbb{R}^m \\ x(0) = x^0 \\ f_1(x(t), u(t)) - \frac{d}{dt}x_1(t) = 0, \quad t \in [0, t_f] \\ \vdots \\ f_n(x(t), u(t)) - \frac{d}{dt}x_n(t) = 0, \quad t \in [0, t_f] \end{array} \right. \quad \left(l(x(t_f)) + \int_0^{t_f} L(x(t), u(t)) dt \right) \quad (3.19)$$

On est en dimension infinie car on minimise une fonctionnelle pour toutes les courbes $t \mapsto (x(t), u(t))$ qui vérifient les contraintes, i.e. pour toutes les trajectoires du système $\frac{d}{dt}x = f(x, u)$ qui démarrent en $t = 0$ de x^0 .

On voit que les contraintes $f_i(x(t), u(t)) - \frac{d}{dt}x_i(t) = 0$ dépendent d'un indice discret i et d'un indice continu t . Il est donc logique d'introduire les multiplicateurs $\lambda_i(t)$ avec ces deux types d'indice. On forme le *Lagrangien* \mathcal{L} avec une somme discrète sur i et une somme continue sur t (une intégrale donc) des produits contrainte d'indice (i, t) fois multiplicateur d'indice (i, t)

$$\begin{aligned} \mathcal{L}(x, u, \lambda) &= l(x(t_f)) + \int_0^{t_f} L(x(t), u(t)) dt \\ &\quad + \sum_{i=1}^n \int_0^{t_f} \lambda_i(t) \left(f_i(x(t), u(t)) - \frac{d}{dt}x_i(t) \right) dt \end{aligned}$$

Le Lagrangien \mathcal{L} est une fonctionnelle qui aux fonctions x, u et λ sur $[0, t_f]$ associe un réel $\mathcal{L}(x, u, \lambda)$ calculé avec une intégrale sur t . On note de façon plus compacte

$$\mathcal{L}(x, u, \lambda) = l(x(t_f)) + \int_0^{t_f} \left[L(x(t), u(t)) + \lambda^T(t) \left(f(x(t), u(t)) - \frac{d}{dt}x(t) \right) \right] dt$$

Les conditions de stationnarité du premier ordre s'obtiennent alors, comme dans le cas de la dimension finie, en explicitant qu'au premier ordre la variation $\delta\mathcal{L}$ de \mathcal{L} est nulle pour toute variation infinitésimale δx , δu et $\delta\lambda$ des courbes paramétrées $t \mapsto x(t)$, $t \mapsto u(t)$ et $t \mapsto \lambda(t)$. Puisqu'on n'a pas introduit de multiplicateur pour la contrainte $x(0) = x^0$, nous imposons à la variation δx de vérifier la contrainte $\delta x(0) = 0$.

Commençons par la variation de λ . On doit avoir, pour toute variation de la courbe $\delta\lambda$, une variation de $\delta\mathcal{L}$ nulle au premier ordre. En d'autres termes, pour toute fonction $t \mapsto \delta\lambda(t) \in \mathbb{R}^n$ on doit avoir

$$\delta\mathcal{L} = \int_0^{t_f} \delta\lambda^T(t) \left(f(x(t), u(t)) - \frac{d}{dt}x(t) \right) dt = 0$$

La seule possibilité⁹ est qu'à chaque instant $t \in [0, t_f]$, $f(x(t), u(t)) - \frac{d}{dt}x(t) = 0$. On retrouve bien les contraintes comme en dimension finie.

8. On pourra alors regarder les conditions au second ordre.

9. On se reportera au lemme de duBois-Reymond.

Poursuivons avec la variation de u . On doit avoir, pour toutes variations du contrôle δu , une variation de $\delta \mathcal{L}$ nulle au premier ordre. Pour toute fonction $t \mapsto \delta u(t) \in \mathbb{R}^m$, on doit avoir

$$\delta \mathcal{L} = \int_0^{t_f} \left(\frac{\partial L}{\partial u} \Big|_{(x(t), u(t))} \delta u(t) + \lambda^T(t) \frac{\partial f}{\partial u} \Big|_{(x(t), u(t))} \delta u(t) \right) dt = 0$$

En mettant δu en facteur on obtient

$$\delta \mathcal{L} = \int_0^{t_f} \left(\frac{\partial L}{\partial u} \Big|_{(x(t), u(t))} + \lambda^T(t) \frac{\partial f}{\partial u} \Big|_{(x(t), u(t))} \right) \delta u(t) dt = 0$$

Ceci donne la condition de stationnarité sur u

$$\frac{\partial L}{\partial u}(x, u) + \lambda^T(t) \frac{\partial f}{\partial u}(x, u) = 0, \quad t \in [0, t_f] \quad (3.20)$$

Terminons par la variation de x . On doit avoir, pour toutes variations δx vérifiant $\delta x(0) = 0$, une variation de $\delta \mathcal{L}$ nulle au premier ordre. Pour toute fonction $t \mapsto \delta x(t) \in \mathbb{R}^n$ telle que $\delta x(0) = 0$, on doit avoir

$$\begin{aligned} \delta \mathcal{L} &= \frac{\partial l}{\partial x}(x(t_f)) \delta x(t_f) + \\ &\quad \int_0^{t_f} \left[\frac{\partial L}{\partial x} \Big|_{(x(t), u(t))} \delta x(t) + \lambda^T(t) \left(\frac{\partial f}{\partial x} \Big|_{(x(t), u(t))} \delta x(t) - \frac{d}{dt} \delta x(t) \right) \right] dt = 0 \end{aligned}$$

Pour mettre, comme avec $\delta u, \delta x$ en facteur dans l'intégrale, il nous faut éliminer $\frac{d}{dt} \delta x$. La seule possibilité est de faire une intégration par partie. Cela fait apparaître $\frac{d}{dt} \lambda$. C'est ici que réside principalement la nouveauté par rapport à la dimension finie. On a

$$\begin{aligned} - \int_0^{t_f} \lambda^T(t) \frac{d}{dt} \delta x(t) dt &= -[\lambda^T \delta x]_0^{t_f} + \int_0^{t_f} \frac{d}{dt} \lambda^T(t) \delta x(t) dt \\ &= -\lambda^T(t_f) \delta x(t_f) + \int_0^{t_f} \frac{d}{dt} \lambda^T(t) \delta x(t) dt \end{aligned}$$

car $\delta x(0) = 0$. On a

$$\begin{aligned} \delta \mathcal{L} &= \left[\frac{\partial l}{\partial x}(x(t_f)) - \lambda^T(t_f) \right] \delta x(t_f) + \\ &\quad \int_0^{t_f} \left[\frac{\partial L}{\partial x} \Big|_{(x(t), u(t))} + \lambda^T(t) \frac{\partial f}{\partial x} \Big|_{(x(t), u(t))} + \frac{d}{dt} \lambda^T(t) \right] \delta x(t) dt = 0 \end{aligned}$$

Donc, pour toute fonction $t \mapsto \delta x(t)$ telle que $\delta x(0) = \delta x(t_f) = 0$, on a

$$\int_0^{t_f} \left[\frac{\partial L}{\partial x} \Big|_{(x(t), u(t))} + \lambda^T(t) \frac{\partial f}{\partial x} \Big|_{(x(t), u(t))} + \frac{d}{dt} \lambda^T(t) \right] \delta x(t) dt = 0$$

On en déduit

$$\frac{\partial L}{\partial x} \Big|_{(x(t), u(t))} + \lambda^T(t) \frac{\partial f}{\partial x} \Big|_{(x(t), u(t))} + \frac{d}{dt} \lambda^T(t) = 0$$

c.-à-d.

$$\left(\frac{\partial L}{\partial x} \right)_{(x,u)}^T + \left(\frac{\partial f}{\partial x} \right)_{(x,u)}^T \lambda + \frac{d}{dt} \lambda = 0, \quad t \in [0, t_f] \quad (3.21)$$

Enfin, avec $\delta x(t_f) \neq 0$ on obtient de $\delta \mathcal{L} = 0$ la condition finale

$$\lambda(t_f) = \frac{\partial l}{\partial x}(x(t_f))$$

En somme, les conditions d'optimalité du premier ordre pour le problème (3.19) sont pour $t \in [0, t_f]$

$$\begin{cases} \frac{d}{dt} x(t) = f(x(t), u(t)) \\ \frac{d}{dt} \lambda(t) = - \left(\frac{\partial f}{\partial x} \right)_{(x(t), u(t))}^T \lambda(t) - \left(\frac{\partial L}{\partial x} \right)_{(x(t), u(t))}^T \\ 0 = \left(\frac{\partial L}{\partial u} \right)_{(x(t), u(t))} + \lambda^T \left(\frac{\partial f}{\partial u} \right)_{(x(t), u(t))} \end{cases} \quad (3.22)$$

avec comme condition au bord

$$x(0) = x^0, \quad \lambda(t_f) = \frac{\partial l}{\partial x}(x(t_f)) \quad (3.23)$$

Il s'agit d'un *problème aux deux bouts*, ce n'est pas un *problème de Cauchy*, car la condition "initiale" est séparée entre les temps $t = 0$ et $t = t_f$ comme le précise l'équation (3.23). On appelle souvent λ *état adjoint*. Nous allons maintenant utiliser ces formules dans le cas où f est linéaire et où les fonctions L et l sont quadratiques. C'est le cas dit *linéaire quadratique*. Sous ces hypothèses, les conditions de stationnarité ci-dessus sont alors des conditions nécessaires et suffisantes d'optimalité.

3.4.2 Problème aux deux bouts dans le cas linéaire quadratique

La caractérisation de la commande optimale pour le problème de minimisation du critère (3.18) pour une condition initiale $x(0) = x^0$ donnée passe par le calcul des conditions de stationnarité (3.22) avec $f = Ax + Bu$, $L = \frac{1}{2}(x^T Rx + u^T Qu)$ et $l = \frac{1}{2}x^T S_f x$. Le problème aux deux bouts correspondant prend la forme

$$\left. \begin{array}{l} \frac{d}{dt} x(t) = A(t)x(t) - BQ^{-1}B^T \lambda(t) \\ \frac{d}{dt} \lambda(t) = -Rx(t) - A^T(t)\lambda(t) \end{array} \right\} \quad (3.24)$$

avec les conditions limites bilatérales

$$x(0) = x^0, \quad \lambda(t_f) = S_f x(t_f)$$

L'état adjoint λ est de la même dimension que x . La commande optimale est alors donnée par la dernière équation de (3.22)

$$u(t) = -Q^{-1}B^T(t)\lambda(t)$$

La particularité des équations (3.24) est qu'elles admettent une solution explicite sous forme linéaire

$$\lambda(t) = S(t)x(t)$$

avec $S(t_f) = S_f$. Nous allons détailler ce point.

En substituant dans (3.24), on obtient

$$\begin{aligned} \frac{d}{dt}S(t)x(t) + S(t)A(t)x(t) - S(t)B(t)Q^{-1}B^T(t)S(t)x(t) \\ = -Rx(t) - A^T(t)S(t)x(t) \end{aligned}$$

Il suffit alors de choisir S solution de l'*équation différentielle* matricielle *de Riccati* en temps rétrograde (quadratique en l'inconnue S)

$$\left. \begin{aligned} \frac{d}{dt}S(t) &= -S(t)A(t) + S(t)B(t)Q^{-1}B^T(t)S(t) - R - A^T(t)S(t) \\ S(t_f) &= S_f \end{aligned} \right\} \quad (3.25)$$

pour obtenir finalement la commande optimale

$$u(t) = -Q^{-1}B^T(t)S(t)x(t) \quad (3.26)$$

Le fait majeur à retenir ici est que la commande optimale (3.26) s'exprime précisément comme une loi de feedback instationnaire dont le gain peut-être calculé hors-ligne¹⁰ (i.e. avant de réaliser la moindre expérience) par la résolution de l'*équation de Riccati* (3.25).

Exemple 20 ([13]). Soit la dynamique $\frac{d}{dt}x = v$, $\frac{d}{dt}v = u$ et le critère à minimiser

$$J = \frac{1}{2} (c_1 v(t_f)^2 + c_2 x(t_f)^2) + \frac{1}{2} \int_0^{t_f} u^2(t) dt$$

où c_1 , c_2 et t_f sont des constantes positives. La commande optimale est

$$u(t) = -\lambda_x(t)x(t) - \lambda_v(t)v(t)$$

où

$$\lambda_x = \frac{1/c_2 + 1/c_1(t_f - t)^2 + 1/3(t_f - t)^3}{D(t_f - t)}$$

$$\lambda_v = \frac{1/c_1(t_f - t) + 1/2(t_f - t)^2}{D(t_f - t)}$$

avec

$$D(t_f - t) = \left(\frac{1}{c_2} + \frac{1}{3}(t_f - t)^3 \right) \left(\frac{1}{c_1} + t_f - t \right) - \frac{1}{4}(t_f - t)^4$$

De manière intéressante, il est possible d'exprimer simplement la valeur du coût J (3.18) associée à la commande optimale u (3.26). Considérons l'égalité

$$\int_0^{t_f} \frac{d}{dt} (x^T(t)S(t)x(t)) dt = x^T(t_f)S_f x(t_f) - x^T(0)S(0)x(0).$$

10. On peut montrer que le problème linéaire instationnaire avec coût quadratique considéré ici donne la solution au deuxième ordre d'un problème non-linéaire plus précis dont il est l'approximation. La loi de feedback que nous venons de calculer procure alors le calcul des extrémales de voisinages (voir [13]).

Il vient alors

$$\begin{aligned} J - \frac{1}{2}x^T(0)S(0)x(0) &= \frac{1}{2} \int_0^{t_f} \frac{d}{dt} (x^T(t)S(t)x(t)) dt \\ &\quad + \frac{1}{2} \int_0^{t_f} (x^T(t)Rx(t) + u^T(t)Qu(t)) dt \end{aligned}$$

En développant et en explicitant (3.26), on a

$$J - \frac{1}{2}x^T(0)S(0)x(0) = \frac{1}{2} \int_0^{t_f} (x^T(t) (R - SBQ^{-1}B^T S + SA + A^T S + \dot{S}) x(t)) dt$$

En utilisant (3.25), on obtient finalement

$$J - \frac{1}{2}x^T(0)S(0)x(0) = 0.$$

On vient de calculer que le coût associé à la commande optimale (3.26) est

$$J^{opt} = \frac{1}{2}x^T(0)S(0)x(0) \tag{3.27}$$

Nous pouvons maintenant récapituler nos résultats comme suit

Théorème 25 (contrôleur LQR)

Soit le système linéaire instationnaire (3.16) $\frac{d}{dt}x(t) = A(t)x(t) + B(t)u$ avec $x(0)$ comme condition initiale, et le critère à minimiser (3.18)

$$J = \frac{1}{2}x^T(t_f)S_fx(t_f) + \frac{1}{2} \int_0^{t_f} (x^T(t)Rx(t) + u^T(t)Qu(t)) dt$$

où $A(t)$ est une matrice $n \times n$, $B(t)$ est une matrice $n \times m$, S_f et R sont symétriques positives, Q est symétrique définie positive. La solution à ce problème de minimisation est la loi de *feedback optimal* (3.26)

$$u(t) = -Q^{-1}B^T(t)S(t)x(t)$$

où S est définie par l'*équation différentielle de Riccati* rétrograde (3.25), et la valeur du critère qui lui est associée est $J^{opt} = \frac{1}{2}x^T(0)S(0)x(0)$.

3.4.3 Planification de trajectoires

À l'objectif de minimiser un critère quadratique du type (3.17), on peut ajouter une *contrainte finale*, permettant d'exprimer le souhait de se rendre exactement en un point donné. On se retrouve alors avec un problème de planification de trajectoire optimale pour un *système linéaire instationnaire*. La contrainte que nous considérons ici est

$$x(t_f) = \psi \tag{3.28}$$

où $\psi \in \mathbb{R}^n$. Il est possible bien sûr de considérer des contraintes plus générales, et/ou ne portant que sur certaines composantes de l'état, on pourra se référer à [13] pour un exposé des méthodes se rapportant à ces cas.

Les conditions de stationnarité sont modifiées par l'adjonction de la contrainte (3.28). Il suffit de remplacer la condition finale dans (3.23) par la condition finale $x(t_f) = \psi$. On obtient la commande optimale par la résolution des équations suivantes

$$\frac{d}{dt}S(t) = -S(t)A(t) + S(t)B(t)Q^{-1}B^T(t)S(t) - R - A^T(t)S(t) \quad (3.29)$$

$$S(t_f) = 0 \quad (3.30)$$

$$\frac{d}{dt}V(t) = - (A^T(t) - S(t)B(t)Q^{-1}B^T(t)) V(t) \quad (3.31)$$

$$V(t_f) = I \quad (3.32)$$

$$\frac{d}{dt}H(t) = V^T(t)B(t)Q^{-1}B^T(t)V(t) \quad (3.33)$$

$$H(t_f) = 0 \quad (3.34)$$

où I est la matrice identité de taille n . On notera que (3.29) est une *équation différentielle de Riccati* qu'on devra résoudre en temps rétrograde depuis sa condition terminale (3.30). Une fois cette équation résolue, on pourra calculer $V(t)$ en résolvant l'équation linéaire (3.31) à partir de la condition terminale (3.32). Ensuite, on pourra calculer $H(t)$ à travers (3.33) et (3.34) par une simple intégration en temps rétrograde. La matrice $H(t)$ est toujours négative car $H(t_f) = 0$ et $\frac{d}{dt}H \geq 0$. Elle n'est pas singulière lorsque le système est commandable (voir [13]). Finalement, la commande optimale est

$$u(t) = -Q^{-1}B^T(t) (S(t) - V(t)H^{-1}(t)V^T(t)) x(t) - Q^{-1}B^T(t)V(t)H^{-1}(t)\psi$$

C'est encore une fois une loi de feedback instationnaire, complétée par un terme de *feedforward* lui aussi instationnaire.

Exemple 21 (Critère quadratique avec cible finale). *Considérons le système dynamique*

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) \triangleq \begin{pmatrix} 0 & 1 \\ -1/2 & 0 \end{pmatrix}x(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix}u(t)$$

On souhaite calculer une commande $[0, 3] \ni t \mapsto u(t) \in \mathbb{R}$ transférant le système de $x^0 = [-0.5 \ 1]^T$ à $x_f = [-1 \ 0]^T \triangleq \psi$ en minimisant le critère quadratique

$$J = \frac{1}{2} \int_0^3 u(t)^2 dt$$

On forme le système

$$\frac{d}{dt}S(t) = -S(t)A + S(t)BQ^{-1}B^T S(t) - A^T S(t) \quad (3.35)$$

$$S(3) = 0 \quad (3.36)$$

$$\frac{d}{dt}V(t) = - (A^T - S(t)BB^T) V(t) \quad (3.37)$$

$$V(3) = I \quad (3.38)$$

$$\frac{d}{dt}H(t) = V^T(t)BB^T V(t) \quad (3.39)$$

$$H(3) = 0 \quad (3.40)$$

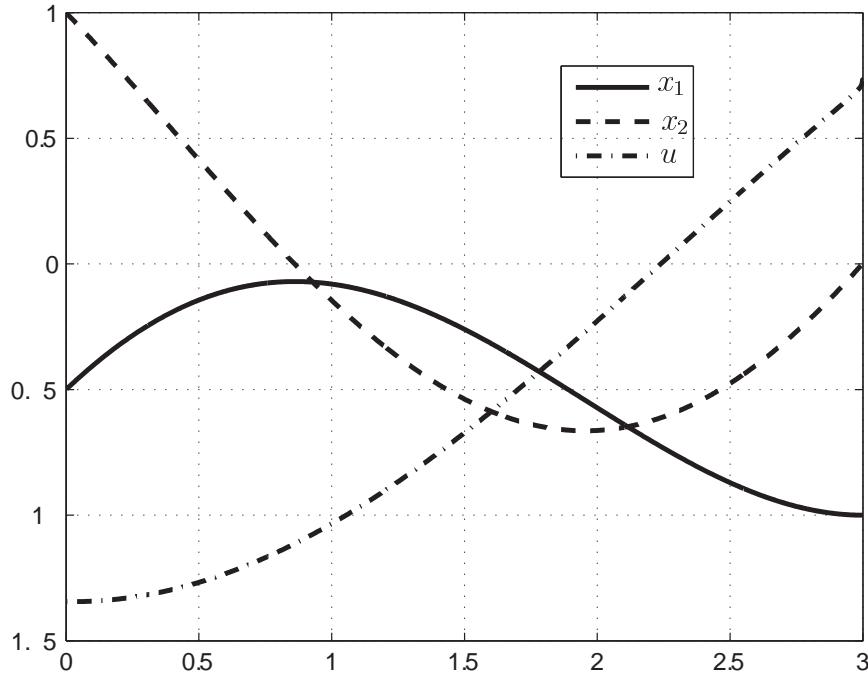


FIGURE 3.10 – Trajectographie sous contraintes finales (*planification de trajectoire*) avec minimisation quadratique.

qu'on résout en cascade comme expliqué à la Section 3.4.3. Tout d'abord, on remarque que $S(t) = 0$. L'équation (3.37) se réécrit alors $\frac{d}{dt}V(t) = -A^T V(t)$, qu'on peut résoudre explicitement avec la condition terminale (3.38). Il vient

$$V(t) = \exp(-(t-3)A^T)$$

Ensuite, on intègre numériquement (3.39) ce qui donne

$$H(t) = - \int_t^{t_f} V^T(\tau) B B^T V(\tau) d\tau$$

pour finalement calculer le contrôle optimal

$$u(t) = B^T V(t) H^{-1}(t) V^T(t) x(t) - B^T(t) V(t) H^{-1}(t) \psi$$

La résolution effective de ces calculs donne les trajectoires de la Figure 3.10.

3.4.4 Régulateur LQR

On utilise souvent une version plus simple de la commande linéaire quadratique que celle que nous venons de présenter en considérant $t_f \rightarrow +\infty$. Le coût à minimiser est alors

$$J = \int_0^{+\infty} (x^T(t) R x(t) + u^T(t) Q u(t)) dt \quad (3.41)$$

où R est une matrice symétrique positive, Q est une matrice symétrique définie positive et la dynamique que nous considérons est linéaire stationnaire

$$\frac{d}{dt}x(t) = Ax(t) + Bu \quad (3.42)$$

où l'état $x \in \mathbb{R}^n$, la commande $u \in \mathbb{R}^m$ et les matrices A et B sont de tailles $n \times n$ et $n \times m$, respectivement.

La résolution de ce problème de commande optimale est plus simple. Elle aboutit à un *feedback stationnaire* qui possède d'intéressantes propriétés.

Théorème 26 (Régulateur LQR)

Considérons le problème de minimisation du coût quadratique (3.41) pour le système (3.42) sous les hypothèses suivantes :

1. (A, B) est commandable
2. R est symétrique positive
3. Q est symétrique définie positive
4. Il existe une racine de R telle que $(A, R^{1/2})$ est *observable* (voir Chapitre 4)^a.

La solution à ce problème de minimisation est la loi de feedback optimal

$$u(t) = -Q^{-1}B^T S^0 x(t)$$

où S^0 est l'unique solution symétrique stabilisante (i.e. telle que $\frac{d}{dt}x = (A - BQ^{-1}B^T S^0)x(t)$ est asymptotiquement stable) de l'*équation de Riccati algébrique*

$$0 = SA + A^T S - SBQ^{-1}B^T S + R \quad (3.43)$$

et la valeur du critère qui lui est associée est $x^T(0)S^0x(0)$.

^a. On entend par racine de R , notée $R^{1/2}$ une matrice telle que $R = (R^{1/2})^T R^{1/2}$. Toute matrice hermitienne positive admet une telle factorisation.

Démonstration. La preuve de ce résultat s'articule en plusieurs points. La principale difficulté réside dans la preuve de l'existence et de l'unicité d'une solution symétrique stabilisante à l'équation de Riccati algébrique (3.43).

i) Commençons par montrer l'existence d'une telle solution. Considérons l'équation différentielle de Riccati obtenue à partir de (3.25) en inversant le temps

$$\frac{d\Sigma}{dt} = \Sigma A + A^T \Sigma - \Sigma B Q^{-1} B^T \Sigma + R \quad (3.44)$$

et notons $\Sigma_0(t)$ la solution correspondant à la condition initiale $\Sigma_0(0) = 0$. Comme on l'a vu à la Proposition 25, on a alors, quel que soit $x(0)$,

$$\min_u \int_0^t (x^T(t)R x(t) + u^T(t)Q u(t)) dt = x^T(0)\Sigma_0(t)x(0)$$

Cette fonction est croissante en t , car

$$\begin{aligned} \min_u \int_0^{t_2 \geq t_1} & (x^T(t)Rx(t) + u^T(t)Qu(t)) dt \\ & \geq \min_u \int_0^{t_1} (x^T(t)Rx(t) + u^T(t)Qu(t)) dt \end{aligned}$$

D'autre part, cette fonction est majorée. En effet, la paire (A, B) étant commandable, il existe, par le Théorème de placement de pôles 24 un feedback linéaire stationnaire $u(t) = -Kx(t)$ procurant la convergence exponentielle au système bouclé $\frac{d}{dt}x = (A-BK)x$. L'état, et donc la commande calculée par ce feedback, convergent tous deux exponentiellement vers zéro. On en déduit que, pour un certain $\alpha > 0$ correspondant à la valeur du critère obtenue en utilisant ce feedback linéaire stationnaire, on a

$$\min_u \int_0^t (x^T(t)Rx(t) + u^T(t)Qu(t)) dt \leq \alpha$$

La fonction $t \mapsto x^T(0)\Sigma_0(t)x(0)$ est croissante et majorée donc elle converge vers une limite, nécessairement quadratique, que nous notons $x^T(0)\Sigma_\infty x(0)$ avec Σ_∞ symétrique. Quel que soit $x(0)$, on a $\lim_{t \rightarrow \infty} x^T(0)\Sigma_0(t)x(0) = x^T(0)\Sigma_\infty x(0)$. On en déduit $\lim_{t \rightarrow \infty} \Sigma_0(t) = \Sigma_\infty$. En utilisant l'équation différentielle de Riccati (3.44), on en déduit que, lorsque $t \rightarrow \infty$, $\frac{d}{dt}\Sigma(t)$ converge lui aussi, et que sa limite est nécessairement nulle. On en déduit finalement, que Σ_∞ satisfait l'équation de Riccati algébrique

$$0 = \Sigma_\infty A + A^T \Sigma_\infty - \Sigma_\infty B Q^{-1} B^T \Sigma_\infty + R \quad (3.45)$$

Montrons que Σ_∞ est définie positive (dans le but de l'utiliser ensuite pour construire une fonction de Lyapounov). Supposons qu'il existe un $x(0)$ non nul tel que $x^T(0)\Sigma_\infty x(0) = 0$, alors, quel que soit $t \geq 0$, on a

$$\min_u \int_0^t (x^T(t)Rx(t) + u^T(t)Qu(t)) dt = 0$$

Or Q est symétrique définie positive, on en déduit $u = 0$ et, à l'optimum, on a alors

$$\int_0^t (R^{1/2}x(t))^T R^{1/2}x(t) dt = 0$$

On a alors nécessairement, que pour tout $t \geq 0$, $R^{1/2}x(\tau) = 0$ pour tout $\tau \in [0, t]$. L'équation différentielle satisfaite pour cet optimum est $\frac{d}{dt}x = Ax$ car, comme nous venons de le voir, dans ce cas la commande est nulle. On en déduit, par dérivations successives,

$$R^{1/2}x(0) = 0 = R^{1/2}Ax(0) = R^{1/2}A^2x(0) = \dots = R^{1/2}A^{n-1}x(0)$$

Or la paire $(A, R^{1/2})$ est observable (voir le critère de Kalman pour l'observabilité donné au Théorème 28). On en déduit $x(0) = 0$, d'où la contradiction. Donc, Σ_∞ est symétrique définie positive.

Posons $A_c = A - B Q^{-1} B^T \Sigma_\infty$ la matrice obtenue en fermant la boucle. On peut déduire de l'équation de Riccati (3.45) l'égalité

$$\Sigma_\infty A_c + A_c^T \Sigma_\infty = -R - \Sigma_\infty B Q^{-1} B^T \Sigma_\infty$$

Considérons la fonction (candidate à être de Lyapounov)

$$V(x) = x^T \Sigma_\infty x$$

Cette fonction continûment différentiable est strictement positive, sauf en 0 où elle est nulle. En dérivant par rapport à t , on obtient

$$\frac{d}{dt}V(x) = x^T(\Sigma_\infty A_c + A_c^T \Sigma_\infty)x = -x^T(R + \Sigma_\infty B Q^{-1} B^T \Sigma_\infty)x$$

L'ensemble des points tels que $\frac{d}{dt}V(x) = 0$ est donc défini par les équations $B^T \Sigma_\infty x = 0$ et $R^{1/2}x = 0$. L'ensemble invariant contenu dans ce sous-ensemble est réduit à l'espace des x satisfaisant

$$R^{1/2}x(0) = 0 = R^{1/2}A_c x(0) = R^{1/2}Ax(0) = \dots = R^{1/2}A^{n-1}x(0)$$

Puisque la paire $(A, R^{1/2})$ est observable, ce sous-ensemble est réduit à zéro. Par le Théorème de LaSalle 11, on en déduit la convergence asymptotique du système vers zéro. La matrice Σ_∞ est donc stabilisante.

ii) Pour prouver l'unicité nous allons utiliser le lemme suivant.

Lemme 1. *L'équation matricielle d'inconnue X*

$$XA + BX = -C$$

où A , B et C sont des matrices carrées de dimensions bien choisies, possède une unique solution si et seulement si A et $-B$ n'ont pas de valeur propre commune.

On pourra se référer à [39] pour la démonstration de ce résultat.

Considérons donc deux solutions symétriques distinctes S_1 et S_2 de l'équation (3.43) telles que

$$A_1 = A - B Q^{-1} B^T S_1 \text{ et } A_2 = A - B Q^{-1} B^T S_2$$

sont toutes deux stables. Calculons alors

$$\begin{aligned} (S_1 - S_2)A_1 + A_2^T(S_1 - S_2) \\ = S_1 A - S_1 B Q^{-1} B^T S_1 - S_2 A + S_2 B Q^{-1} B^T S_1 \\ + (A - B Q^{-1} B^T S_2)^T(S_1 - S_2) = -R + R = 0 \end{aligned}$$

en utilisant deux fois l'équation (3.43).

On en déduit l'équation matricielle

$$(S_1 - S_2)A_1 + A_2^T(S_1 - S_2) = 0$$

qui est justement de la forme évoquée dans le Lemme 1. Or, par hypothèse, A_1 et $-A_2$ ne possèdent pas de valeur propre commune car A_1 et A_2 sont stables. On en déduit

$$S_1 = S_2$$

d'où l'unicité.

iii) Calculons enfin la valeur obtenue pour la commande optimale. Considérons la commande optimale candidate $u(t) = Q^{-1} B^T S^0 x(t)$ et calculons la valeur du critère J pour une autre commande v . Cette commande devant être stabilisante, on a

$$\int_0^\infty \frac{d}{dt}(x^T S^0 x) dt = -x^T(0) S^0 x(0)$$

Poursuivons,

$$\begin{aligned} J &= x^T(0)S^0x(0) + \int_0^\infty \left(x^T(t)Rx(t) + v^T(t)Qv(t) + \frac{d}{dt}(x^T S^0 x) \right) dt \\ &= x^T(0)S^0x(0) + \\ &\quad \int_0^\infty (x^T(R + A^T S^0 + S^0 A)x + v^T Qv + v^T B^T S^0 x + x^T S^0 Bv) dt \end{aligned}$$

en utilisant l'équation de Riccati algébrique (3.43)

$$\begin{aligned} &= x^T(0)S^0x(0) + \\ &\quad \int_0^\infty (x^T S^0 B Q^{-1} B^T S^0 x + v^T Qv + v^T B^T S^0 x + x^T S^0 Bv) dt \\ &= \int_0^\infty (Q^{-1} B^T S^0 x + v)^T Q (Q^{-1} B^T S^0 x + v) dt \end{aligned}$$

Or Q est par hypothèse symétrique définie positive. Donc, l'optimum est atteint pour la commande optimale

$$-Q^{-1}B^T S^0 x$$

et la valeur de cet optimum est $x^T(0)S^0x(0)$.

□

Utilisation pratique de la commande LQR

Si (A, B) est commandable, Q diagonale à termes strictement positifs et R symétrique positive (par exemple diagonale à termes positifs), alors le théorème est applicable. Les matrices Q et R permettent d'exprimer un arbitrage entre l'importance des erreurs sur l'état et l'effort sur la commande. Plus un terme est grand, plus l'erreur s'y rapportant prend d'importance dans la fonction coût et plus elle a tendance à être réduite par la solution optimale. En résolvant l'équation de Riccati, on obtient la loi de commande optimale.

Exemple 22. Considérons le système $\frac{d}{dt}x = \frac{-1}{\tau}x + u$ et le critère quadratique $J = \frac{1}{2} \int_0^{+\infty} (ax^2 + bu^2) dt$ où $a \geq 0, b > 0$. L'équation de Riccati algébrique associée est

$$0 = \frac{2}{\tau}S - a + \frac{S^2}{b}$$

Elle possède deux solutions $S_{\pm} = \frac{-b}{\tau} \pm \sqrt{\frac{b^2}{\tau^2} + ab}$. La commande associée est $u = \left(\frac{1}{\tau} \mp \sqrt{\frac{1}{\tau^2} + \frac{a}{b}}\right) x$. La commande optimale correspond à S_+ .

Résolution de l'équation de Riccati algébrique

Lorsqu'on cherche effectivement à résoudre l'équation Riccati algébrique (3.43) on peut utiliser deux approches. La première consiste à résoudre l'équation de Riccati différentielle (3.25) en temps rétrograde (i.e. en rajoutant un signe $-$ devant le second membre) en partant de la condition initiale nulle. La solution converge pour $t \rightarrow +\infty$ vers la solution de l'équation Riccati algébrique (3.43) recherchée. Cette méthode est dûe à Kalman. C'est d'ailleurs l'idée utilisée dans la preuve du Théorème 26. Une autre méthode repose sur la décomposition de Schur de la matrice recherchée (i.e. une

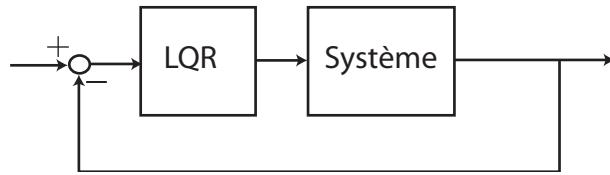


FIGURE 3.11 – Utilisation du régulateur LQR en boucle fermée.

décomposition sous la forme d'une matrice unitaire et d'une matrice triangulaire supérieure). C'est ce genre d'approche qui est utilisée dans les logiciels de calcul scientifique. On pourra se référer à différents articles sur le sujet dont [8].

Marges de stabilité du régulateur LQR

On s'intéresse ici aux marges de stabilité, définies au Chapitre 2, procurées par le régulateur LQR du Théorème 26.

Utilisé en boucle fermée tel que sur la Figure 3.11, on se retrouve à étudier le lieu de Nyquist de la fonction de transfert

$$1 + Q^{-1}B^T S^0(sI - A)^{-1}B$$

. Ainsi dans les calculs ci-dessous, $s = \omega$ est imaginaire pur et donc $s^* = -s$. Notons $K = Q^{-1}B^T S^0$. D'après l'*équation de Riccati algébrique* (3.43), on a, en faisant apparaître $(sI - A)$,

$$0 = S^0(sI - A) - (sI + A^T)S^0 - R + K^T QK$$

d'où

$$0 = -(sI + A^T)^{-1}S^0 + S^0(sI - A)^{-1} - (sI + A^T)^{-1}(K^T QK - R)(sI - A)^{-1} \quad (3.46)$$

car $((sI - A)^{-1})^T = (s^* I - A^T)^{-1} = -(sI + A^T)^{-1}$ puisque T correspond alors à la transposition hermitienne.

Calculons maintenant

$$\begin{aligned} & \overline{(I + K(sI - A)^{-1}B)}^T Q(I + K(sI - A)^{-1}B) \\ &= (I + B^T(-sI - A^T)^{-1}K^T)Q(I + K(sI - A)^{-1}B) \end{aligned}$$

qui donne après substitution partielle de K

$$\begin{aligned} &= Q + B^T(-(sI + A^T)^{-1}S^0 + S^0(sI - A)^{-1})B \\ &\quad - B^T(sI + A^T)^{-1}K^T QK(sI - A)^{-1}B \end{aligned}$$

et, en utilisant (3.46), il vient

$$= Q + \overline{(sI - A)^{-1}B}^T R(sI - A)^{-1}B$$

Nous venons d'établir dans un cadre général multivariable l'équation (dite de retour et valable uniquement pour $s = \omega$ imaginaire)

$$\overline{(I + K(sI - A)^{-1}B)}^T Q(I + K(sI - A)^{-1}B) = Q + \overline{(sI - A)^{-1}B}^T R(sI - A)^{-1}B \quad (3.47)$$

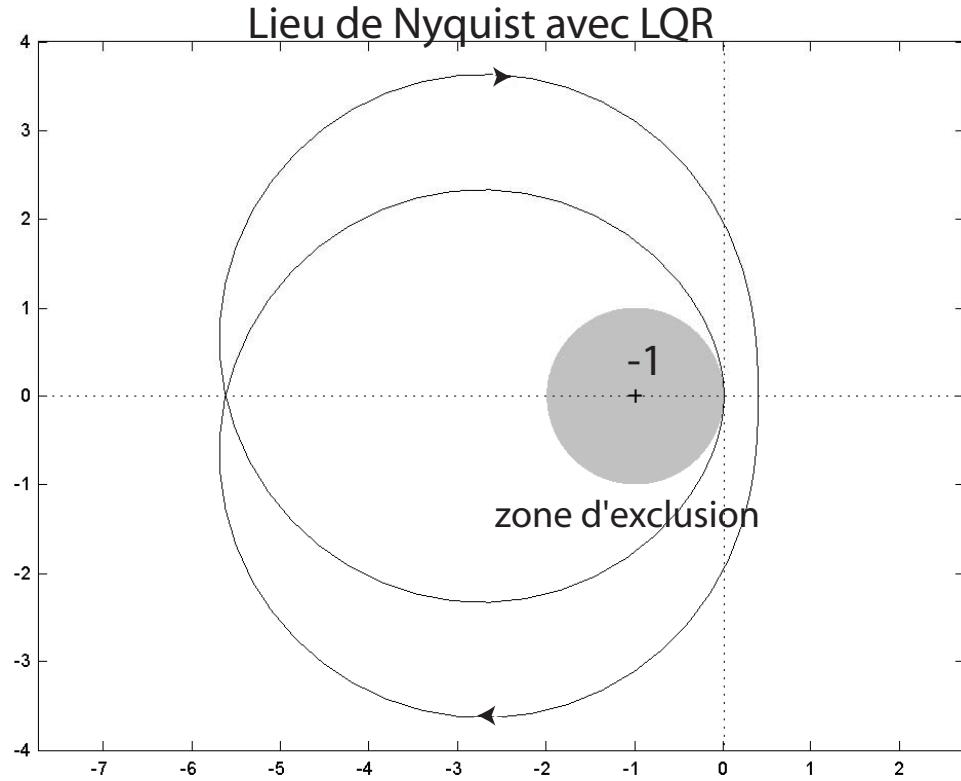


FIGURE 3.12 – Utilisation du LQR en boucle fermée.

Exprimons la dans le cas mono-entrée. Alors, Q est un scalaire strictement positif, K est un vecteur ligne, B est un vecteur colonne. Il vient alors

$$\|1 + K(sI - A)^{-1}B\|^2 = 1 + \mathcal{Q} \quad (3.48)$$

où \mathcal{Q} est un réel positif ou nul (car R n'est que positive et non pas positive définie).

Ceci implique que le *lieu de Nyquist* du système est à une distance plus grande que 1 du point -1 . Ceci procure donc toujours une marge de phase supérieure à 60 degrés, une marge de gain (positive) infinie, et une marge de gain (négative) d'au moins $1/2$, voir Figure 3.12.

Exemple 23. Considérons le système $\frac{d}{dt}x = Ax + Bu$ où

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0.3 \end{pmatrix}, \quad B = \begin{pmatrix} 0.25 \\ 1 \end{pmatrix}$$

Avec les matrices de pondération $Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $R = 1$, on obtient le régulateur LQR de gains $[0.8168 \quad 1.4823]$. Le lieu de Nyquist du système ainsi bouclé est représenté sur la Figure 3.12. Il entoure 2 fois le point -1 dans le sens indirect. C'est le nombre de pôles instables de la fonction de transfert

$$K(sI - A)^{-1}B = \frac{1.687s + 0.385}{s^2 - 0.3s + 1}$$

Le régulateur LQR assure la stabilité. Les marges de stabilité sont visibles sur la Figure 3.12.

3.5 Compléments

3.5.1 Linéarisation par bouclage

Equivalence statique

La relation d'équivalence qui permet de mettre un système linéaire $\frac{d}{dt}x = Ax + Bu$ commandable sous forme de Brunovsky peut être prolongée au cas des systèmes non linéaires de la manière suivante. Au lieu de considérer des transformations du type

$$\begin{bmatrix} x \\ u \end{bmatrix} \mapsto \begin{bmatrix} Mx \\ Kx + Nu \end{bmatrix}$$

avec M et N matrices inversibles, considérons des transformations inversibles plus générales et non linéaires suivantes

$$\begin{bmatrix} x \\ u \end{bmatrix} \mapsto \begin{bmatrix} z = \phi(x) \\ v = k(x, u) \end{bmatrix}$$

où ϕ est un difféomorphisme et à x bloqué, $u \mapsto k(x, u)$ également. On va considérer les systèmes non linéaires de la forme $\frac{d}{dt}x = f(x, u)$ et leur classification, comme nous l'avons fait pour les systèmes linéaires, modulo le groupe de transformations ci-dessus. La relation d'équivalence qui en résulte est appelée équivalence par bouclage statique régulier et changement de coordonnées (d'une façon plus abrégée *équivalence statique*). Décider si deux systèmes avec les mêmes nombres d'états et des commandes, $\frac{d}{dt}x = f(x, u)$ et $\frac{d}{dt}z = g(z, v)$, (f, g régulières) sont équivalents, est un problème de géométrie difficile et largement ouvert. En revanche, il existe une caractérisation explicite des systèmes non linéaires équivalents aux systèmes linéaires commandables.

Exemple 24. Soit le système de la Figure 3.6. On suppose ici que le ressort est non linéaire. Dans (3.14) la raideur k est fonction de $x_1 - x_2$: $k = k_0 + a(x_1 - x_2)^2$ avec k_0 et $a > 0$. On peut montrer que le système reste commandable et calculer sa sortie non linéaire de Brunovsky (la sortie plate).

Exemple 25. Reprenons le système (3.6) en ne considérant que les deux équations différentielles relatives à x_1 et T (nous ne considérons que la partie commandable). On peut montrer (formellement) que ce sous-système à deux états et une commande est commandable (la quantité $y = x_1$ joue le rôle de sortie non linéaire de Brunovsky (la sortie plate)) et en déduire le bouclage statique qui linéarise le système.

CNS de linéarisation statique

L'intérêt pratique est le suivant. Les équations issues de la physique $\frac{d}{dt}x = f(x, u)$ sont en général non linéaires dans les coordonnées de modélisation x et u . La question “Existe-t-il des coordonnées, $z = \phi(x)$ et $v = k(x, u)$, qui rendent les équations linéaires, $\frac{d}{dt}z = Az + Bv$ avec (A, B) commandable ?” est alors d'importance. En effet, une réponse positive signifie que le système est non linéaire seulement en apparence : le système est alors dit *linéarisable par bouclage statique*. Il suffit de changer de “repère” pour que tout devienne linéaire.

À partir de maintenant, nous considérons le système

$$\frac{d}{dt}x = f(x, u), \quad x \in \mathbb{R}^n, \quad u \in \mathbb{R}^m$$

avec f régulière et $f(0, 0) = 0$. Notre point de vue sera local autour de l'équilibre $(x, u) = (0, 0)$.

Lemme 2. *Les deux propositions suivantes sont équivalentes*

1. *Le système étendu*

$$\begin{cases} \frac{d}{dt}x = f(x, u) \\ \frac{d}{dt}u = \bar{u} \end{cases} \quad (3.49)$$

est linéarisable par bouclage statique (\bar{u} est ici la commande)

2. *Le système*

$$\frac{d}{dt}x = f(x, u) \quad (3.50)$$

est linéarisable par bouclage statique.

Démonstration. Si $x = \phi(z)$ et $u = k(z, v)$ transforment (3.50) en un système linéaire commandable $\frac{d}{dt}z = Az + Bv$, alors $(x, u) = (\phi(z), k(z, v))$ et $\bar{u} = \frac{\partial k}{\partial z}(Az + Bv) + \frac{\partial k}{\partial v}\bar{v}$ transforment (3.49) en

$$\frac{d}{dt}z = Az + Bv, \quad \frac{d}{dt}v = \bar{v} \quad (3.51)$$

système linéaire commandable. Ainsi la seconde proposition implique la première.

Supposons maintenant la première proposition vraie. Comme tout système linéaire commandable peut s'écrire sous la forme (3.51) avec (A, B) commandable, (*forme de Brunovsky*) il existe une transformation $(x, u) = (\phi(z, v), \psi(z, v))$ et $\bar{u} = k(z, v, \bar{v})$ qui transforme (3.49) en (3.51) avec $\dim(z) = \dim(x)$. Cela veut dire que pour tout (z, v, \bar{v})

$$\frac{\partial \phi}{\partial z}(z, v)(Az + Bv) + \frac{\partial \phi}{\partial v}\bar{v} = f(\phi(z, v), \psi(z, v)).$$

Donc ϕ ne dépend pas de v et la transformation inversible $x = \phi(z)$, $u = \psi(z, v)$ transforme (3.50) en $\frac{d}{dt}z = Az + Bv$. \square

Quitte à étendre l'état en posant $\frac{d}{dt}u = \bar{u}$ et en prenant comme entrée \bar{u} , on peut toujours supposer que f est affine en u , i.e., que le système admet les équations

$$\frac{d}{dt}x = f(x) + u_1g_1(x) + \dots + u_mg_m(x) \quad (3.52)$$

où f et les g_i sont des champs de vecteurs réguliers. Il est alors facile de voir que les transformations $x = \phi(z)$ et $u = k(z, v)$ qui rendent le système linéaire sont nécessairement affines en v , i.e., $k(x, v) = \alpha(x) + \beta(x)v$ avec β inversible pour tout x .

Prenons maintenant un changement régulier de variables $x = \phi(z)$ d'inverse $\psi = \phi^{-1}$, $z = \psi(x)$. Considérons le système défini par (3.52) dans le repère x . Dans le repère z , on obtient

$$\frac{d}{dt}z = (D\psi \cdot f + u_2D\psi \cdot g_1 + \dots + u_mD\psi \cdot g_m)_{x=\phi(z)} \quad (3.53)$$

où $D\psi$ est la matrice Jacobienne de $\psi : \left(\frac{\partial \psi_i}{\partial x_j} \right)_{i,j}$. Ainsi, f (respectivement g_k) devient $D\psi \cdot f$ (respectivement $D\psi \cdot g_k$).

À partir de ces *champs de vecteurs* définissant (3.52), on construit, par la récurrence suivante, une suite croissante d'espaces vectoriels indexés par x

$$E_0 = \{g_1, \dots, g_m\}, \quad E_i = \{E_{i-1}, [f, E_{i-1}]\} \quad i \geq 1$$

où $[f, g]$ est le *crochet de Lie* de deux champs de vecteurs f et g et où $\{ \}$ signifie espace vectoriel engendré par les vecteurs à l'intérieur des parenthèses. On rappelle que le crochet de deux champs de vecteurs f et g , de composantes $(f_1(x), \dots, f_n(x))$ et $(g_1(x), \dots, g_n(x))$ dans les coordonnées (x_1, \dots, x_n) , admet comme composantes dans les mêmes coordonnées x

$$[f, g]_i = \sum_{k=1}^n \frac{\partial f_i}{\partial x_k} g_k - \frac{\partial g_i}{\partial x_k} f_k$$

Un simple calcul montre que si $z = \psi(x)$ est un changement régulier de variables on obtient les composantes du crochet $[f, g]$ dans les coordonnées z par les mêmes formules que dans les coordonnées x . En d'autres termes

$$D\psi.[f, g] = [D\psi.f, D\psi.g]$$

On sait faire du *calcul différentiel intrinsèque* sans passer par un choix particulier de repère. Les E_k deviennent, dans les coordonnées z , $D\psi.E_k$. On appelle ce type d'objet des distributions (rien à voir avec les distributions de L. Schwartz). Ce sont des objets intrinsèques car la méthode de construction de E_k ne dépend pas du système de coordonnées choisi pour faire les calculs.

Théorème 27 (CNS de linéarisation statique)

Autour de l'équilibre $(x, u) = (0, 0)$, le système (3.52) est linéarisable par bouclage statique régulier si et seulement si les distributions $E_i, i = 1, \dots, n-1$ définies ci-dessus sont involutives (stables par le crochet de Lie), de rang constant autour de $x = 0$ et le rang de E_{n-1} vaut n , la dimension de x .

Une distribution E est dite involutive si et seulement si pour tous champs de vecteurs f et g dans E (pour tout x , $f(x)$ et $g(x)$ appartiennent à l'espace vectoriel $E(x)$) le crochet $[f, g]$ reste aussi dans E .

Démonstration. Il est évident que les distributions E_i restent inchangées par bouclage statique $u = \alpha(x) + \beta(x)v$ avec $\beta(x)$ inversible. Comme pour un système linéaire commandable $\frac{d}{dt}x = Ax + Bu$, E_i correspond à l'image de (B, AB, \dots, A^iB) , les conditions sur les E_i sont donc nécessaires.

Leur côté suffisant repose essentiellement sur le théorème de Frobenius [28]. Ce résultat classique de géométrie différentielle dit que toute distribution involutive E de rang constant m correspond, dans des coordonnées adaptées $w = (w_1, \dots, w_n)$, à l'espace vectoriel engendré par les m premières composantes. On a l'habitude de noter $\partial/\partial w_k$ le champ de vecteurs de composantes $(\delta_{i,k})_{1 \leq i \leq n}$ dans les coordonnées w . Alors $E = \{\partial/\partial w_1, \dots, \partial/\partial w_m\}$.

Si les E_i vérifient les conditions du théorème, alors il existe un système de coordonnées locales (x_1, \dots, x_n) autour de 0 tel que

$$E_i = \left\{ \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_{\sigma_i}} \right\}$$

où σ_i est le rang de E_i . Dans ces coordonnées locales, $\frac{d}{dt}x_i$ pour $i > \sigma_0$ ne dépend pas de la commande u . Ainsi, en remplaçant u par $\alpha(x) + \beta(x)u$ avec une matrice inversible β bien choisie, la dynamique (3.52) s'écrit nécessairement sous la forme

$$\begin{aligned} \frac{d}{dt}x_i &= u_i, \quad i = 1, \dots, \sigma_0 \\ \frac{d}{dt}x_i &= f_i(x), \quad i = \sigma_0 + 1, \dots, n \end{aligned}$$

Un raisonnement simple montre que, pour $i > \sigma_1$, f_i ne dépend pas de $(x_1, \dots, x_{\sigma_0})$ car E_1 est involutive. Ainsi nous avons la structure suivante

$$\begin{aligned}\frac{d}{dt}x_i &= u_i, \quad i = 1, \dots, \sigma_0 \\ \frac{d}{dt}x_i &= f_i(x_1, \dots, x_n), \quad i = \sigma_0 + 1, \dots, \sigma_1 \\ \frac{d}{dt}x_i &= f_i(x_{\sigma_0+1}, \dots, x_n), \quad i = \sigma_1 + 1, \dots, n\end{aligned}$$

De plus, le rang de $(f_{\sigma_0+1}, \dots, f_{\sigma_1})$ par rapport à $(x_1, \dots, x_{\sigma_0})$ vaut $\sigma_1 - \sigma_0$. Donc $\sigma_0 \leq \sigma_1 - \sigma_0$. Quitte à faire des permutations sur les σ_0 premières composantes de x , on peut supposer que $(x_1, \dots, x_{\sigma_1-\sigma_0}) \mapsto (f_{\sigma_0+1}, \dots, f_{\sigma_1})$ est inversible. Cela permet de définir un nouveau système de coordonnées en remplaçant les $\sigma_1 - \sigma_0$ premières composantes de x par $(f_{\sigma_0+1}, \dots, f_{\sigma_1})$. Dans ces nouvelles coordonnées, après bouclage statique régulier $u \mapsto \beta(x)u$ avec $\beta(x)$ inversible bien choisi, nous obtenons la structure suivante (les notations avec u , x et f sont conservées)

$$\begin{aligned}\frac{d}{dt}x_i &= u_i, \quad i = 1, \dots, \sigma_0 \\ \frac{d}{dt}x_i &= x_{i-\sigma_0}, \quad i = \sigma_0 + 1, \dots, \sigma_1 \\ \frac{d}{dt}x_i &= f_i(x_{\sigma_0+1}, \dots, x_n), \quad i = \sigma_1 + 1, \dots, n\end{aligned}$$

On sait que ce système est linéarisable si et seulement si le système réduit

$$\begin{aligned}\frac{d}{dt}x_i &= x_{i-\sigma_0}, \quad i = \sigma_0 + 1, \dots, \sigma_1 \\ \frac{d}{dt}x_i &= f_i(x_{\sigma_0+1}, \dots, x_n), \quad i = \sigma_1 + 1, \dots, n\end{aligned}$$

l'est avec $(x_1, \dots, x_{\sigma_1-\sigma_0})$ comme commande. Comme les distributions E_i associées à ce système réduit se déduisent simplement de celles du système étendu en éliminant les champs de vecteurs $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_{\sigma_0}}$, on voit qu'elles vérifient, elles aussi, les conditions du théorème. Il est donc possible de réduire encore le système. À chaque étape, la linéarisation du système étendu est équivalente à celle du système réduit. Au bout de cette élimination (en au plus de $n - 1$ étapes), la linéarisation du système de départ est alors équivalente à celle d'un système réduit de la forme

$$\frac{d}{dt}x = f(x, u)$$

où le rang de f par rapport à u est égale à la dimension de x , linéarisation qui est alors triviale. \square

Bouclage dynamique

Le Lemme 2 est trompeur. Il semble suggérer que le fait d'étendre un système en rajoutant des dérivées de la commande dans l'état ne rajoute rien pour la linéarisation. Ceci est vrai si on rajoute le même nombre d'intégrateurs sur toutes les commandes (c.-à-d. si on effectue une "prolongation totale"). Par contre, des nombres différents peuvent permettre de gagner quelque chose. Par exemple le système

$$\ddot{x} = -u_1 \sin \theta, \quad \ddot{z} = u_1 \cos \theta - 1, \quad \ddot{\theta} = u_2$$

n'est pas linéarisable par bouclage statique bien que le système étendu

$$\ddot{x} = -u_1 \sin \theta, \quad \ddot{z} = u_1 \cos \theta - 1, \quad \ddot{u}_1 = \bar{u}_1, \quad \ddot{\theta} = u_2$$

de commande (\bar{u}_1, u_2) le soit. Ce fait n'est nullement en contradiction avec le Lemme 2 puisque seule l'entrée u_1 a été prolongée par des intégrateurs deux fois. Pour un système à une seule commande, on ne gagne évidemment rien.

Cette remarque est à l'origine de la technique de *linéarisation par bouclage dynamique*. Un système $\frac{d}{dt}x = f(x, u)$ est dit linéarisable par bouclage dynamique régulier si et seulement si il existe un compensateur dynamique régulier

$$\frac{d}{dt}\xi = a(x, \xi, v), \quad u = k(x, \xi, v),$$

tel que le système bouclé

$$\frac{d}{dt}x = f(x, k(x, \xi, v)), \quad \frac{d}{dt}\xi = a(x, \xi, v)$$

soit *linéarisable par bouclage statique* régulier. Noter que la dimension de ξ est libre. La dimension de l'espace dans lequel on doit travailler peut a priori être arbitrairement grande. Noter également que les compensateurs dynamiques qui consistent à ne prolonger que les entrées, sont des compensateurs particuliers. Ils sont insuffisants car ils ne permettent pas de linéariser certains systèmes comme celui ci

$$\begin{aligned} \ddot{x} &= \varepsilon u_2 \cos \theta - u_1 \sin \theta \\ \ddot{z} &= \varepsilon u_2 \sin \theta + u_1 \cos \theta - g \\ \ddot{\theta} &= u_2 \end{aligned}$$

On peut montrer que, quel que soit le compensateur dynamique de la forme $u_1^{(\alpha_1)} = \bar{u}_1, u_2^{(\alpha_2)} = \bar{u}_2$ (α_1 et α_2 entiers arbitraires), le système étendu n'est pas linéarisable par bouclage statique. En revanche, il est linéarisable par des bouclages dynamiques plus généraux.

Cette question est à l'origine des *systèmes plats*, les systèmes linéarisables par des bouclages dynamiques dits endogènes et auxquels est associée une relation d'équivalence (i.e., une géométrie). Pour en savoir plus, on se reportera à [57].

3.5.2 Stabilisation par méthode de Lyapounov et backstepping

On a déjà découvert cette méthode de stabilisation autour d'un point d'équilibre lors de l'exemple représenté sur la Figure 3.1. On reprend ici simplement l'idée de base pour un système non linéaire¹¹.

On part d'un système dynamique avec un seul contrôle u (pour simplifier),

$$\frac{d}{dt}x = f(x) + ug(x)$$

et d'une fonction $V(x)$ positive et qui tend à l'infini quand $x \in \mathbb{R}^n$ tend vers l'infini en norme. On suppose ici que toutes les fonctions sont dérivables autant de fois que nécessaire.

11. L'idée s'applique aussi aux systèmes de dimension infinie, i.e., gouvernés par des équations aux dérivées partielles avec un contrôle sur la frontière.

On suppose de plus que pour $u = 0$, la dérivée de V le long des trajectoires $\frac{d}{dt}x = f(x)$ est négative ou nulle. Pour les systèmes mécaniques, V est souvent l'énergie mécanique. Ainsi par hypothèse pour tout x on a

$$\frac{\partial V}{\partial x}f(x) \leq 0$$

Avec $u \neq 0$, on a

$$\frac{d}{dt}V = \frac{\partial V}{\partial x}f(x) + u\frac{\partial V}{\partial x}g(x)$$

Le feedback est alors construit avec n'importe quelle fonction $K : \mathbb{R} \mapsto \mathbb{R}$ assez régulière qui s'annule en 0 uniquement et telle que $K(\xi) > 0$ si $\xi < 0$ et $K(\xi) < 0$ si $\xi > 0$

$$u = K\left(\frac{\partial V}{\partial x}g(x)\right)$$

Le principe d'invariance de LaSalle s'applique alors (voir le Théorème 11). L'ensemble vers lequel convergent les trajectoires du système en boucle fermée est donné en résolvant le système sur-déterminé suivant

$$\frac{d}{dt}x = f(x), \quad \frac{\partial V}{\partial x}f(x) = 0, \quad \frac{\partial V}{\partial x}g(x) = 0.$$

La méthode consiste à dériver un certain nombre de fois les deux équations “algébriques” $\frac{\partial V}{\partial x}f(x) = 0$ et $\frac{\partial V}{\partial x}g(x) = 0$ par rapport au temps et à remplacer à chaque fois $\frac{d}{dt}x$ par $f(x)$. La fonction V est souvent appelée “control Lyapounov function” ou CLF.

La méthode dite du “backstepping” résout le problème suivant. Les données sont

- un système affine en contrôle (on suppose toujours, pour simplifier l'exposé, un seul contrôle scalaire)

$$\frac{d}{dt}x = f(x) + ug(x)$$

- une fonction de Lyapounov V pour $u = 0$ (une CLF donc)

On peut donc utiliser la méthode ci-dessous pour construire le feedback, noté $u = k(x)$, qui stabilise le système. On souhaite en déduire un feedback stabilisant pour le *système étendu*

$$\frac{d}{dt}x = f(x) + \nu g(x), \quad \frac{d}{dt}\nu = u$$

où maintenant le contrôle n'est plus ν mais sa dérivée u . On considère la fonction suivante

$$W(x, \nu) = V(x) + \frac{1}{2}(\nu - k(x))^2$$

Ainsi W est bien infinie à l'infini et reste positive. Calculons sa dérivée par rapport au temps

$$\frac{d}{dt}W = \frac{\partial V}{\partial x}(f(x) + \nu g(x)) + p(\nu - k(x)) \left(u - \frac{\partial k}{\partial x}(f(x) + \nu g(x)) \right)$$

Utilisons le fait qu'avec $\nu = k(x)$ la dérivée $\frac{d}{dt}V = \frac{\partial V}{\partial x}(f(x) + k(x)g(x))$ est négative. Cela conduit au réarrangement suivant du second membre

$$\frac{d}{dt}W = \frac{\partial V}{\partial x}(f(x) + k(x)g(x)) + (\nu - k(x)) \left(u - \frac{\partial k}{\partial x}(f(x) + \nu g(x)) + \frac{\partial V}{\partial x}g(x) \right)$$

Il suffit, avec p paramètre positif, de poser

$$u = \frac{\partial k}{\partial x}(f(x) + \nu g(x)) - \frac{\partial V}{\partial x}g(x) - p(\nu - k(x))$$

pour avoir

$$\frac{d}{dt}W = \frac{\partial V}{\partial x}(f(x) + k(x)g(x)) - p(\nu - k(x))^2 \leq 0$$

On conclut l'analyse toujours par l'invariance de LaSalle. Pour un exposé détaillé des possibilités de ces méthodes nous renvoyons le lecteur intéressé à [67].

Chapitre 4

Observabilité, estimation et adaptation

Dans le chapitre précédent, nous avons vu comment stabiliser un système par retour d'état (feedback). La réalisation pratique de tels bouclages nécessite la connaissance à chaque instant de l'état x . Or, il est fréquent que seule une partie de l'état soit directement mesurée. On est alors confronté au problème suivant. Connaissant les équations du système (i.e., ayant un modèle), $\frac{d}{dt}x = f(x, u)$, les relations entre les mesures y et l'état, $y = h(x)$, les entrées $t \mapsto u(t)$ et les mesures $t \mapsto y(t)$, il nous faut estimer x . Cela revient à résoudre le problème suivant

$$\frac{d}{dt}x = f(x, u), \quad y = h(x)$$

où x est l'inconnue (une fonction du temps) et où u et y sont des fonctions connues du temps. Il est clair que ce problème est *sur-déterminé*. L'unicité de la solution correspond à la propriété d'observabilité que nous allons définir. L'existence d'une solution provient du fait que y et u ne peuvent pas être des fonctions du temps indépendantes l'une de l'autre. Elles doivent vérifier des relations de compatibilité qui prennent la forme d'équations différentielles (le modèle).

Ce chapitre aborde ces questions. Tout d'abord nous donnons, dans un cadre général, les définitions et les critères assurant l'existence et l'unicité de la solution. Pour les systèmes linéaires nous présentons une méthode très économique en calculs pour obtenir x avec un observateur asymptotique. Nous montrons aussi comment paramétriser de "façon optimale" un tel observateur asymptotique grâce au *filtre de Kalman*. Le couplage avec la loi de feedback est alors simple : il suffit de remplacer dans les formules donnant la loi de rétro-action les valeurs de l'état par leur estimées. On montre alors que le contrôleur ainsi obtenu, celui qui part des sorties, estime l'état et utilise cette estimation pour calculer le contrôle (observateur contrôleur, commande modale, bouclage dynamique de sortie), permet de stabiliser asymptotiquement tout système linéaire à coefficients constants commandable et observable (c'est le *principe de séparation*). Enfin, nous proposons quelques extensions aux cas non linéaires avec la synthèse d'observateurs asymptotiques via l'injection de sortie et la notion de contraction.

4.1 Un exemple

Cet exemple est représentatif de ce que, dans certains domaines industriels, les ingénieurs appellent les *capteurs logiciels*. Il s'agit, pour un moteur électrique, d'estimer la vitesse de rotation et son couple de charge via les signaux de courants (mesure) et de tension (contrôle). Il est possible d'obtenir des estimations de ces deux variables mécaniques, coûteuses à mesurer, en partant des variables électriques, qui sont simples et faciles à mesurer. Pour simplifier, nous traitons ici le cas des

moteurs à courant continu. Pour les moteurs à induction, le même problème se pose mais il est nettement plus complexe, non complètement résolu et fait encore l'objet de travaux de recherches et de développement¹.

Les capteurs logiciels sont des outils de traitement en temps-réel de l'information. Ils fournissent (idéalement) des informations non bruitées sur des grandeurs mesurées ou non. Sur l'exemple choisi ici, il s'agit, à partir de la mesure des tensions et des courants qui traversent le moteur, d'estimer de façon causale sa vitesse mécanique et son couple de charge. L'intérêt pratique est évident : les informations électriques sont toujours disponibles car les capteurs sont simples et fiables. Au contraire, les informations mécaniques nécessitent une instrumentation complexe, chère et peu fiable. On cherche à s'en passer. Pour des raisons de coût mais aussi de sécurité, déduire des courants et tensions, la vitesse de rotation est un enjeu technologique important en électro-technique.

Dans d'autres domaines, on rencontre des problèmes très similaires. Pour les débits, températures et pressions sont faciles à avoir par des capteurs simples et robustes alors que les compositions sont plus difficiles à mesurer (temps de retard de l'analyse, ...). Un traitement de l'information contenue dans les signaux de températures, pressions et débits permet souvent d'obtenir des estimations précieuses sur les compositions. Un autre exemple intéressant (également évoqué dans l'Exemple 28) est l'estimation de l'orientation relative d'un mobile par rapport à un référentiel terrestre. Les mesures sont de deux types : les gyromètres donnent de façon précise et rapide les vitesses angulaires ; le GPS ou les images issues caméra embarquées donnent des informations basse fréquence sur les positions et orientations. On peut déduire grâce aux relations cinématiques, une estimation robuste de l'orientation du mobile (ses trois angles d'Euler, i.e., une matrice de rotation), de sa vitesse et de sa position. Ce problème est central dans les techniques de guidage de missiles et de drones.

L'exemple du moteur à courant continu que nous présentons en détails illustre la notion d'*observabilité* (voir la Définition 16), la technique des *observateurs asymptotiques* (voir la Section 4.3.2), et l'*observateur-contrôleur* (voir la Section 4.4).

4.1.1 Un modèle simple de moteur à courant continu

Un premier modèle de moteur à courant continu est le suivant

$$\begin{aligned} J \frac{d}{dt} \omega &= k \iota - p \\ L \frac{d}{dt} \iota &= -k\omega - R\iota + u \end{aligned}$$

où ω est la vitesse de rotation du moteur, ι le courant, u la tension, $L > 0$ la self, $R > 0$ la résistance, k la constante de couple, p le couple de charge et J l'inertie de la partie tournante (moteur + charge).

Nous supposons les paramètres $J > 0$, $k > 0$, $L > 0$ et $R > 0$ connus et constants. En revanche, seule l'intensité ι est mesurée. La charge p est une constante inconnue. Il nous faut concevoir un algorithme qui ajuste en temps réel la tension u de façon à suivre une vitesse de référence $\omega_r(t)$ variable dans le temps. Pour cela nous ne disposons que d'un capteur de courant.

1. Le système est inobservable au premier ordre et très sensible aux paramètres à basse vitesse.

4.1.2 Estimation de la vitesse et de la charge

Comme p est un paramètre constant, $\frac{d}{dt}p = 0$ et on peut l'inclure dans les variables d'état. On a alors le système

$$\begin{aligned}\frac{d}{dt}p &= 0 \\ J \frac{d}{dt}\omega &= k\iota - p \\ L \frac{d}{dt}\iota &= -k\omega - R\iota + u\end{aligned}$$

avec comme commande u et comme sortie $y = \iota$. Étudier l'observabilité de ce système revient à se poser la question suivante : connaissant $t \mapsto (\iota(t), u(t))$ et les équations du système, est-il possible de calculer ω et p ? La réponse est positive et immédiate car

$$\omega = (u - L \frac{d}{dt}\iota - R\iota)/k, \quad p = k\iota - (J/k) \left(\frac{d}{dt}u - L \frac{d^2}{dt^2}\iota - R \frac{d}{dt}\iota \right)$$

Le système est donc observable. On pourrait aussi reprendre le critère de Kalman (Théorème 28). Dans l'esprit, il est issu du même calcul.

Cependant les mesures de courant sont bruitées. Il est donc hors de question de dériver ce signal. Le fait d'être en théorie observable ne donne pas un algorithme d'estimation réaliste. Il nous faut concevoir un algorithme qui soit insensible au bruit, i.e., qui les filtre sans introduire de déphasage comme le ferait un simple filtre passe-bas. Ici apparaît une idée centrale : l'*observateur asymptotique*. Cette idée consiste à copier la dynamique du système en lui rajoutant des termes correctifs liés à l'erreur entre la prédiction et la mesure. Cela donne ici l'observateur suivant

$$\begin{aligned}\frac{d}{dt}\hat{p} &= -L_p(\hat{\iota} - \iota) \\ J \frac{d}{dt}\hat{\omega} &= k\hat{\iota} - \hat{p} - L_\omega(\hat{\iota} - \iota) \\ L \frac{d}{dt}\hat{\iota} &= -k\hat{\omega} - R\hat{\iota} + u - L_\iota(\hat{\iota} - \iota)\end{aligned}$$

Il est d'usage de rajouter un “ \wedge ” sur les estimées. On parle souvent pour le paramètre p d'*identification*, \hat{p} étant la valeur identifiée et pour ι et ω d'*estimation* ou de *filtrage*, $\hat{\iota}$ et $\hat{\omega}$ étant les valeurs estimées et filtrées (c'est à dire débarrassées des bruits hautes fréquences). En choisissant correctement les gains L_p , L_ω et L_ι les écarts entre les estimées et les vraies valeurs tendent vers zéro. En effet la dynamique des erreurs $e_p = \hat{p} - p$, $e_\omega = \hat{\omega} - \omega$ et $e_\iota = \hat{\iota} - \iota$ s'écrit

$$\begin{aligned}\frac{d}{dt}e_p &= -L_p e_\iota \\ J \frac{d}{dt}e_\omega &= -e_p - (k - L_\omega)e_\iota \\ L \frac{d}{dt}e_\iota &= -(L_\iota + R)e_\iota - k e_\omega\end{aligned}$$

Il s'agit d'un *système linéaire stationnaire* dont les valeurs propres sont les racines du polynôme en s de degré 3 suivant

$$s^3 + \frac{L_\iota + R}{L}s^2 + \frac{k(k - L_\omega)}{LJ}s + \frac{L_p k}{LJ}$$

Étant donné qu'en jouant sur les L_p , L_ω et L_ι , on peut donner n'importe quelles valeurs aux fonctions symétriques des racines, il est possible de les choisir comme l'on veut. Si L_p , L_ω et L_ι vérifient

$$\begin{aligned}-\frac{L_\iota + R}{L} &= r_1 + r_2 + r_3 \\ \frac{k(k - L_\omega)}{LJ} &= r_1 r_2 + r_1 r_3 + r_2 r_3 \\ -\frac{L_p k}{LJ} &= r_1 r_2 r_3\end{aligned}$$

alors les racines seront r_1, r_2 et r_3 . On peut choisir pour les pôles d'observation (on le verra en détail dans le Théorème 29 de *placement des pôles de l'observateur*) les valeurs suivantes

$$r_1 = -\frac{R}{L}(1 + \sqrt{-1}), \quad r_2 = -\frac{R}{L}(1 - \sqrt{-1}), \quad r_3 = -\sqrt{\frac{k^2}{LJ}}$$

Ce choix correspond aux échelles de temps caractéristiques du système en boucle ouverte.

Les valeurs $\hat{p}, \hat{\omega}$ et $\hat{\iota}$ ainsi calculées convergent vers p, ω et ι , quelle que soit la loi horaire $t \mapsto u(t)$.

4.1.3 Prise en compte des échelles de temps

Supposons, et c'est souvent le cas, que la dynamique électrique est nettement plus rapide que la dynamique mécanique. Cela revient à dire que la self L est très petite, positive mais mal connue. Ainsi, tout ce que l'on sait c'est que $L \approx \varepsilon$ où ε est un petit paramètre positif. Il est alors facile de voir que les résultats théoriques sur les perturbations singulières (voir Section 1.4.1) s'appliquent ici : le système reste stable en boucle ouverte avec une dynamique du courant convergeant immédiatement vers son régime quasi-statique

$$\iota = (u - k\omega)/R$$

La dynamique lente de la vitesse étant alors

$$J \frac{d}{dt} \omega = -(k^2/R)\omega + (k/R)u - p.$$

Il suffit de remarquer que, puisque $\omega = (u - R\iota)/k$, la vitesse est indirectement connue en combinant la tension et la mesure de courant.

L'observateur asymptotique aura alors la forme plus simple suivante

$$\begin{aligned} \frac{d}{dt} \hat{p} &= -L_p(\hat{\omega} - z(t)) \\ J \frac{d}{dt} \hat{\omega} &= -(k^2/R)\hat{\omega} + (k/R)u - \hat{p} - L_\omega(\hat{\omega} - z(t)) \end{aligned} \tag{4.1}$$

où la mesure $z(t) = (u(t) - R\iota(t))/k$ correspond à la vitesse ω , suite à l'approximation quasi-statique pour le courant. Les gains $-L_p$ et L_ω doivent être choisis positifs pour garantir la stabilité et pas trop grands à cause de l'approximation quasi-statique sur le courant.

Le suivi de la référence en vitesse $t \mapsto \omega_r(t)$ sera alors assuré par le feedback u solution de

$$-(k^2/R)\omega + (k/R)u - p = J \left(\frac{d}{dt} \omega_r - (\omega - \omega_r)/\tau \right)$$

si on avait directement accès à ω et p . Ici $\tau > 0$ est la constante de temps de suivi. τ ne doit pas être trop petit toujours à cause de l'approximation quasi-statique sur le courant. Il est alors naturel de remplacer ω et p par leur estimées pour calculer u

$$-(k^2/R)\hat{\omega} + (k/R)u - \hat{p} = J \left(\frac{d}{dt} \omega_r - (\hat{\omega} - \omega_r)/\tau \right)$$

Nous avons obtenu le bouclage dynamique de sortie $y = \iota$ avec $y_r = \omega_r(t)$ une référence continûment dérivable en t (la partie dynamique étant alors formée par l'observateur de ω et p)

$$\begin{aligned} \frac{d}{dt} \hat{p} &= -L_p(\hat{\omega} - z(t)) \\ J \frac{d}{dt} \hat{\omega} &= -(k^2/R)\hat{\omega} + (k/R)u(t) - \hat{p} - L_\omega(\hat{\omega} - z(t)) \\ z(t) &= (u(t) - Ry(t))/k \\ u(t) &= k\hat{\omega} + \frac{R}{k}\hat{p} + J \left(\frac{d}{dt} y_r(t) - \frac{\hat{\omega} - y_r(t)}{\tau} \right) \end{aligned}$$

4.1.4 Contraintes sur les courants

Il est important pour des raisons de sécurité de garantir un courant ι borné. Ainsi nous avons comme contrainte

$$|\iota| \leq \iota_{max}$$

où $\iota_{max} > 0$ est le courant maximum supporté par le variateur et le moteur. Les algorithmes précédents ne garantissent pas le respect de cette contrainte d'état. Néanmoins, il est possible de prendre en compte cette contrainte en ne modifiant que le contrôleur de la section précédente sans toucher à l'observateur. On considère donc un premier bouclage grand gain en courant :

$$u = -R\bar{\iota} - k\hat{\omega} + \frac{1}{\eta}(\bar{\iota} - \iota)$$

où $\bar{\iota}$ est une référence de courant et η tel que $\eta R \ll 1$. Un tel bouclage rend la dynamique du courant stable et bien plus rapide que celle de la vitesse. On pourra remarquer que, même si L est petit, ce bouclage ne fait que renforcer la rapidité du courant sans le déstabiliser. En effet on a

$$L \frac{d}{dt}\iota = ke_\omega - (R + 1/\eta)(\iota - \bar{\iota})$$

et donc $\iota \approx \bar{\iota}$ est une très bonne approximation. La dynamique de la vitesse se réduit à

$$J \frac{d}{dt}\omega = k\bar{\iota} - p$$

Une justification mathématique de cette réduction relève encore de la théorie des perturbations singulières.

La référence de courant peut alors être calculée ainsi

$$\bar{\iota} = \frac{J\dot{\omega}_r - J(\hat{\omega} - \omega_r)/\tau + \hat{p}}{k}$$

où $\tau > 0$ est un temps nettement supérieur l'échelle de temps de la dynamique du courant, i.e., $\tau \gg L/(R + 1/\eta)$ avec $\hat{\omega}$ et \hat{p} solution de (4.1).

Si la référence de courant $\bar{\iota}$ ainsi calculée ne vérifie pas la contrainte, alors il suffit de la saturer en valeur absolue à ι_{max} en préservant son signe. Il est possible de voir qu'une telle saturation ne peut pas déstabiliser la vitesse. Elle assure de fait le suivi au mieux de la référence ω_r .

4.2 Observabilité non linéaire

Considérons les systèmes non linéaires de la forme

$$\begin{cases} \frac{d}{dt}x = f(x, u) \\ y = h(x) \end{cases} \quad (4.2)$$

avec $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ et $y \in \mathbb{R}^p$, les fonctions f et h étant régulières.

4.2.1 Définition

Pour définir l'observabilité, il convient d'abord de définir la notion de distinguabilité.

Définition 16 (Distinguabilité). Deux états initiaux x et \tilde{x} sont dits indistinguables (notés $x \sim \tilde{x}$) si pour tout $t \geq 0$, les sorties $y(t)$ et $\tilde{y}(t)$ sont identiques pour toute entrée $u(t)$ admissible². Ils sont dits distinguables sinon.

L'indistinguabilité est une relation d'équivalence. Notons $I(x)$ la classe d'équivalence de x . L'observabilité est alors définie de la manière suivante

Définition 17 (Observabilité globale). Le système (4.2) est dit observable en x si $I(x) = \{x\}$ et il est observable si $I(x) = \{x\}$ pour tout x .

En fait, le système est observable si pour tous les états initiaux x et \tilde{x} , il existe une entrée admissible u qui distingue x et \tilde{x} , c'est à dire telle que $y(t) \neq \tilde{y}(t)$ pour au moins un temps $t \geq 0$.

Il peut exister des entrées qui ne distinguent pas certains points. Cependant, le système peut être malgré tout observable. Par exemple

$$\begin{cases} \frac{d}{dt}x_1 &= ux_2 \\ \frac{d}{dt}x_2 &= 0 \\ y &= x_1 \end{cases}$$

est observable (pour $u = 1$ par exemple). Cependant l'entrée $u = 0$ ne distingue pas les points x et \tilde{x} tels que $x_1 = \tilde{x}_1$ et $x_2 \neq \tilde{x}_2$. Notons que l'observabilité ne signifie pas que toute entrée distingue tous les états. L'observabilité est un concept global. Il peut être nécessaire d'aller très loin dans le temps et dans l'espace d'état pour distinguer deux états initiaux. Pour cela nous introduisons le concept plus fort qui suit.

Définition 18 (Observabilité locale en temps et en espace). L'état x de (4.2) est localement observable, si pour tout $\varepsilon > 0$ et pour tout voisinage U de x , il existe $\eta > 0$ plus petit que ε et un voisinage V de x contenu dans U , tel que pour tout $\tilde{x} \in V$, il existe une entrée $[0, \eta] \ni t \mapsto u(t)$ qui distingue x et \tilde{x} , i.e. telle que $y(\eta) \neq \tilde{y}(\eta)$. Le système (4.2) est localement observable s'il l'est pour tout x .

Intuitivement, le système (4.2) est localement observable si on peut instantanément distinguer chaque état de ses voisins en choisissant judicieusement l'entrée u .

4.2.2 Critère

La seule façon effective de tester l'observabilité d'un système est de considérer l'application qui à x associe y et ses dérivées en temps. Nous supposerons dans cette section que y et u sont des fonctions régulières du temps. Nous supposerons également que les rangs en x des fonctions de $(x, u, \frac{d}{dt}u, \dots)$ qui apparaissent ci-dessous sont constants.

Considérons donc (4.2). On note $h_0(x) := h(x)$. En dérivant y par rapport au temps on a

$$\frac{d}{dt}y = D_x h(x) \frac{d}{dt}x = D_x h(x) \cdot f(x, u) := h_1(x, u)$$

Des dérivations successives conduisent donc à une suite de fonctions

$$h_k(x, u, \dots, u^{(k-1)})$$

2. $y(t)$ (resp. $\tilde{y}(t)$) correspond à la sortie de (4.2) avec l'entrée $u(t)$ et la condition initiale x (resp. \tilde{x}).

définie par la récurrence

$$h_{k+1} = \frac{d}{dt}(h_k), \quad h_0(x) = h(x)$$

Si pour un certain k , le rang en x du système

$$\left\{ \begin{array}{l} h_0(x) = y \\ h_1(x, u) = \frac{d}{dt}y \\ \vdots \\ h_k(x, u, \dots, u^{(k-1)}) = y^{(k)} \end{array} \right.$$

vaut $n = \dim(x)$ alors le système est *localement observable*. Il suffit d'utiliser le théorème d'inversion locale pour calculer x en fonction de $(y, \dots, y^{(k)})$ et $(u, \dots, u^{(k-1)})$. Si à partir d'un certain k , h_{k+1} ne fait plus apparaître de nouvelle relation en x , i.e., si le rang en x de $(h_0, \dots, h_k)'$ est identique à celui de $(h_0, \dots, h_k, h_{k+1})'$, alors il en est de même pour $k+2, k+3, \dots$. Ainsi, il n'est pas nécessaire de dériver plus de $n-1$ fois y pour savoir si un système est localement observable ou non. Ce raisonnement est valable autour d'un état générique, nous ne traitons pas les singularités qui peuvent apparaître en des états et entrées particulières. Nous renvoyons à [27] pour les cas plus généraux avec singularités.

Ce calcul élémentaire montre aussi que y et u sont reliés par des équations différentielles. Elles correspondent aux relations de compatibilité associées au système sur-déterminé (4.2) où l'inconnue est x et les données sont u et y . On obtient toutes les relations possibles en éliminant x du système

$$\left\{ \begin{array}{l} h_0(x) = y \\ h_1(x, u) = \frac{d}{dt}y \\ \vdots \\ h_n(x, u, \dots, u^{(n-1)}) = y^{(n)} \end{array} \right.$$

On peut montrer que pour un système localement observable, u et y sont reliés par $p = \dim(y)$ équations différentielles indépendantes. Ces équations font intervenir y dérivé au plus n fois et u dérivé au plus $n-1$ fois.

La mise en forme des idées précédentes est assez fastidieuse mais néanmoins instructive. Nous nous contenterons de retenir qu'en général l'observabilité signifie que *l'état peut être exprimé en fonction des sorties, des entrées et d'un nombre fini de leur dérivées en temps*. Dans ce cas, y et u sont reliés par p équations différentielles d'ordre au plus n en y et $n-1$ en u .

Exemple 26. Pour conclure, reprenons l'exemple du réacteur chimique (3.6) afin d'illustrer l'analyse formelle précédente. Nous ne considérons que x_1 et T car l'invariant chimique $x_1 + x_2$ est supposé égal à x_1^{in} . Nous supposons que la température T est mesurée (thermo-couple) mais pas la concentration x_1 . Nous avons donc à résoudre le système sur-déterminé (les quantités autres que (x_1, u, y, T) sont des constantes connues)

$$\begin{aligned} \frac{dx_1}{dt} &= D(x_1^{in} - x_1) - k_0 \exp(-E/RT)x_1 \\ \frac{dT}{dt} &= D(T^{in} - T) + \alpha \Delta H \exp(-E/RT)x_1 + u \\ y(t) &= T \end{aligned}$$

On a facilement x_1 en fonction de $(y, \frac{d}{dt}y)$ et u

$$x_1 = \frac{\frac{d}{dt}y - D(T^{in} - y) - u}{\alpha\Delta H \exp(-E/Ry)} \quad (4.3)$$

Le système est donc observable. y et u sont reliés par une équation différentielle du second ordre en y et du premier ordre en u . On l'obtient en utilisant l'équation donnant $\frac{d}{dt}x_1$

$$\begin{aligned} & \frac{d}{dt} \left(\frac{\frac{d}{dt}y - D(T^{in} - y) - u}{\alpha\Delta H \exp(-E/Ry)} \right) \\ &= Dx_1^{in} - (D + k_0 \exp(-E/Ry)) \frac{\frac{d}{dt}y - D(T^{in} - y) - u}{\alpha\Delta H \exp(-E/Ry)} \end{aligned} \quad (4.4)$$

Il s'agit d'une condition de compatibilité entre y et u . Si elle n'est pas satisfaite alors le système sur-déterminé de départ n'admet pas de solution. On conçoit très bien que ces relations de compatibilité sont à la base du diagnostic et de la détection de panne.

4.2.3 Observateur, estimation, moindres carrés

Savoir que le système est observable est bien. Calculer x à partir de y et u est encore mieux. Cependant, la démarche formelle précédente ne répond en pratique qu'à la première question. En effet, avoir x en fonction de dérivées des mesures s'avère d'une utilité fort limitée dès que l'ordre de dérivation dépasse 2 et/ou dès que les signaux sont bruités. Il convient en fait de calculer x en fonction d'intégrales de y et u . Dans ce cas, le bruit sur les signaux est beaucoup moins gênant. La synthèse d'observateur pose des problèmes supplémentaires (et nettement plus difficiles en fait) que la caractérisation des systèmes observables.

Revenons à (4.2). Nous avons un nombre infini d'équations en trop. En effet, puisque l'entrée u est connue, l'état x est entièrement donné par sa condition initiale x^0 grâce au flot ϕ_t^u de $\frac{d}{dt}x = f(x, u(t))$: $\phi_t^u(x^0)$ est la solution de $\frac{d}{dt}x = f(x, u(t))$ qui démarre en x^0 à $t = 0$. Ainsi x^0 vérifie à chaque instant t , p équations, p étant donc le nombre de mesures

$$y(t) = h(\phi_t^u(x^0))$$

Il est très tentant de résoudre ce système par les moindres carrés, même si, pour un système non-linéaire cela n'a pas beaucoup de sens. Considérons un intervalle d'observation $[0, T]$. x^0 peut être calculé comme l'argument du minimum de

$$J(\xi) = \int_0^T (y(t) - h(\phi_t^u(x^0)))^2 dt$$

x^0 est ainsi obtenu comme on obtient un paramètre à partir de données expérimentales et d'un modèle où ce paramètre intervient : en minimisant l'erreur quadratique entre l'observation $y(t)$ et la valeur prédictive par le modèle $\phi_t^u(x^0)$. Les problèmes d'observateurs sont fondamentalement proches des problèmes d'estimation pour lesquels l'optimisation joue un rôle important. Cependant, les difficultés ne sont pas pour autant aplanies : la résolution de l'équation différentielle $\frac{d}{dt}x = f(x, u)$ ne peut se faire que numériquement en général ; la fonction J n'a aucune raison d'avoir les propriétés de convexité qui assurent la convergence des principaux algorithmes d'optimisation (voir par exemple [63]). La

synthèse d'observateurs reste donc une question difficile en général bien que très importante en pratique. Noter enfin que l'identification de paramètres θ sur un modèle $\frac{d}{dt}x = f(x, u, \theta)$ en est un sous-problème : l'identifiabilité correspond alors à l'observabilité du système étendu

$$\frac{d}{dt}x = f(x, u, \theta), \quad \frac{d}{dt}\theta = 0, \quad y = x$$

d'état (x, θ) et de sortie $y = x$.

Dans le cas linéaire, $f = Ax + Bu$ et $h = Cx$, $\phi_t^u(x^0)$ est une fonction affine en x^0

$$\phi_t^u(x^0) = \exp(tA)x^0 + \int_0^t \exp((t-s)A)Bu(s) ds$$

À partir d'un intervalle d'observations $[0, T]$, x^0 peut être calculé comme l'argument du minimum de

$$J(\xi) = \int_0^T (z(t) - C \exp(tA)x^0)^2 dt \quad (4.5)$$

où $z(t) = y(t) - C \int_0^t \exp((t-s)A)Bu(s) ds$. Dans ces conditions J est quadratique. On est en train de retrouver, dans un cadre déterministe pour les moindres carrés, le filtre de Kalman traité dans la Section 4.5. Nous allons maintenant aborder l'observabilité des systèmes linéaires avec un point de vue moins classique qui met l'accent sur les observateurs asymptotiques. Ces derniers fournissent, avec des calculs très économiques, directement $x = \phi_t^u(x^0)$ en fonction de y, u et leurs intégrales.

4.3 Observabilité linéaire

On considère ici le système, d'entrée u , d'état x et de sortie y suivant

$$\frac{d}{dt}x = Ax + Bu, \quad y = Cx \quad (4.6)$$

où A est une matrice $n \times n$, B une matrice $n \times m$ et C une matrice $p \times n$.

4.3.1 Le critère de Kalman

Théorème 28 (Critère d'observabilité de Kalman)

Le système (4.6) $\frac{d}{dt}x = Ax + Bu$, $y = Cx$ est observable au sens de la Définition 18 si et seulement si le rang de la *matrice d'observabilité*

$$\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}$$

est égal à $n = \dim(x)$.

Pour abréger, on dit souvent que *la paire* (A, C) est observable lorsque le rang de la matrice d'observabilité \mathcal{O} est maximum. On notera la dualité avec le critère de commandabilité du Théorème 22.

Démonstration. Dérivons y et d'utilisons l'équation d'état. Une première dérivation donne

$$\frac{d}{dt}y = C \frac{d}{dt}x = CAx + CBu.$$

Donc x est nécessairement solution du système (les fonctions y et u sont connues)

$$\begin{aligned} Cx &= y \\ CAx &= \frac{d}{dt}y - CBu. \end{aligned}$$

À ce niveau, tout se passe comme si la quantité $\bar{y}_1 = \frac{d}{dt}y - CBu$ était une nouvelle sortie. En la dérivant de nouveau, nous avons $CA^2x = \dot{\bar{y}}_1 - CABu$. Maintenant, x est nécessairement solution du système étendu

$$\begin{aligned} Cx &= \bar{y}_0 = y \\ CAx &= \bar{y}_1 = \frac{d}{dt}y - CBu \\ CA^2x &= \bar{y}_2 = \dot{\bar{y}}_1 - CABu. \end{aligned}$$

Il est alors facile de voir que, ainsi de suite, x est nécessairement solution des équations

$$CA^kx = \bar{y}_k \tag{4.7}$$

où les quantités connues \bar{y}_k sont définies par la récurrence $\bar{y}_k = \frac{d}{dt}\bar{y}_{k-1} - CA^{k-1}Bu$ pour $k \geq 1$ et $\bar{y}_0 = y$.

Si le rang de la matrice d'observabilité est maximum et égal à n , elle admet un inverse à gauche (non nécessairement unique), P matrice $n \times pn$ vérifiant

$$P \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} = 1_n$$

Ainsi,

$$x = P \begin{pmatrix} \bar{y}_0 \\ \vdots \\ \bar{y}_{n-1} \end{pmatrix}$$

La condition de rang est donc suffisante.

Supposons maintenant que la matrice d'observabilité, de taille $pn \times n$, soit de rang $r < n$. Nous allons montrer qu'il existe, au moins, deux trajectoires différentes avec les mêmes commandes, donnant la même sortie. Cela montrera que la condition est aussi nécessaire.

Soit $w \in \mathbb{R}^n$ un élément non nul du noyau de la matrice d'observabilité. Pour $k = 0, \dots, n-1$, $CA^k w = 0$. Par un raisonnement identique à celui fait lors de la preuve de la Proposition 8 avec les noyaux à gauche de $A^k B$, on a nécessairement $CA^k w = 0$, pour toute $k \geq n$. Donc, w est dans le noyau de toutes les matrices CA^k . Prenons comme première trajectoire $[0, T] \ni t \mapsto (x, u) = 0$. Il vient, $y = 0$. Prenons maintenant comme seconde trajectoire, celle qui, à commande nulle, démarre en w : $[0, T] \ni t \mapsto (x, u) = (\exp(tA)w, 0)$. Sa sortie vaut

$$C \exp(tA)w = \sum_{i=0}^{+\infty} \frac{t^i}{i!} CA^i w = 0$$

car chaque terme de la série est nul. Ceci conclut la preuve. \square

4.3.2 Observateurs asymptotiques

Comme on l'a déjà dit, il est classique de noter par $\hat{x}(t)$ une estimation de la quantité $x(t)$. Nous cherchons ici à obtenir une estimation de l'état sans utiliser les dérivées de y et u . La première idée qui vient à l'esprit est de copier la dynamique du système. On intègre directement

$$\frac{d}{dt}\hat{x} = A\hat{x} + Bu$$

à partir d'une condition initiale \hat{x}^0 et des valeurs connues du contrôle u à chaque instant. Si la matrice A est stable, alors \hat{x} peut être pris comme estimation de x car l'erreur $e_x = \hat{x} - x$ tend naturellement vers 0 puisque $\frac{d}{dt}e_x = Ae_x$.

Si A est instable cette méthode ne marchera pas. En effet, une petite erreur initiale $e_x(0)$ sera amplifiée exponentiellement. Intuitivement, si l'erreur $\hat{x} - x$ devient grande alors, le système étant observable, l'erreur sur les sorties $\hat{y} - y$ deviendra grande également³. Comme y est connue, il est alors tentant de modifier $\frac{d}{dt}\hat{x} = A\hat{x} + Bu$ par l'ajout d'un terme du type $-L(\hat{y} - y)$ qu'on connaît et qui correspond à l'erreur d'observation. Le problème suivant se pose : peut-on choisir la matrice L de façon à ce que la solution \hat{x} du système

$$\frac{d}{dt}\hat{x} = A\hat{x} + Bu(t) - L(\hat{y} - y(t)), \quad \hat{y} = C\hat{x}$$

converge vers x ? Puisque $y = Cx$, la question se pose ainsi : peut-on ajuster la matrice L de façon à obtenir une équation différentielle d'erreur stable

$$\frac{d}{dt}e_x = (A - LC)e_x ?$$

Soyons plus exigeants encore, par un choix judicieux de L , peut-on imposer à $A - LC$ d'avoir toutes ses valeurs propres à partie réelle strictement négative?

Or, les valeurs propres restent inchangées par la transposition : $A - LC$ admet le même spectre que $A^T - C^T L^T$. De plus la paire (A, C) est observable si, et seulement si, la paire (A^T, C^T) est commandable (comme on l'a déjà évoqué, on obtient le critère de Kalman de commandabilité en transposant celui de l'observabilité). Ainsi le théorème 24 se transpose de la manière suivante.

Théorème 29 (Placement de pôles. Observateur asymptotique)

Si (A, C) est observable, il existe L , matrice $n \times p$, telle que le spectre de $A - LC$ soit le même que celui de n'importe quelle matrice réelle $n \times n$ librement choisie.

4.3.3 Observateurs réduits de Luenberger

Supposons que C soit de rang maximum $p = \dim(y)$ et que la paire (A, C) soit observable. On peut toujours supposer, quitte à faire un changement de variable sur x , que y correspond aux p premières composantes de l'état x : $x = (y, x_r)$. L'équation d'état $\frac{d}{dt}x = Ax + Bu$ s'écrit alors sous forme blocs

$$\begin{aligned} \frac{d}{dt}y &= A_{yy}y + A_{yrr}x_r + B_yu \\ \frac{d}{dt}x_r &= A_{ry}y + A_{rr}x_r + B_ru \end{aligned}$$

3. On a noté $\hat{y} = C\hat{x}$.

Il est facile de montrer, en revenant, par exemple à la définition de l'observabilité, que (A, C) est observable si, et seulement si, (A_{rr}, A_{yr}) l'est : en effet connaître y et u implique la connaissance de $A_{yr}x_r = \frac{d}{dt}y - A_{yy}y - B_yu$, qui peut être vue comme une sortie du système $\frac{d}{dt}x_r = A_{rr}x_r + (B_ru + A_{ry}y)$.

En ajustant correctement la matrice des gains d'observation L_r , le spectre de $A_{rr} - L_r A_{yr}$ coïncide avec celui de n'importe quelle matrice réelle carrée d'ordre $n - p = \dim(x_r)$. Considérons alors la variable $\xi = x_r - L_r y$ au lieu de x_r . Un simple calcul montre que

$$\frac{d}{dt}\xi = (A_{rr} - L_r A_{yr})\xi + (A_{ry} - L_r A_{yy} - (A_{rr} - L_r A_{yr})L_r)y + (B_r - L_r B_y)u$$

Ainsi en choisissant L_r , de façon à avoir $A_{rr} - L_r A_{yr}$ stable, nous obtenons un observateur d'ordre réduit $n - p$ pour ξ (donc pour $x_r = \xi + L_r y$) en recopiant cette équation différentielle

$$\frac{d}{dt}\hat{\xi} = (A_{rr} - L_r A_{yr})\hat{\xi} + (A_{ry} - L_r A_{yy} - (A_{rr} - L_r A_{yr})L_r)y(t) + (B_r - L_r B_y)u(t)$$

La dynamique de l'erreur sur ξ , $e_\xi = \hat{\xi} - \xi$ vérifie l'équation autonome stable

$$\frac{d}{dt}e_\xi = (A_{rr} - L_r A_{yr})e_\xi.$$

Cet observateur réduit est intéressant lorsque $n - p$ est petit, typiquement $n - p = 1, 2$: la stabilité d'un système de dimension 1 ou 2 est en outre très simple à étudier.

4.4 Observateur-contrôleur

4.4.1 Version état multi-entrée multi-sortie (MIMO)⁴

En regroupant les résultats sur la commandabilité et l'observabilité linéaires, nous savons comment résoudre de façon robuste par rapport à de petites erreurs de modèle et de mesures, le problème suivant : amener, à l'aide de la commande u , l'état x du système de p à q pendant le temps T en ne mesurant que y sachant que : $\frac{d}{dt}x = Ax + Bu$, $y = Cx$, (A, B) commandable et (A, C) observable.

En effet, comme (A, B) est commandable, nous savons avec la *forme de Brunovsky* construire explicitement une trajectoire de référence $[0, T] \ni t \mapsto (x_r(t), u_r(t))$ pour aller de p à q . Le respect de certaines contraintes peut-être important à ce niveau et être un guide dans le choix de cette trajectoire de référence (et intervenir dans la définition du critère pour la commande optimale).

Toujours à cause de la commandabilité, nous savons construire un bouclage statique sur $x, -Kx$, de façon à ce que la matrice $A - BK$ soit stable (placement de pôle). La matrice K est souvent appelée *matrice des gains de la commande*.

Grâce à l'observabilité, nous savons construire un observateur asymptotique sur x en choisissant les *gains d'observation* L de façon à avoir $A - LC$ stable.

Alors le bouclage dynamique de sortie

$$\begin{aligned} u(t) &= u_r(t) - K(\hat{x} - x_r(t)) && \text{contrôleur} \\ \frac{d}{dt}\hat{x} &= A\hat{x} + Bu(t) - L(C\hat{x} - y(t)) && \text{observateur} \end{aligned}$$

assure le suivi asymptotique de la trajectoire de référence $[0, T] \ni t \mapsto (x_r(t), u_r(t))$. Avec ce bouclage, appelé *commande modale* ou encore *observateur-contrôleur*, les petites erreurs de conditions

4. MIMO : pour Multi-Input Multi-Output

initiales sont amorties lorsque t croît et les petites erreurs de modèle et de mesures ne sont pas amplifiées au cours du temps.

En effet, comme $\frac{d}{dt}x = Ax + Bu$ et $y = Cx$, on a pour la dynamique du système bouclé

$$\begin{aligned}\frac{d}{dt}x &= Ax + B(u_r(t) - K(\hat{x} - x_r(t))) \\ \frac{d}{dt}\hat{x} &= A\hat{x} + B(u_r(t) - K(\hat{x} - x_r(t))) - L(C\hat{x} - Cx)\end{aligned}$$

où l'état est maintenant (x, \hat{x}) . Comme (x_r, u_r) est une trajectoire du système, $\frac{d}{dt}x_r = Ax_r + Bu_r$, on a en prenant comme variables d'état $(\Delta x = x - x_r(t), e_x = \hat{x} - x)$ au lieu de (x, \hat{x}) , la forme triangulaire suivante :

$$\begin{aligned}\frac{d}{dt}(\Delta x) &= (A - BK)\Delta x - BK e_x \\ \frac{d}{dt}e_x &= (A - LC)e_x\end{aligned}$$

Ce qui montre que e_x et Δx tendent vers 0 exponentiellement en temps. Le spectre du système d'état $(\Delta x, e_x)$ se compose du spectre de $A - LC$ et de celui de $A - BK$. C'est le *principe de séparation*.

4.4.2 Version transfert mono-entrée mono-sortie (SISO)⁵

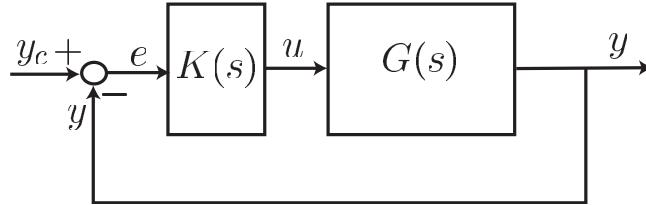


FIGURE 4.1 – Le transfert en boucle fermée, $e = y_c - y = \frac{1}{1+GK}y_c$.

Dans le cas mono-variable où y et u sont de dimension un, il est instructif d'interpréter l'observateur contrôleur ci-dessus sous sa version état, en terme de transfert. On part donc du transfert rationnel $y = G(s)u$, avec $G(s) = \frac{P(s)}{Q(s)}$ strictement causal ($d^0 P < d^0 Q$), P et Q premiers entre eux (pas de racines communes dans \mathbb{C}). On cherche, comme illustré sur la figure 4.1, un contrôleur $K(s)$ aussi sous la forme d'un transfert rationnel causal, $K(s) = \frac{N(s)}{D(s)}$ ($d^0 N < d^0 D$), de sorte que le système bouclé soit stable. Nous allons voir que l'observateur-contrôleur répond à la question. Pour cela on passe à la forme d'état. Pour éviter tout confusion avec les fonctions de transfert, les matrices sont notées en caractères calligraphiques.

Il existe trois matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ telles que $\frac{d}{dt}x = \mathcal{A}x + \mathcal{B}u$, $y = \mathcal{C}x$ avec $(\mathcal{A}, \mathcal{B})$ commandable, $(\mathcal{A}, \mathcal{C})$ observable et $G(s) = \mathcal{C}(s\mathcal{I} - \mathcal{A})^{-1}\mathcal{B}$. L'observateur contrôleur

$$u = -\mathcal{K}\hat{x}, \quad \frac{d}{dt}\hat{x} = \mathcal{A}\hat{x} + \mathcal{B}u - \mathcal{L}(\mathcal{C}\hat{x} - (y - y_c))$$

correspond au transfert $u = K(s)(y_c - y)$ donné par

$$K(s) = \mathcal{K}(s\mathcal{I} - \mathcal{A} + \mathcal{B}\mathcal{K} + \mathcal{L}\mathcal{C})^{-1}\mathcal{L}.$$

5. SISO : pour Single-Input Single-Output

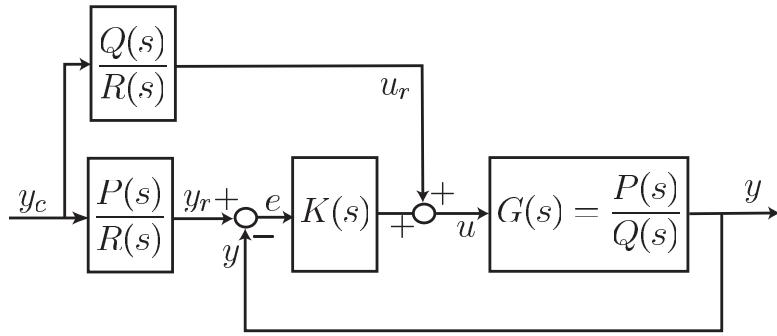


FIGURE 4.2 – Boucle fermée avec générations de trajectoires par un filtre passe bas sur la consigne y_c et dont l'ordre est au moins égal à celui du système ($d^0 R \geq d^0 Q$, R stable).

L'identité matricielle suivante⁶

$$[\mathcal{I} - \mathcal{C}\mathcal{M}^{-1}\mathcal{N}(\mathcal{M} + \mathcal{N} + \mathcal{LC})^{-1}\mathcal{L}]^{-1} = \mathcal{I} + \mathcal{C}(\mathcal{M} + \mathcal{N})^{-1}\mathcal{N}(\mathcal{M} + \mathcal{LC})^{-1}\mathcal{L}.$$

conduit, avec $\mathcal{M} = s\mathcal{I} - \mathcal{A}$, $\mathcal{N} = \mathcal{B}\mathcal{K}$, à la formule explicite suivante :

$$\frac{1}{1 + G(s)K(s)} = 1 - \mathcal{C}(s\mathcal{I} - \mathcal{A} + \mathcal{B}\mathcal{K})^{-1}\mathcal{B}\mathcal{K}(s\mathcal{I} - \mathcal{A} + \mathcal{LC})^{-1}\mathcal{L}$$

Ainsi le transfert en boucle fermée admet comme pôles ceux du contrôleur et de l'observateur, pôles que l'on peut choisir où l'on veut puisque $(\mathcal{A}, \mathcal{B})$ est commandable et $(\mathcal{A}, \mathcal{C})$ est observable.

Le gain statique en boucle fermée $\frac{1}{1+G(0)K(0)} = 1 - \mathcal{C}(\mathcal{A} - \mathcal{B}\mathcal{K})^{-1}\mathcal{B}\mathcal{K}(\mathcal{A} - \mathcal{LC})^{-1}\mathcal{L}$ entre $e = y_c - y$ et y_c n'a aucune raison d'être nul. Pour avoir cela, soit on rajoute un effet intégral lent (recalage), soit on rajoute à u , u_c vérifiant $Q(0)y_c = P(0)u_c$.

On peut aussi générer en temps réel et de façon causale des références y_r et u_r qui vérifient identiquement $Q(s)y_r = P(s)u_r$ et qui suivent asymptotiquement n'importe quelle consigne vérifiant $Q(0)y_c = P(0)u_c$. Il suffit de poser

$$y_r = \frac{P(s)}{R(s)}y_c, \quad u_r = \frac{Q(s)}{R(s)}y_c$$

avec R polynôme stable, $d^0 R \geq d^0 Q$ et $R(0) = 1$. Ainsi, $1/R(s)$ est un filtre passe-bas de gain statique unitaire, et on considère le contrôleur

$$u = u_r - K(s)(y - y_r).$$

qui donne comme transfert en boucle fermée $y = y_r = \frac{Q(s)}{R(s)}y_c$. On parle alors de filtrage de la consigne y_c . Ces petits calculs de génération et suivi de trajectoires correspondent au schéma bloc fonctionnel de la figure 4.2. Si $Q(0) \neq 0$, alors, avec le changement d'échelle $y_c \mapsto y_c/Q(0)$ on voit que y_r est bien une valeur filtrée de y_c .

4.5 Filtre de Kalman

Nous proposons ici une présentation simplifiée de la technique du filtrage de Kalman. On se restreint au cas des filtres à gains constants. Le filtre de Kalman à gains constants, tel que nous le présentons ici, est un algorithme utilisé pour reconstruire les variables d'état d'un système continu soumis

6. Cette identité est un excellent exercice de calcul dans un cadre non commutatif.

à des perturbations stochastiques en utilisant des mesures incomplètes entâchées de bruit. Cette technique (voir [40]) a été utilisée de manière intensive dans le domaine de la navigation inertielle [23]. Il en existe de nombreuses variantes : version instationnaire, version discrète, version étendue pour tenir compte de la variation du point de fonctionnement, etc. On pourra se reporter à [61] pour une présentation complète et à [9] pour des compléments au sujet de la robustesse dans le cadre stationnaire.

Le lien avec les observateurs asymptotiques de la section précédente est simple : ces derniers peuvent être interprétés comme des filtres de Kalman où les matrices de covariance des bruits sur l'état et la sortie paramétrisent la matrice des gains L du théorème 29. Au lieu de définir L à partir des pôles de l'observateur, i.e., des valeurs propres de $A - LC$, L est défini à partir des matrices de covariance. Comme pour la commande LQR, il difficile d'avoir alors des formules analytiques pour L sauf pour les systèmes à un ou deux états⁷.

4.5.1 Formalisme

On considère un système sous forme d'état linéaire stationnaire

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) + D\xi(t) \quad (4.8)$$

où l'état est $x(t) \in \mathbb{R}^n$, l'entrée est un signal (déterministe) connu $u(t) \in \mathbb{R}^m$, et $\xi(t) \in \mathbb{R}^q$ représente la valeur d'un signal stochastique qui agit comme une *perturbation* sur le système. On suppose que ce dernier signal possède les propriétés statistiques suivantes (qui en font un *bruit blanc gaussien centré*) :

$$\begin{aligned} &\text{pour tout } t, \xi(t) \text{ est une variable gaussienne de moyenne } E(\xi(t)) = 0, \\ &\text{cov}(\xi(t), \xi(\tau)) = M_\xi \delta(t - \tau) \end{aligned}$$

où δ est la fonction Dirac (voir [44]), et M_ξ est une matrice constante de $\mathcal{M}_q(\mathbb{R})$ symétrique définie positive. On rappelle que

$$\text{cov}(\xi(t), \xi(\tau)) = E(\xi(t)\xi^T(\tau))$$

où $\xi(t)\xi^T(\tau)$ est la matrice carré $(\xi_i(t)\xi_j(\tau))_{1 \leq i,j \leq q}$.

Sous ces hypothèses, $x(t)$ solution de (4.8) est un processus stochastique gaussien *coloré*.

En ce qui concerne la mesure, nous disposons de

$$y(t) = Cx(t) + \varrho(t) \quad (4.9)$$

où $y(t) \in \mathbb{R}^p$, et $\varrho(t) \in \mathbb{R}^p$ représente lui aussi un *bruit blanc gaussien centré*, indépendant de $\xi(t)$ tel que

$$\begin{aligned} &E(\varrho(t)) = 0, \text{ pour tout } t \\ &\text{cov}(\varrho(t), \varrho(\tau)) = M_\varrho \delta(t - \tau) \end{aligned}$$

où M_ϱ est une matrice constante de $\mathcal{M}_p(\mathbb{R})$ symétrique définie positive.

On se place dans le cas où on connaît les entrées et les mesures jusqu'au temps t . On note alors

$$U(t) = \{u(\tau), \tau \in]-\infty, t]\}$$

7. Ce qui n'est pas le cas si L est défini en plaçant les valeurs propres de $A - LC$.

$$Y(t) = \{y(\tau), \tau \in]-\infty, t]\}$$

Étant donné $U(t)$ et $Y(t)$, on cherche une estimation $\hat{x}(t)$ de l'état $x(t)$ qui soit optimale en un certain sens. Le critère à minimiser est la covariance totale de l'erreur $\tilde{x}(t) = x(t) - \hat{x}(t)$

$$J = E(\tilde{x}^T(t)\tilde{x}(t)) = \text{trace}[E(\tilde{x}(t)\tilde{x}^T(t))]$$

sous la contrainte d'estimation *non biaisée*

$$E(\tilde{x}(t)) = 0$$

4.5.2 Hypothèses et définition du filtre

Hypothèses

Pour pouvoir construire le filtre de Kalman, on suppose que les deux hypothèses (fortes) sont vérifiées

- (H1) : la paire (A, D) est commandable
- (H2) : la paire (A, C) est observable

(H1) signifie que le bruit $\xi(t)$ excite tout l'état du système. (H2) implique que la mesure non bruitée Cx contient des informations sur tout l'état. On peut relâcher ces deux hypothèses en n'imposant seulement que

- (H1') : la paire (A, D) est stabilisable
- (H2') : la paire (A, C) est détectable

en imposant que les modes non excités par le bruit $\xi(t)$ soient asymptotiquement stables et que les modes non observés soient asymptotiquement stables.

Minimisation de la covariance

Le filtre de Kalman que nous présentons est un filtre stationnaire de dimension égale à celle de l'état du système à observer. Il admet comme entrée le signal $u(t)$ et la mesure bruitée $y(t)$. Il s'écrit, comme pour l'observateur asymptotique de Luenberger, au moyen d'une matrice de gain L sous la forme suivante

$$\frac{d}{dt}\hat{x}(t) = A\hat{x}(t) + Bu(t) - L(C\hat{x}(t) - y(t)) \quad (4.10)$$

Le terme $C\hat{x}(t) - y(t)$ est nommé *innovation*. La stabilité de ce filtre dépend de la stabilité de la matrice $A - LC$. D'après (H2), (ou (H2')) il est possible de stabiliser cette matrice par un choix approprié de L . La particularité du filtre de Kalman réside dans le calcul de L à partir des matrices de covariance sur les bruits ξ et ϱ .

Par définition de l'équation (4.10), l'équation satisfaite par l'erreur est

$$\frac{d}{dt}\tilde{x}(t) = (A - LC)\tilde{x}(t) + D\xi(t) - L\varrho(t)$$

En utilisant la *matrice de transition* $\phi(t, t_0) = \exp((A - LC)(t - t_0))$ entre un temps initial t_0 et t , on peut écrire la solution de ce système comme-suit

$$\tilde{x}(t) = \phi(t, t_0)x(t_0) + \int_{t_0}^t \phi(t, \tau)(D\xi(\tau) - L\varrho(\tau))(\tau)d\tau$$

En passant aux espérances, il vient

$$E(\tilde{x}(t)) = \phi(t, t_0)E(x(t_0))$$

car les signaux $\xi(t)$ et $\varrho(t)$ sont d'espérance nulle par hypothèse. Si $A - LC$ est asymptotiquement stable, on en déduit que lorsque le filtre de Kalman est initialisé au temps $t_0 = -\infty$, on a

$$E(\tilde{x}(t)) = 0$$

et on conclut que l'estimation est, quel que soit le choix de L (stabilisant), non biaisée⁸. Un calcul semblable permet de calculer la covariance de l'erreur, notée $\Sigma \triangleq E(\tilde{x}(t)\tilde{x}^T(t))$ comme l'unique solution symétrique positive de l'*équation matricielle de Lyapounov* suivante

$$(A - LC)\Sigma + \Sigma(A - LC)^T + DM_\xi D^T + LM_\varrho L^T = 0 \quad (4.11)$$

Ainsi, pour toute valeur de la matrice de gain L , on peut calculer la covariance de l'erreur par l'équation précédente. Le critère que nous cherchons à minimiser est justement $J = \text{trace}(\Sigma)$. Pour le filtre de Kalman, L est choisi comme l'argument du minimum de cette fonction.

Gain du filtre

Le gain du filtre de Kalman est

$$L = \Sigma C^T M_\varrho^{-1} \quad (4.12)$$

où Σ est l'unique solution symétrique positive de l'équation de Riccati algébrique

$$0 = A\Sigma + \Sigma A^T + DM_\xi D^T - \Sigma C^T (M_\varrho)^{-1} C\Sigma$$

On peut montrer, par une démonstration très semblable au cas de la commande quadratique, que le gain L est stabilisant, i.e. est tel que $A - LC$ a toutes ses valeurs propres à partie réelle strictement négative. Ceci justifie a posteriori les hypothèses que nous avons formulées.

Exemple 27. Considérons un système autonome mono-dimensionnel très simple

$$\frac{d}{dt}x = -x + u + \xi$$

$$y = x + \varrho$$

où $E(\xi(t)) = 0$, pour tout t , $\text{cov}(\xi(t), \xi(\tau)) = \delta(t - \tau)$, $E(\varrho(t)) = 0$, pour tout t , $\text{cov}(\varrho(t), \varrho(\tau)) = k^2\delta(t - \tau)$. Le filtre de Kalman est de la forme

$$\frac{d}{dt}\hat{x}(t) = -\hat{x} + u - L(\hat{x} - y(t))$$

8. Si on initialise le filtre de Kalman en $t = 0$ avec un état $x(0) = x_0$ quelconque, la stabilité permettra de conclure à l'absence de biais de manière asymptotique, lorsque $t \rightarrow +\infty$.

où L est calculé par

$$L = \Sigma(k^2)^{-1}$$

où Σ est l'unique solution positive de

$$0 = -2\Sigma + 1 - \Sigma^2 (k^2)^{-1}$$

Il vient $\Sigma = \sqrt{k^2 + k^4} - k^2$ et l'équation du filtre de Kalman est alors

$$\frac{d}{dt}\hat{x}(t) = -\sqrt{1 + \frac{1}{k^2}}\hat{x} + u + \left(\sqrt{1 + \frac{1}{k^2}} - 1\right)y$$

À titre d'illustration, on a représenté sur la figure 4.3 des résultats de simulation permettant de comprendre l'effet du filtre de Kalman. On peut comparer sur la même figure ses résultats avec un autre réglage non optimal.

Exemple 28 (Fusion de capteurs pour robot à roues). Considérons un robot mobile à roues indépendantes tel que le MobileRobots™ Pioneer 3-AT représenté sur la Figure 4.4. Pour contrôler ce véhicule, on peut librement agir sur les quatres roues indépendantes. Pour le rendre autonome, c.-à-d. capable de suivre une trajectoire prévue à l'avance par exemple, la principale difficulté consiste à estimer (à bord) précisément la position et l'orientation du véhicule. Pour résoudre ce problème, on peut équiper le véhicule d'un certain nombre de capteurs embarqués : odomètres, récepteur GPS, gyroscope entre autres.

Chaque capteur présente des défauts bien spécifiques. Le GPS fournit un signal actualisé à relativement basse fréquence, entâché d'un fort bruit de mesure basse fréquence. Le gyroscope fournit des informations fréquemment réactualisées mais dégradées par un fort bruit de mesure haute fréquence. Enfin, les odomètres sont relativement fiables, mais ont une résolution limitée.

Les équations de la dynamique du véhicule sont

$$\begin{cases} \frac{d}{dt}x = \frac{(v_1 + v_2)}{2} \cos \theta, & \frac{d}{dt}y = \frac{(v_1 + v_2)}{2} \sin \theta, \\ \frac{d}{dt}\theta = \frac{(v_2 - v_1)}{2l} \end{cases} \quad (4.13)$$

où v_1, v_2 représentent la vitesse du train roulant gauche et droit respectivement, x, y représentent la position du centre de masse du véhicule et θ est l'orientation du chassis, l est la demi largeur de l'engin. En conditions normales d'utilisation, le modèle (4.13) est fiable.

Dans ce modèle, on dispose de la connaissance de v_1 et v_2 par l'utilisation des odomètres (ici des capteurs incrémentaux situés dans les arbres des roues). On mesure x et y par le récepteur GPS et enfin on mesure $\frac{d}{dt}\theta$ grâce à un gyroscope⁹ à un axe.

Afin d'estimer l'état $(x, y, \theta)^T$ du système, on peut utiliser un observateur, ou un filtre de Kalman. Ces techniques de fusion de capteurs utilisent la redondance de l'information et l'accord avec un modèle dynamique, pour fournir une estimation meilleure que l'information fournie par chacun des capteurs séparément. Par exemple, si on intègre (en boucle ouverte) les données du gyroscope et des odomètres, ce qui théoriquement permet de calculer les positions, on observe très rapidement (après quelques dizaines de secondes) une importante dérive de l'estimation. Si on n'utilise que le GPS, on obtient assez rapidement des aberrations signalant par exemple des mouvements latéraux de plusieurs mètres (alors que le véhicule en est totalement incapable).

9. Ce gyroscope de technologie MEMS mesure une vitesse angulaire.

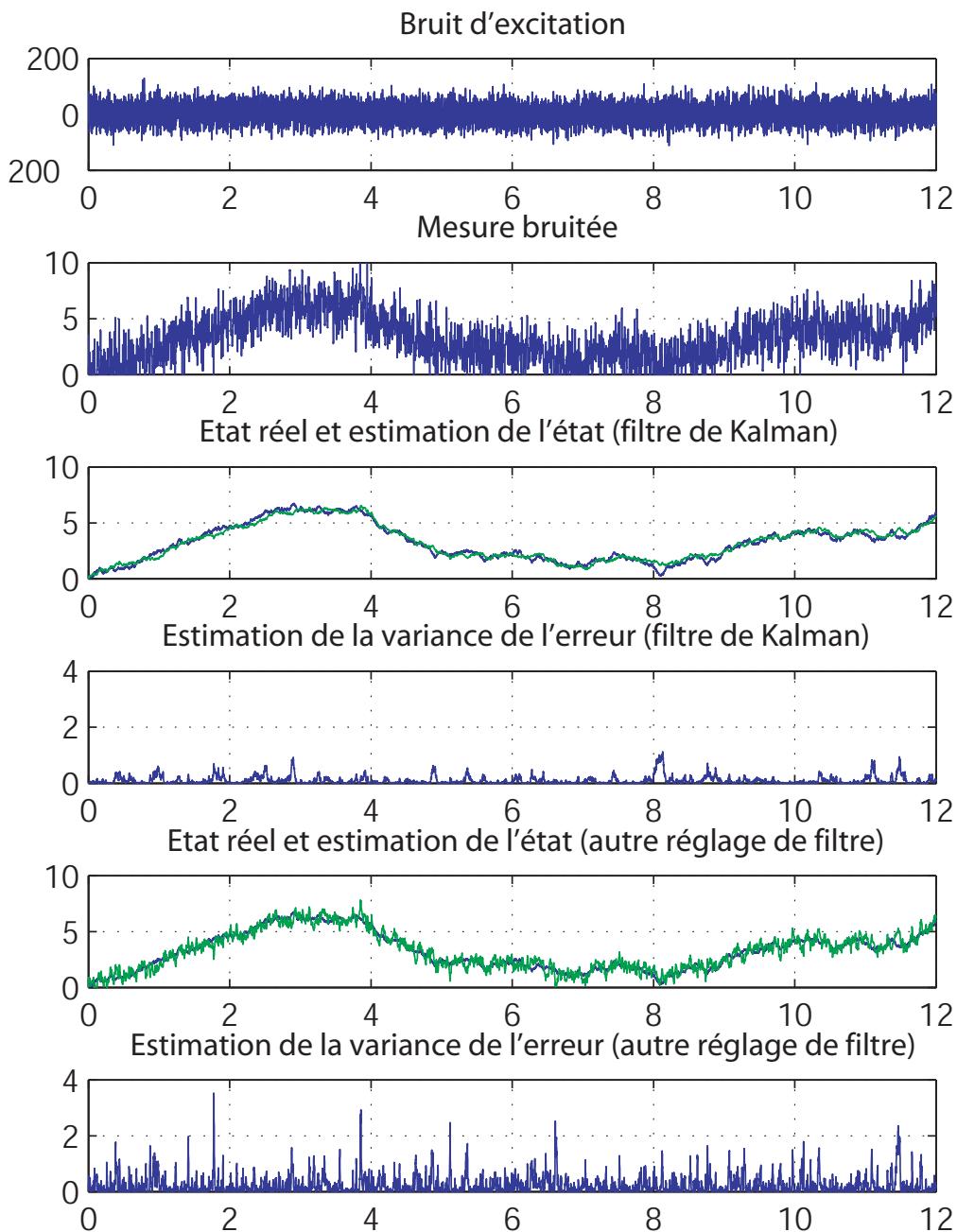


FIGURE 4.3 – Utilisation d'un filtre de Kalman.



FIGURE 4.4 – Véhicule autonome à roues indépendantes (remerciements au Laboratoire de recherches balistiques et aérodynamiques de la Délégation Générale pour l’Armement).

On a représenté sur la Figure 4.5, des résultats de suivi en boucle fermée d'une trajectoire de consigne. On donne aussi à titre informatif, l'enregistrement des valeurs données par le récepteur GPS qu'on pourra comparer à l'estimation par fusion de capteurs (on dit aussi hybridation) de la position du véhicule.

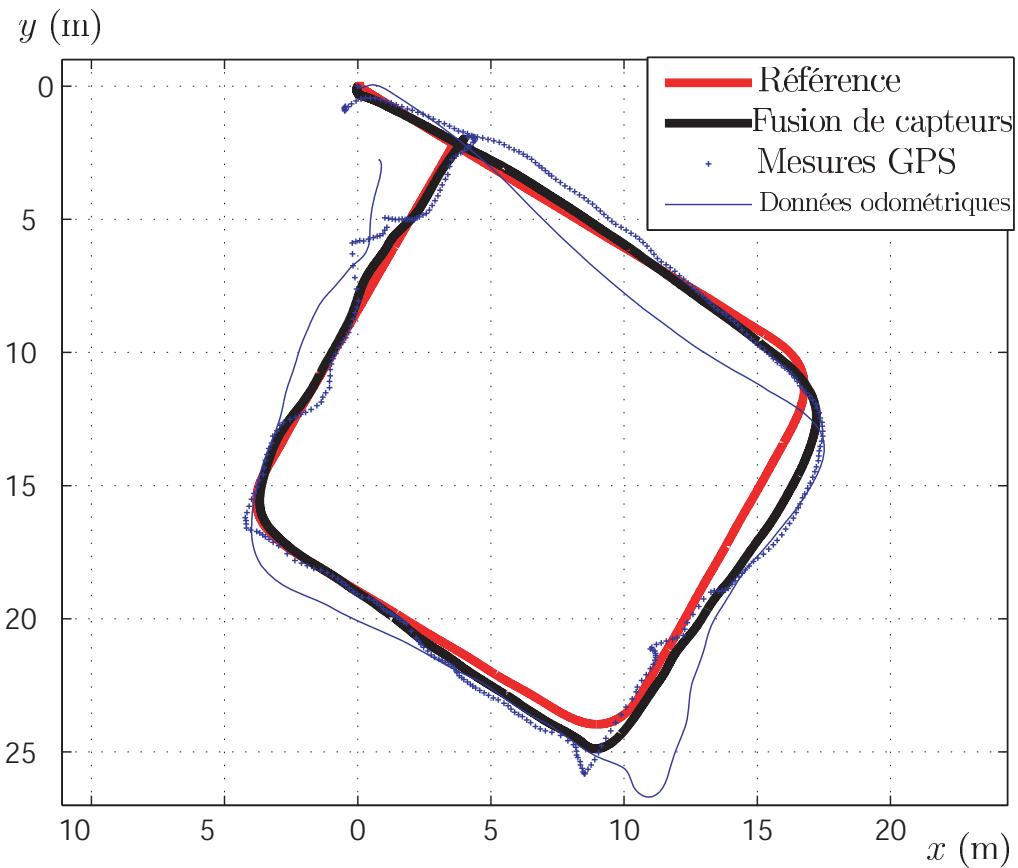


FIGURE 4.5 – Suivi de trajectoire en boucle fermée pour un véhicule à roues en utilisant une fusion de capteurs : GPS, gyroscope et odomètres.

Utilisation dans un cadre déterministe

Lorsque le système n'est pas affecté par des bruits ou que ceux-ci sont mal connus, on peut continuer à utiliser le filtre de Kalman pour régler l'observateur asymptotique stationnaire qu'il définit. Les matrices M_ϱ et M_ξ agissent comme des pondérations (on remarquera l'analogie avec la commande quadratique une fois de plus) avec les effets suivants :

1. lorsqu'on augmente M_ϱ , on a tendance à faire moins confiance aux mesures (ou à certaines d'entre elles si on agit seulement sur certains coefficients dominants de la matrice) et donc naturellement l'influence de ces mesures sur la dynamique du filtre sera réduite à travers le calcul de la matrice de gain (4.12) ;
2. lorsqu'on augmente M_ξ , on a tendance à faire moins confiance au modèle de la dynamique non bruitée, naturellement on va accélérer la dynamique de l'observateur pour en tenir compte en privilégiant dans (4.10) l'innovation aux dépens du modèle. Ceci est apparent dans la contribution de M_ξ au gain L à travers la covariance dans l'équation (4.11).

4.6 Compléments

4.6.1 Estimation de paramètres et commande adaptative

Soit le système d'état x et de paramètre p

$$\frac{d}{dt}x = f(x, t) + p g(x, t), \quad x \in \mathbb{R}^n, \quad p \in \mathbb{R}$$

On suppose que les trajectoires $t \mapsto x(t)$ sont bornées et donc définies pour tout temps t positif. On mesure ici tout l'état x . On souhaite estimer p .

On considère l'estimateur suivant où K et λ sont des paramètres constants > 0 (sorte de moindres carrés récursifs)

$$\frac{d}{dt}\hat{x} = f(x(t), t) + \hat{p} g(x(t), t) - K(\hat{x} - x(t)), \quad \frac{d}{dt}\hat{p} = -\lambda \langle g(x(t), t), (\hat{x} - x(t)) \rangle$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire dans \mathbb{R}^n .

Cet estimateur donne en temps réel d'une part une valeur filtrée de x , \hat{x} et d'autre part reconstruit asymptotiquement le paramètre p lorsqu'il existe $\alpha > 0$ tel que $\|g(x, t)\| \geq \alpha$ pour tout $x \in \mathbb{R}^n$ et tout temps t . En effet, il suffit de considérer la fonction de Lyapounov suivante

$$V = \frac{1}{2}(\hat{x} - x)^2 + \frac{1}{2\lambda}(\hat{p} - p)^2$$

On voit que

$$\frac{d}{dt}V = -K(\hat{x} - x)^2 \leq 0$$

Ainsi \hat{x} tend vers x quand t tend vers l'infini. Et donc, intuitivement $\frac{d}{dt}x - \frac{d}{dt}\hat{x}$ converge vers 0 soit donc $(p - \hat{p})g(x, t)$ converge zéro. Comme le vecteur $g(x, t)$ est toujours de norme plus grande que $\alpha > 0$ on en déduit que \hat{p} converge vers p .

Il est possible de généraliser cet estimateur pour un nombre arbitraire m de paramètres. On part de

$$\frac{d}{dt}x = f(x, t) + \sum_{i=1}^m p_i g_i(x, t)$$

On considère avec $\lambda_i > 0$ paramètre ($i = 1, \dots, m$) l'estimateur

$$\begin{aligned}\frac{d}{dt}\hat{x} &= f(x(t), t) + \sum_{i=1}^m \hat{p}_i g_i(x(t), t) - K(\hat{x} - x(t)) \\ \frac{d}{dt}\hat{p}_1 &= -\lambda_1 \langle g_1(x(t), t), (\hat{x} - x(t)) \rangle \\ &\vdots \\ \frac{d}{dt}\hat{p}_m &= -\lambda_m \langle g_m(x(t), t), (\hat{x} - x(t)) \rangle\end{aligned}$$

Il est alors facile de voir que la fonction

$$V = \frac{1}{2}(\hat{x} - x)^2 + \sum_{i=1}^m \frac{1}{2\lambda_i}(\hat{p}_i - p_i)^2$$

est décroissante

$$\frac{d}{dt}V = -K(\hat{x} - x)^2 \leq 0$$

Donc on a toujours \hat{x} qui converge vers x . En revanche, la convergence des \hat{p}_i vers les p_i n'est pas garantie : tout dépend des $g_i(x, t)$: s'ils forment pour chaque x et t un système libre de vecteurs de \mathbb{R}^n , système "bien conditionné", alors la convergence des paramètres est encore assurée. Dans les autres cas, cela dépend de la trajectoire suivie par x . En revanche, il est très intéressant de voir que x converge toujours vers \hat{x} . L'estimateur précédent est une sorte de filtre non linéaire et adaptatif de x .

Cet estimateur est aussi à la base de la *commande adaptative* qui estime p en même temps que l'on contrôle le système via u . Prenons un seul paramètre p et un seul contrôle u (la généralisation est simple à partir de là)

$$\frac{d}{dt}x = f_0(x) + u f_1(x) + p g(x)$$

Supposons que nous ayons un feedback $u = k(x, p)$ qui stabilise le système en $x = 0$ mais où le paramètre p inconnu intervient. Alors, il est naturel de considérer le feedback construit à partir des estimées $u = k(x, \hat{p})$ ou $u = k(\hat{x}, \hat{p})$ avec

$$\begin{aligned}\frac{d}{dt}\hat{x} &= f_0(x(t)) + u(t) f_1(x(t)) + \hat{p} g(x(t), t) - K(\hat{x} - x(t)) \\ \frac{d}{dt}\hat{p} &= -\lambda \langle g(x(t), t), (\hat{x} - x(t)) \rangle\end{aligned}$$

Sous des hypothèses raisonnables sur la dynamique en boucle ouverte pour x (pas d'explosion en temps fini essentiellement) on montre que \hat{x} tend vers x et ensuite que le feedback assure la convergence de x vers 0. Sans conditions supplémentaires, tout ce que l'on peut dire c'est que \hat{p} reste borné. Ainsi, la connaissance de p n'est pas nécessairement indispensable pour faire converger x vers 0 avec le contrôle u . C'est le principal paradoxe du contrôle adaptatif. Un lecteur intéressé pourra consulter [43]. On notera que l'une des hypothèses très importante est la dépendance affine de la dynamique par rapport au paramètre p . Dans le cas de dépendance non linéaire très peu de résultats existent.

4.6.2 Linéarisation par injection de sortie

Il arrive parfois que, dans les bonnes coordonnées d'état, coordonnées notées x ici, les équations du système s'écrivent ainsi :

$$\frac{d}{dt}x = Ax + f(Cx, t), \quad y = Cx$$

avec la paire (A, C) observable et $f(y, t)$ est une fonction a priori non linéaire. Il est clair qu'il faut avoir un peu de flair pour choisir les bonnes variables x pour que le système s'écrive ainsi. Ce n'est en général pas possible. Parfois c'est possible et même évident comme pour le pendule commandé :

$$\frac{d^2}{dt^2}\theta = -\frac{g}{l} \sin \theta + u(t), \quad y = \theta.$$

En effet dans ce cas on a

$$x = \begin{pmatrix} \theta \\ \frac{d}{dt}\theta \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \end{pmatrix},$$

$$f(Cx, t) = \begin{pmatrix} 0 \\ -\frac{g}{l} \sin \theta + u(t) \end{pmatrix}$$

Il est alors facile de construire un observateur pour x . Il suffit de choisir L telle que $A + LC$ soit une matrice réelle stable. Alors l'observateur

$$\frac{d}{dt}\hat{x} = Ax + L(C\hat{x} - y(t)) + f(y(t), t)$$

est exponentiellement convergent car lorsque l'on regarde la dynamique de l'erreur $e_x = \hat{x} - x$, les termes non linéaires disparaissent et on a

$$\frac{d}{dt}e_x = (A + LC)e_x.$$

4.6.3 Contraction

La notion de contraction [51, 33] pour un système, avec une dynamique $\dot{x} = f(x, t)$, peut être interprétée comme la décroissance exponentielle, avec le temps, de la longueur de tout segment de conditions initiales transporté par le flot.

Définition 19 (Contraction stricte). Soit un système dynamique $\frac{d}{dt}x = f(x, t)$ régulier (C^1 par exemple) défini sur une variété M régulière. Soit g une métrique sur M . Soit $U \subset M$ un sous ensemble de M . La dynamique f est dite strictement contractante sur U au sens de la métrique g , si la partie symétrique de sa matrice Jacobienne est une matrice définie négative, c'est à dire, s'il existe $\lambda > 0$ tel que, dans des coordonnées locales x sur U , nous avons pour tout t et pour tout x ,

$$\frac{\partial f^T}{\partial x} g(x) + g(x) \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x} f(x, t) \leq -\lambda g(x)$$

Nous avons le résultat suivant qui justifie cette définition et cette terminologie.

Théorème 30

Soit un système dynamique $\frac{d}{dt}x = f(x, t)$ régulier défini sur une variété M régulière. Soit $g(x)$ une métrique sur M . Soit $X(x, t)$ le flot associé à f

$$\begin{aligned}\frac{d}{dt}X(x, t) &= f(X(x, t), t) \quad \forall t \in [0, T[\quad \text{avec} \quad T \leq +\infty. \\ X(x, 0) &= x\end{aligned}$$

Considérons deux points x_0 and x_1 dans M et une géodésique $\gamma(s)$ qui joint $x_0 = \gamma(0)$ et $x_1 = \gamma(1)$. Si

- f est une stricte contraction sur un sous ensemble $U \subset M$, avec λ la constante présente dans la définition 19,
- et si $X(\gamma(s), t)$ appartient à U pour tout $s \in [0, 1]$ et pour tout $t \in [0, T[$.

alors

$$d_g(X(x_0, t), X(x_1, t)) \leq e^{-\frac{\lambda}{2}t} d_g(x_0, x_1) \quad \forall t \in [0, T[$$

où d_g est la distance géodésique associée à la métrique g .

Démonstration. La preuve est la suivante¹⁰. Soit $l(t)$ la longueur de la courbe $(X(\gamma(s), t), s \in [0, 1])$ au sens de la métrique g

$$l(t) = \int_0^1 \sqrt{\frac{dX(\gamma(s), t)}{ds}^T g(X(\gamma(s), t)) \frac{dX(\gamma(s), t)}{ds}} ds$$

Nous avons alors

$$\frac{d}{dt}l(t) = \int_0^1 \frac{\frac{d}{dt} \left\{ \frac{dX(\gamma(s), t)}{ds}^T g(X(\gamma(s), t)) \frac{dX(\gamma(s), t)}{ds} \right\}}{2 \sqrt{\frac{dX(\gamma(s), t)}{ds}^T g(X(\gamma(s), t)) \frac{dX(\gamma(s), t)}{ds}}} ds$$

Comme

$$\frac{d}{dt} \frac{dX(\gamma(s), t)}{ds} = \frac{d}{ds} f(X(\gamma(s), t), t) = \frac{\partial}{\partial X} f(X(\gamma(s), t), t) \frac{d}{ds} X(\gamma(s), t)$$

nous obtenons

$$\frac{d}{dt} \left\{ \frac{dX(\gamma(s), t)}{ds}^T g(X(\gamma(s), t)) \frac{dX(\gamma(s), t)}{ds} \right\} = \frac{dX(\gamma(s), t)}{ds}^T P(s, t) \frac{dX(\gamma(s), t)}{ds}$$

avec

$$P(s, t) = \frac{\partial f(X)}{\partial X}^T g(X) + g(X) \frac{\partial f(X)}{\partial X} + \frac{\partial g(X)}{\partial X} f(X, t)$$

et

$$X = X(\gamma(s), t)$$

Puisque f est une contraction sur U , il existe $\lambda > 0$ tel que

$$P(s, t) \leq -\lambda g(X(\gamma(s), t))$$

10. Cette preuve s'inspire de calculs faits par Laurent Praly

On obtient alors l'inégalité suivante pour la dérivée $\frac{d}{dt}l(t)$

$$\frac{d}{dt}l(t) \leq -\frac{\lambda}{2}l(t)$$

qui conduit à

$$l(t) \leq l(0)e^{-\frac{\lambda}{2}t} \quad \forall t \in [0, T[$$

Puisque $d_g(X(x_0, t), X(x_1, t)) \leq l(t)$ et $l(0) = d_g(x_0, x_1)$ (en effet γ est la géodésique qui joint les deux points x_0 et x_1), le théorème est démontré. \square

Lorsque la variété Riemannienne M est l'espace Euclidien \mathbb{R}^n , la dynamique $\frac{d}{dt}x = f(x, t)$ est une contraction lorsque la partie symétrique de la matrice Jacobienne est négative

$$\frac{\partial f}{\partial x} + \left(\frac{\partial f}{\partial x}\right)^T \leq 0$$

Enfin, l'intérêt pour la construction d'observateurs non-linéaires asymptotiques vient du fait qu'avec une dynamique contractante, il suffit de simuler le système pour avoir une estimation de x

$$\frac{d}{dt}\hat{x} = f(\hat{x}, t)$$

la dépendance en temps étant alors due aux entrées et/ou aux termes d'erreur entre la sortie estimée et la mesure.

L'exemple typique est le suivant. On considère un système mécanique à un degré de liberté x_1 , de vitesse x_2 obéissant à l'équation de Newton suivante

$$\frac{d}{dt}x_1 = x_2, \quad \frac{d}{dt}x_2 = F(x_1, x_2, t)$$

avec comme seule mesure $y = x_1$ et on suppose que $\frac{\partial F}{\partial x_2} \leq 0$. On considérer alors l'observateur

$$\begin{aligned} \frac{d}{dt}\hat{x}_1 &= \hat{x}_2 - \lambda(\hat{x}_1 - y(t)) \\ \frac{d}{dt}\hat{x}_2 &= F(y(t), \hat{x}_2) - \mu(\hat{x}_1 - y(t)) \end{aligned}$$

avec $\lambda, \mu > 0$ deux paramètres de réglage. On écrit symboliquement cet observateur $\frac{d}{dt}\hat{x} = f(\hat{x}, t)$ où la dépendance en t vient de la mesure $y(t) = x_1(t)$. Cet observateur s'écrit formellement $\frac{d}{dt}\hat{x} = f(\hat{x}, t)$. Avec comme produit scalaire celui qui est associé à la matrice constante symétrique définie positive

$$g = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\mu} \end{pmatrix}$$

on a

$$\left(\frac{\partial f}{\partial \hat{x}}\right)^T g + g \frac{\partial f}{\partial \hat{x}} = \begin{pmatrix} -2\lambda & 0 \\ 0 & \frac{2\frac{\partial F}{\partial x_2}}{\mu} \end{pmatrix} \leq 0$$

Nous n'avons pas une contraction stricte comme dans la Définition 19 mais au sens large. Il est cependant facile de voir que si $\frac{\partial F}{\partial x_2} \leq \beta < 0$ uniformément pour tout x avec $\beta > 0$, on a une contraction stricte et donc la convergence exponentielle de \hat{x} vers x .

Annexe A

Théorème de Cauchy-Lipchitz

On considère donc le problème de Cauchy suivant :

$$x(0) = x^0, \quad \frac{d}{dt}x(t) = f(x(t), t). \quad (\text{A.1})$$

Pour établir des résultats sur ces propriétés (essentiellement, les Théorèmes 1 et 2) nous allons utiliser la notion de solution approchante qui s'appuie simplement sur le schéma numérique d'Euler explicite

$$x_{k+1} = x_k + \delta f(x_k, k\delta), \quad x_0 = x(0), \quad k \in \mathbb{N}$$

où δ est un petit temps > 0 et x_k une approximation de $x(k\delta)$.

Considérons le *problème de Cauchy* (A.1). On suppose, de manière générale, que f est continue dans une région $R = \{\|x - x^0\| \leq b, |t| \leq a\}$. En général on prend comme norme $\|\cdot\|$, la norme euclidienne sur \mathbb{R}^n , mais ce n'est pas une obligation. Par la suite, sauf mention contraire $\|\cdot\|$ désigne la norme euclidienne. Il en découle que f est bornée sur R par une certaine constante $M > 0$.

Définition 20 (Solution approchante). *On appelle solution approchante à ϵ près du problème de Cauchy (A.1), une fonction $t \mapsto x(t)$ continue, dérivable par morceaux, telle que pour tout $(t, x(t)) \in R$ on a*

$$x(0) = x^0, \quad \left\| \frac{d}{dt}x(t) - f(x(t), t) \right\| \leq \epsilon$$

en tout point où $\frac{d}{dt}x$ est défini.

En d'autres termes, $\frac{d}{dt}x(t)$ peut ne pas être défini en un nombre fini de points de $|t| \leq a$.

De telles solutions approchantes existent comme le précise le résultat suivant.

Théorème 31

Soit le problème de Cauchy (A.1) où f est continue sur $R = \{\|x - x^0\| \leq b, |t| \leq a\}$ et majorée par $|f| \leq M$. Pour tout $\epsilon > 0$, on peut construire une solution approchante à ϵ près définie sur l'intervalle $|t| \leq \min(a, b/M) \triangleq h$.

Démonstration. L'application f est continue sur R , elle est donc uniformément continue. Pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que

$$|f(x_1, t_1) - f(x_2, t_2)| \leq \epsilon$$

quels que soient $(x_1, t_1) \in R$ et $(x_2, t_2) \in R$ avec $\|x_1 - x_2\| \leq \delta, |t_1 - t_2| \leq \delta$.

Considérons maintenant, δ étant donné, un ensemble de temps intermédiaires entre 0 et h

$$t_0 = 0 < t_1 < t_2 \dots < t_m = h$$

vérifiant

$$|t_i - t_{i-1}| \leq \min(\delta, \delta/M), \quad i = 1, \dots, m \quad (\text{A.2})$$

(par exemple, cette famille de temps intermédiaires peut être de plus en plus nombreuse lorsque δ diminue). Nous allons montrer que la fonction définie par la relation de récurrence (dite de Cauchy-Euler)

$$x(t) = x_{i-1} + (t - t_{i-1})f(x_{i-1}, t_{i-1}), \quad t_{i-1} \leq t \leq t_i, \quad x_i = x(t_i)$$

est une solution approchante à ϵ près. Cette fonction est continue, et elle possède une dérivée continue par morceaux (elle est en fait affine par morceaux par construction). Sa dérivée est définie partout sauf aux points $(t_i)_{i=1, \dots, m-1}$. En dehors de ces points, on a, pour $t_{i-1} \leq t \leq t_i$,

$$\left\| \frac{d}{dt} x(t) - f(x(t), t) \right\| = \|f(x_{i-1}, t) - f(x(t), t)\|$$

Par (A.2), et d'après la relation de récurrence, on a

$$\|x(t) - x_{i-1}\| = M(t - t_{i-1}) < M \frac{\delta}{M} = \delta \quad (\text{A.3})$$

D'après la continuité uniforme de f déjà évoquée, on déduit de (A.2)-(A.3)

$$\|f(x_{i-1}, t_{i-1}) - f(x(t), t)\| \leq \epsilon$$

On a donc établi, que pour tout $|t| < a$ en dehors des points t_1, \dots, t_m on a

$$\left\| \frac{d}{dt} x(t) - f(x(t), t) \right\| \leq \epsilon$$

Enfin, par construction $x(0) = x^0$, ce qui achève la démonstration. \square

Nous allons maintenant avoir besoin de la propriété Lipschitz permettant de prouver une inégalité de croissance de la propagation des approximations.

Définition 21 (Fonction Lipschitz). Une fonction scalaire $(x, t) \in R \subset \mathbb{R}^n \times \mathbb{R} \rightarrow f(x, t) \in \mathbb{R}$ est Lipschitz avec la constante $k > 0$ (ou k -Lipschitz) en x si

$$\|f(x_1, t) - f(x_2, t)\| \leq k \|x_1 - x_2\|$$

pour tout $(x_1, t), (x_2, t)$ dans R .

Définition 22 (Fonction Lipschitz vectorielle). Une fonction $(x, t) \in \mathbb{R}^n \times \mathbb{R} \rightarrow f(x, t) \in \mathbb{R}^n$ est Lipschitz en x si et seulement si chacune de ses composantes l'est (les constantes Lipschitz de chaque composante peuvent différer).

On notera, grâce à la proposition suivante, que la propriété Lipschitzienne est très générale.

Proposition 10. Soit $\mathbb{R}^n \times \mathbb{R} \ni (x, t) \mapsto f(x, t) \in \mathbb{R}^n$. Supposons que chaque dérivée partielle $\frac{\partial f(x, t)}{\partial x_i}$ pour $i = 1, \dots, n$ existe et soit de norme bornée par un scalaire c , alors $f(x, t)$ est Lipschitzienne en x avec comme constante $c\sqrt{n}$.

Démonstration. Quels que soient $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ deux vecteurs de \mathbb{R}^n , on peut écrire,

$$f(y, t) = f(x, t) + \sum_{i=1, \dots, n} \int_0^1 \frac{\partial f}{\partial x_i} \Big|_{(y+\lambda(x-y), t)} (x_i - y_i) d\lambda$$

Il suffit alors d'utiliser l'inégalité de Cauchy-Schwartz, l'inégalité triangulaire et l'hypothèse $\left\| \frac{\partial f(x, t)}{\partial x_i} \right\| \leq n$ pour avoir

$$\|f(y, t) - f(x, t)\| \leq c\sqrt{n} \|x - y\|$$

□

Nous allons maintenant pouvoir énoncer, sous une hypothèse de Lipschitz, un résultat important sur la propagation des différentes entre les solutions approchantes.

Théorème 32 (Propagation de la distance entre deux solutions approchantes)

Étant donnée $f(x, t)$ continue sur une région R et k-Lipschitz en x . Soient $x_1(t)$ et $x_2(t)$ deux solutions approchantes à ϵ_1 et ϵ_2 près du problème de Cauchy (A.1), définies sur $|t| \leq b$. La différence $d(t) = x_1(t) - x_2(t)$ satisfait, pour tout $|t| \leq b$

$$\|d(t)\| \leq \exp(k|t|) \|d(0)\| + \frac{\epsilon_1 + \epsilon_2}{k} (\exp(k|t|) - 1) \quad (\text{A.4})$$

Démonstration. La dérivée $\frac{d}{dt}d(t)$ est définie partout sauf en un nombre fini de points. Par définition des solutions approchantes, on a (en notant $\epsilon \triangleq \epsilon_1 + \epsilon_2$)

$$\begin{aligned} \left\| \frac{d}{dt}d(t) \right\| &= \left\| \frac{d}{dt}x_1(t) - \frac{d}{dt}x_2(t) \right\| \leq \|f(x_1(t), t) - f(x_2(t), t)\| + \epsilon \\ &\leq k \|x_1 - x_2\| + \epsilon \leq k \|d(t)\| + \epsilon \end{aligned}$$

où on a utilisé la propriété Lipschitzienne de f . Aux points tels que $\frac{d}{dt}d(t)$ existe et $d(t) \neq 0$, on peut donc conclure d'après le Lemme 3 (donné un peu plus loin),

$$\frac{d}{dt} \|d(t)\| \leq k \|d(t)\| + \epsilon$$

On a alors trois cas :

1. si $d(t)$ est identiquement nul, alors la conclusion est triviale
2. si $d(t)$ ne s'annule jamais sur $] -b, b[$, alors on peut écrire

$$\int_0^t \exp(-kt) \left(\frac{d}{dt} \|d(t)\| - k \|d(t)\| \right) dt \leq \epsilon \int_0^t \exp(-kt) dt$$

où le membre de gauche peut avoir un intégrand discontinu mais possède effectivement une primitive. Il vient

$$\exp(-kt) \|d(t)\| - \|d(0)\| \leq \frac{\epsilon}{k} (1 - \exp(-kt))$$

d'où la conclusion

3. s'il existe \bar{t} tel que $d(\bar{t}) \neq 0$, $|\bar{t}| \leq b$ alors que d s'annule quelque part sur l'intervalle $[-b, \bar{t}]$, alors, par continuité de $\|d(t)\|$, il existe t_1 , avec $-b \leq t_1 < \bar{t} \leq b$ tel que $d(t_1) = 0$ et $d(t) \neq 0$ sur $t_1 < t < \bar{t}$. Par une intégration semblable au cas précédent mais effectuée sur l'intervalle $[t_1, \bar{t}]$, on a alors

$$\|d(t)\| \leq \frac{\epsilon}{k} (\exp(k(t - t_1)) - 1)$$

qui est une majoration plus forte que celle recherchée.

□

Lemme 3. Si $x(t)$ est continûment différentiable sur un intervalle ouvert $]t_1, t_2[$, alors, en tout point t tel que $\|x(t)\| \neq 0$, la dérivée $\frac{d\|x(t)\|}{dt}$ existe et vérifie

$$\left| \frac{d\|x(t)\|}{dt} \right| \leq \left\| \frac{dx(t)}{dt} \right\|$$

Démonstration. Soit $t \in]t_1, t_2$ tel que $\|x(t)\| \neq 0$, avec $\|x(t)\| = \sqrt{\sum_{i=1}^n x_i^2(t)}$, il vient $\frac{d\|x(t)\|}{dt} = \frac{\sum_{i=1}^n x_i \frac{dx_i}{dt}}{\|x(t)\|}$ qui est bien définie. D'autre part, par passage à la limite dans l'inégalité suivante, on obtient le résultat désiré.

$$\frac{1}{\delta t} |\|x(t + \delta t)\| - \|x(t)\|| \leq \frac{1}{\delta t} \|x(t + \delta t) - x(t)\|$$

□

Nous pouvons maintenant nous intéresser aux théorèmes d'existence et d'unicité des solutions du problème de Cauchy.

Théorème 33 (Unicité de la solution au problème de Cauchy)

Soit le problème de Cauchy (A.1) où $f(x, t)$ est continue et Lipschitz en x dans un voisinage de $(x^0, 0)$. Ce problème possède au plus une solution.

Démonstration. Soient deux solutions (exactes) x_1 et x_2 du problème (A.1). D'après le Théorème 32, on a

$$\|x_1(t) - x_2(t)\| \leq 0$$

d'où $x_1 \equiv x_2$.

□

Théorème 34 (Existence d'une solution au problème de Cauchy)

Soit le problème de Cauchy (A.1) où $f(x, t)$ est continue et Lipschitz en x dans la région $R = \{|x - x^0| \leq b, |t| \leq a\}$. Soit M la borne supérieure de $\|f\|$ sur R . Il existe une solution $x(t)$ de (A.1) définie sur l'intervalle $|t| \leq \min(a, \frac{b}{M}) \triangleq h$.

Démonstration. Considérons une suite décroissante positive $(\epsilon_n)_{n \in \mathbb{N}}$ convergeant vers 0. D'après le Théorème 31, on peut construire une suite de fonctions $(x_n)_{n \in \mathbb{N}}$ qui sont chacune solution approchante à ϵ_n près de (A.1). On a alors, pour $|t| \leq h$,

$$\left\| \frac{d}{dt} x_n - f(x_n(t), t) \right\| \leq \epsilon_n, \quad x_n(0) = x^0$$

La dérivée $\frac{d}{dt} x_n$ est définie partout sauf en un nombre fini de points $t_i^{(n)}$, $i = 1, \dots, m_n$. On va prouver trois propriétés

1. la suite de fonctions $(x_n)_{n \in \mathbb{N}}$ converge uniformément sur $|t| \leq h$ vers une fonction continue $x(t)$
2. la suite $\left(\int_0^t f(t, x_n(t)) dt \right)_{n \in \mathbb{N}}$ converge uniformément vers $\int_0^t f(t, x(t)) dt$
3. la fonction $x(t)$ est continûment différentiable et vérifie $x(0) = x^0$ et $\frac{d}{dt} x(t) = f(x(t), t)$. Autrement dit $x(t)$ est solution du problème de Cauchy.

preuve du point 1

Considérons $n < p$ et appliquons l'inégalité du Théorème 32 aux deux solutions approchantes $x_n(t)$ et $x_p(t)$. Il vient, en notant k la constante Lipschitz de $f(x, t)$

$$\|x_n(t) - x_p(t)\| \leq \frac{\epsilon_n + \epsilon_p}{k} (\exp(|t|) - 1) \leq \frac{2\epsilon_n}{k} (\exp(|t|) - 1)$$

donc la suite de fonctions continues $(x_n(t))$ est uniformément de Cauchy, elle converge vers une fonction limite continue¹.

preuve du point 2

Considérons une région $R' = \{|t| \leq h, \|x - x^0\| \leq Mh\}$. L'application $f(x, t)$ est continue dans R' . Donc, pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que, pour tout $(t, x_1), (t, x_2)$ dans R' , tels que $\|x_1 - x_2\| \leq \delta$, on a

$$\|f(x_1, t) - f(x_2, t)\| \leq \epsilon$$

Pour ce δ , il existe d'après le résultat de convergence établi au point 1, $N \in \mathbb{N}$ tel que, pour tout $n > N$ on a

$$\|x_n(t) - x(t)\| \leq \epsilon$$

pour tout $|t| \leq h$. Donc la suite $(f(x_n(t), t))$ converge uniformément vers $f(x(t), t)$ et on a (uniformément en t)

$$\lim_{n \rightarrow \infty} \int_0^t f(x_n(t), t) dt = \int_0^t f(x(t), t) dt$$

1. L'ensemble des fonctions continues bornées de $|t| \leq h$ dans \mathbb{R}^n est un sous espace clos des fonctions bornées et est donc complet.

preuve du point 3

Reprendons l'inégalité $\left\| \frac{d}{dt}x_n(t) - f(x_n(t), t) \right\| \leq \epsilon_n$ pour tout $|t| < h$. En intégrant, on obtient

$$\left\| \int_0^{t_1} \left(\frac{d}{dt}x_n(t) - f(x_n(t), t) \right) dt \right\| \leq \epsilon_n |t_1| \leq \epsilon_n h$$

or $x_n(t)$ est continue, donc l'intégration donne

$$\left\| x_n(t_1) - x^0 - \int_0^{t_1} f(x_n(t), t) dt \right\| \leq \epsilon_n h$$

D'après le point 2, on a alors par passage à la limite

$$x(t) = x^0 + \int_0^{t_1} f(x_n(t), t) dt$$

L'intégrand $f(x(t), t)$ est continu, donc $x(t)$ est continûment différentiable et sa dérivée vaut $\frac{d}{dt}x(t) = f(x(t), t)$. \square

Annexe B

Fonctions de Lyapounov et stabilité des points d'équilibre

Dans cette annexe, nous prendrons comme définition de fonction de Lyapounov la définition ci-dessous qui est adaptée à l'étude de point d'équilibre mais qui est plus restrictive que la définition 5, page 28.

Définition 23. Soit \bar{x} un point d'équilibre d'un système $\frac{d}{dt}x = f(x)$ défini dans un domaine $\Omega \subset \mathbb{R}^n$, où f est Lipschitz. On appelle fonction de Lyapounov centrée en \bar{x} une fonction $V : \Omega \rightarrow \mathbb{R}$ qui vérifie les trois propriétés suivantes

1. V est continue et possède des dérivées partielles continues
2. $V(\bar{x}) = 0$ et $V(x) > 0$ sur $\Omega \setminus \{\bar{x}\}$
3. $\frac{d}{dt}V(x) = \nabla V(x) \cdot f(x) \leq 0$ pour tout $x \in \Omega$ (autrement dit V est décroissante le long des trajectoires).

Théorème 35 (Stabilité et stabilité asymptotique locale par une fonction de Lyapounov)

Avec les notations de la Définition 23, s'il existe une fonction de Lyapounov $V(x)$ dans un voisinage centré en \bar{x} alors \bar{x} est un point d'équilibre stable. Si de plus, $\frac{d}{dt}V < 0$ pour tout $x \in \Omega \setminus \{\bar{x}\}$ (on parle alors de fonction de Lyapounov stricte), alors \bar{x} est un point d'équilibre (localement) asymptotiquement stable.

Démonstration. Reprenons la Définition 1, et montrons tout d'abord la stabilité de \bar{x} . On souhaite montrer que, pour tout $\epsilon > 0$ tel que la boule centrée en \bar{x} de rayon ϵ soit contenue dans Ω , il existe $\eta > 0$ tel que, pour toute condition initiale $\|x^0 - \bar{x}\| \leq \eta$, on a

$$\|x(t) - \bar{x}\| \leq \epsilon, \text{ pour tout } t \geq 0$$

La fonction de Lyapounov va nous fournir ces inégalités qui portent sur la norme Euclidienne. Notons α le minimum atteint par V (qui est continue) sur la frontière (compacte) de la boule B_ϵ centrée en \bar{x} de rayon ϵ . Par hypothèse, $\alpha = \min_{\|x-\bar{x}\|=\epsilon} V(x) > 0$. Considérons maintenant le sous-ensemble

$$E_{\frac{\alpha}{2}} \triangleq \{x \in B_\epsilon / V(x) \leq \frac{\alpha}{2}\}$$

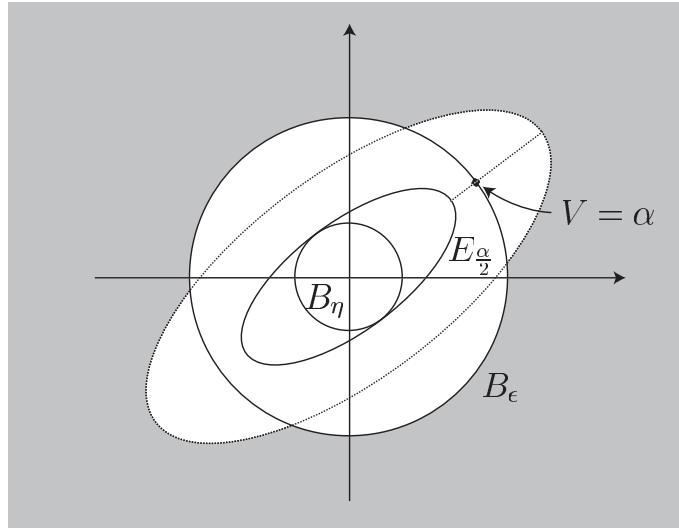


FIGURE B.1 – Liens entre fonction de Lyapounov et la norme Euclidienne.

Ce sous-ensemble est strictement à l'intérieur de B_ϵ et est donc égal à $\{x / V(x) \leq \frac{\alpha}{2}\}$. En effet, si ce n'était pas le cas, il existerait un x^* à la fois dans $E_{\frac{\alpha}{2}}$ et à la frontière de B_ϵ . On aurait alors

$$V(x^*) \leq \frac{\alpha}{2} \text{ et } \|\bar{x} - x^*\| = \epsilon \text{ donc } V(x^*) \geq \alpha$$

d'où la contradiction.

D'autre part, on note que V est continue et vérifie $V(\bar{x}) = 0$. On peut donc finalement construire une boule B_η centrée en \bar{x} de rayon η dans laquelle on a $V(x) \leq \frac{\alpha}{2}$. Cette boule est donc incluse dans $E_{\frac{\alpha}{2}}$.

On vient d'établir (voir Figure B.1)

$$B_\eta \subset E_{\frac{\alpha}{2}} \subset B_\epsilon$$

L'ensemble $E_{\frac{\alpha}{2}}$ est *positivement invariant* : pour toute condition initiale $x(0) = x^0$ appartenant à $E_{\frac{\alpha}{2}}$, on a $x(t) \in E_{\frac{\alpha}{2}}$ pour tout $t \geq 0$. En effet, on déduit de $\frac{d}{dt}V(x) \leq 0$ que $V(x) \leq V(x(0)) \leq \frac{\alpha}{2}$.

On peut alors conclure que pour toute condition initiale $x(0) \in B_\eta$ on a $x(0) \in E_{\frac{\alpha}{2}}$, et donc par l'invariance que nous venons d'évoquer, $x(t) \in E_{\frac{\alpha}{2}}$ pour tout $t \geq 0$ et donc $x(t) \in B_\epsilon$ pour tout $t \geq 0$. D'où la conclusion concernant la stabilité du point d'équilibre \bar{x} .

Considérons maintenant l'hypothèse supplémentaire $\frac{d}{dt}V < 0$ pour tout $x \in \Omega \setminus \{\bar{x}\}$. La fonction $t \mapsto V(x(t))$ est décroissante et bornée inférieurement par 0, donc elle converge. Notons

$$\lim_{t \rightarrow +\infty} V(x(t)) = l$$

Par hypothèse $l \geq 0$, montrons par l'absurde que $l = 0$. Si ce n'était pas le cas, alors $x(t)$ resterait en dehors, lorsque $t \rightarrow +\infty$, de $E_{\frac{l}{2}} = \{x \in /V(x) \leq \frac{l}{2}\}$ et donc, par la continuité de V et l'équation $V(\bar{x}) = 0$ déjà évoquées, en dehors d'une certaine boule B_r de rayon r vérifiant $\epsilon > r > 0$ centrée en \bar{x} . On peut alors calculer $k = -\max_{r \leq \|x\| \leq \epsilon} \frac{d}{dt}V(x)$ qui, par construction, vérifie $k < 0$. Enfin,

$$V(x(t)) = V(x(0)) + \int_0^t \frac{d}{dt}V(x(t))dt \leq V(x(0)) - kt$$

Le second membre de la dernière inégalité converge vers $-\infty$ ce qui est incompatible avec la propriété $V(x) \geq 0$.

Nous venons de montrer que $\lim_{t \rightarrow +\infty} V(x(t)) = 0$. Il nous faut maintenant conclure sur x lui-même. Considérons $\epsilon > 0$, et, comme précédemment, l'ensemble $E_{\frac{\epsilon}{2}} \subset B_\epsilon$. Puisque $\lim_{t \rightarrow +\infty} V(x(t)) = 0$, alors il existe un temps $T > 0$ tel que $x(t) \in E_{\frac{\epsilon}{2}}$ pour tout $t \geq T$. On en déduit que pour tout $\epsilon > 0$, il existe un temps $T > 0$ tel que $x(t) \in B_\epsilon$. Le point \bar{x} est donc asymptotiquement stable. \square

Au cours de la preuve du théorème précédent, nous avons été capables de construire un sous-ensemble de conditions initiales (noté $E_{\frac{\epsilon}{2}}$) fournissant des trajectoires convergeant asymptotiquement vers le point d'équilibre \bar{x} . Ce sous-ensemble fait donc partie d'un ensemble appelé *bassin d'attraction* du point \bar{x} . Ce dernier est défini comme l'ensemble des conditions initiales fournissant une trajectoire convergeant vers \bar{x} . Caractériser précisément cet ensemble est difficile, et nous en avons pour l'instant simplement trouvé une partie. Une question intéressante est de savoir sous quelle hypothèse le bassin d'attraction est l'espace \mathbb{R}^n tout entier, c.-à-d. sous quelle hypothèse \bar{x} est *globalement asymptotiquement stable*.

De manière préliminaire, considérons l'exemple d'un système de dimension 2, avec

$$V(x) = \frac{x_1^2}{1+x_1^2} + x_2^2$$

On peut vérifier que $\min_{\|x\|=\epsilon} V(x) \leq 1$, pour tout $\epsilon > 0$. Ainsi, la construction précédente est limitée à des ensembles $E_{\frac{\epsilon}{2}} \subset \{x / V(x) \leq 1/2\}$ qui ne couvrent pas l'ensemble du plan. On peut, en fait, garantir la possibilité d'une construction de $E_{\frac{\epsilon}{2}}$ aussi vaste que désiré si la fonction V est *non bornée radialement*, c.-à-d. si

$$\lim_{\|x - \bar{x}\| \rightarrow \infty} V(x) = +\infty$$

On obtient alors le théorème suivant, qui sera très utile en pratique

Théorème 36 (Stabilité asymptotique globale par une fonction de Lyapounov)

S'il existe une fonction de Lyapounov $V(x)$ définie sur \mathbb{R}^n non bornée radialement telle que $\frac{d}{dt}V < 0$ pour tout $x \neq \bar{x}$, alors \bar{x} est un point d'équilibre globalement asymptotiquement stable.

Démonstration. Définissons $[0, +\infty[\ni a \mapsto r(a) = \sup_{\{x / V(x) \leq a\}} \|x\|$. Cette fonction est bien définie car V est non bornée radialement. Le long d'une trajectoire ayant $x(0)$ pour condition initiale, on a par hypothèse $\frac{d}{dt}V \leq 0$, et donc $V(x(t)) \leq V(x(0))$ pour tout $t \geq 0$. Donc, pour tout $t \geq 0$, $\|x(t)\| \leq \sup_{\{x / V(x) \leq V(x(0))\}} \|x\| = r(V(x(0)))$. De cette inégalité on déduit que pour toute condition initiale $x(0)$, on peut, en échantillonnant la trajectoire qui en est issue, construire la suite $(x(i))_{i \in \mathbb{N}}$ et que, pour tout $t \geq i$ on a

$$\|x(t)\| \leq r(V(x(i))) \tag{B.1}$$

De la même manière que dans le Théorème 35, on a $\lim_{t \rightarrow +\infty} V(x(t)) = 0$. D'autre part, on va montrer que $\lim_{a \rightarrow 0} r(a) = 0$. La fonction $t \mapsto r(1/t)$ est décroissante et minorée donc elle converge lorsque $t \rightarrow +\infty$. Notons la limite $\lim_{a \rightarrow 0} r(a) = l \geq 0$. Supposons que cette limite est non nulle. Alors, pour toute suite $(a_n)_{n \in \mathbb{N}}$ convergeant vers zéro, il existe une suite $(x_n)_{n \in \mathbb{N}}$ telle que $\|x_n\| \geq l/2$ et $V(x_n) \leq a_n$. On déduit que, pour cette suite, $\lim_{n \rightarrow +\infty} V(x_n) = 0$ avec $\|x_n\| \geq l/2$ ce qui est impossible car V est non nulle en dehors de \bar{x} . On a donc $\lim_{a \rightarrow 0} r(a) = 0$.

On déduit donc de la suite d'inégalités (B.1) que $\lim_{t \rightarrow +\infty} x(t) = 0$. \square

Bien souvent, on trouvera que la condition de stricte négativité de la dérivée $\frac{d}{dt}V$ en dehors du point d'équilibre \bar{x} est trop contraignante. En effet, on pourra souvent trouver une fonction de Lyapounov (ce qui est déjà difficile) telle que l'ensemble des points x où sa dérivée est nulle n'est pas réduite à un point. Notons $Z = \{x / \frac{d}{dt}V(x) = 0\}$. Dans ce cas, le théorème suivant permet de déterminer l'ensemble vers lequel le système converge. Il utilise la notion d'*ensemble positivement invariant*, déjà rencontrée plus haut dans la preuve du Théorème 35 et donc la définition précise est donnée à la page 28.

Théorème 37 (Invariance de LaSalle (version locale))

Soit V fonction de Lyapounov (définition 5) pour le système $\frac{d}{dt}x = f(x)$, $x \in \mathbb{R}^n$. Soient $c > 0$ et E_c le sous-ensemble $\{x \in \mathbb{R}^n / V(x) \leq c\}$. On suppose que E_c est bornée et que $\frac{d}{dt}V(x) \leq 0$ à l'intérieur de E_c . Notons Z l'ensemble des points de E_c pour lesquels $\frac{d}{dt}V(x) = 0$. Toute trajectoire issue d'une condition initiale dans E_c converge lorsque $t \rightarrow +\infty$ vers le plus grand sous-ensemble positivement invariant contenu dans Z .

Exemple 29. Considérons l'oscillateur non linéaire suivant

$$\frac{d^2}{dt^2}x + k \frac{d}{dt}x + g(x) = 0$$

où k est un paramètre strictement positif et g est telle que $xg(x) > 0$ pour $x \neq 0$. On peut réécrire ce système sous forme d'état

$$\begin{aligned}\frac{d}{dt}x &= y \\ \frac{d}{dt}y &= -ky - g(x)\end{aligned}$$

Une fonction candidate à être de Lyapounov est

$$V(x, y) = \int_0^x g(\tau)d\tau + \frac{y^2}{2}$$

Cette fonction est continue et possède des dérivées partielles continues. V possède un unique minimum en $(x, y) = (0, 0)$. Calculons maintenant

$$\frac{d}{dt}V(x, y) = g(x)\frac{d}{dt}x + y\frac{d}{dt}y = -ky^2 \leq 0$$

On peut en déduire la stabilité de l'origine par le Théorème 35. Pour déduire la stabilité asymptotique, il nous faut utiliser le Théorème 37.

Le cas particulier $g(x) = \frac{g}{R} \sin x$, correspond à un pendule de longueur R soumis à la gravité, tel que représenté sur la Figure B.2.

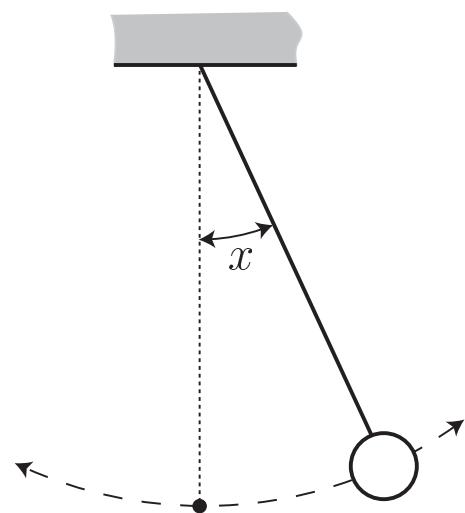


FIGURE B.2 – Pendule amorti.

Annexe C

Moyennisation

C.1 Introduction

On suppose ici que les effets rapides ont un caractère oscillant. La méthode de moyennisation a été utilisée en mécanique céleste depuis longtemps pour déterminer l'évolution des orbites planétaires sous l'influence des perturbations mutuelles entre les planètes et étudier la stabilité du système solaire. Gauss en donne la définition suivante qui est des plus intuitives : *il convient de répartir la masse de chaque planète le long de son orbite proportionnellement au temps passé dans chaque partie de l'orbite et de remplacer l'attraction des planètes par celle des anneaux de matière ainsi définis.*

Dans ce cadre, les équations non perturbées du mouvement de la terre sont celles qui ne prennent en compte que la force d'attraction due au soleil. L'orbite de la terre est alors une ellipse dont le soleil est l'un des foyers. Les équations perturbées sont celles où l'on rajoute les forces d'attraction entre la terre et les autres planètes en supposant que ces dernières décrivent toutes des orbites elliptiques selon les lois de Kepler. Le paramètre ε correspond au rapport de la masse du soleil à celles des planètes : $\varepsilon \approx 1/1000$. L'échelle de temps rapide est de l'ordre d'une période de révolution, quelques années. L'échelle de temps lente est de l'ordre de quelques millénaires. La question est alors de savoir si ces petites perturbations d'ordre ε peuvent entraîner à terme, i.e. à l'échelle du millénaire, une dérive systématique des longueurs du grand axe et du petit axe de la trajectoire de la terre, ce qui aurait des conséquences catastrophiques pour le climat. En fait, les calculs (moyennisation) montrent qu'il n'en est rien. En revanche, l'excentricité des orbites oscille lentement. Ces oscillations influencent le climat.

C.2 Le résultat de base

Considérons le système

$$\begin{cases} \frac{dx}{dt} = \varepsilon f(x, z, \varepsilon) \\ \frac{dz}{dt} = g(x, z, \varepsilon) \end{cases}$$

On passe de cette forme à la forme (Σ^ε) choisie pour énoncer le théorème de Tikhonov (Théorème 15) par un simple changement d'échelle de temps. On remplace t par t/ε . L'échelle de temps rapide correspond maintenant à t d'ordre 1 et l'échelle de temps lente à t de l'ordre de $1/\varepsilon$.

Le régime oscillatoire le plus simple pour y est le régime périodique, de période T :

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \begin{cases} \frac{dy_1}{dt} = y_2 & = g_1(x, y, \varepsilon) \\ \frac{dy_2}{dt} = -\left[\frac{2\pi}{T}\right]^2 y_1 & = g_2(x, y, \varepsilon), \end{cases}$$

Sans changer de notation, on pose $f(x, y(t), \varepsilon) = f(x, t, \varepsilon)$: f est régulière en x et dépend de t de façon périodique (période T). Le système perturbé s'écrit alors

$$\frac{dx}{dt} = \varepsilon f(x, t, \varepsilon), \quad 0 \leq \varepsilon \ll 1 \quad (\text{C.1})$$

Le système moyenné (ou système lent) est alors

$$\frac{dz}{dt} = \varepsilon \frac{1}{T} \int_0^T f(z, t, 0) dt \stackrel{\text{déf}}{=} \varepsilon \bar{f}(z) \quad (\text{C.2})$$

Remplacer les trajectoires du système instationnaire (C.1) par celles du système stationnaire (C.2), revient alors à lisser les trajectoires de (C.1). Comme pour le théorème 16, si le système moyen admet un point d'équilibre hyperboliquement stable (les valeurs propres du tangent sont toutes à partie réelle strictement négative) alors le système perturbé oscille légèrement autour de cet équilibre moyen.

De façon plus rigoureuse, théorème suivant montre qu'à un *point d'équilibre hyperbolique* du système moyen correspond une petite orbite périodique du système perturbé (C.1) (démonstration dans [30]).

Théorème 38 (Moyennisation à une fréquence)

Considérons le système perturbé (C.1) avec f régulière. Il existe un changement de variables, $x = z + \varepsilon w(z, t)$ avec w de période T en t , tel que (C.1) devienne

$$\frac{dz}{dt} = \varepsilon \bar{f}(z) + \varepsilon^2 f_1(z, t, \varepsilon)$$

avec \bar{f} définie par (C.2) et f_1 régulière de période T en t . De plus,

- (i) si $x(t)$ et $z(t)$ sont respectivement solutions de (C.1) et (C.2) avec comme conditions initiales x_0 et z_0 telles que $\|x_0 - z_0\| = O(\varepsilon)$, alors $\|x(t) - z(t)\| = O(\varepsilon)$ sur un intervalle de temps de l'ordre de $1/\varepsilon$.
- (ii) Si \bar{z} est un point fixe hyperbolique stable du système moyenné (C.2), alors il existe $\bar{\varepsilon} > 0$ tel que, pour tout $\varepsilon \in]0, \bar{\varepsilon}]$, le système perturbé (C.1) admet une unique orbite périodique $\gamma_\varepsilon(t)$, proche de \bar{z} ($\gamma_\varepsilon(t) = \bar{z} + O(\varepsilon)$) et asymptotiquement stable (les trajectoires démarrant près de $\gamma_\varepsilon(t)$ ont tendance à s'enrouler autour de cette dernière). L'approximation, à $O(\varepsilon)$ près, des trajectoires du système perturbé (C.1) par celles du système moyenné (C.2) devient valable pour $t \in [0, +\infty[$.

Il est instructif de voir comment est construit le changement de coordonnées $x = z + \varepsilon w(z, t)$ en enlevant à x des termes oscillants d'ordre ε (w de période T en t). On a, d'une part,

$$\frac{dx}{dt} = \frac{dz}{dt} + \varepsilon \frac{\partial w}{\partial z}(z, t) \frac{dz}{dt} + \varepsilon \frac{\partial w}{\partial t}(z, t)$$

et, d'autre part,

$$\frac{dx}{dt} = \varepsilon f(z + \varepsilon w(z, t), t, \varepsilon)$$

Ainsi

$$\begin{aligned}\frac{dz}{dt} &= \varepsilon \left(I + \varepsilon \frac{\partial w}{\partial z}(z, t) \right)^{-1} [f(z + \varepsilon w(z, t), t, \varepsilon) - \frac{\partial w}{\partial t}(z, t)] \\ &= \varepsilon [f(z, t, 0) - \frac{\partial w}{\partial t}(z, t)] + O(\varepsilon^2)\end{aligned}$$

Comme la dépendance en t de w est T -périodique, il n'est pas possible d'annuler complètement le terme d'ordre 1 en ε car il n'y a aucune raison pour que la fonction définie par

$$\int_0^t f(z, s, 0) ds$$

soit T -périodique en temps. En revanche, on peut éliminer la dépendance en temps du terme d'ordre 1 en ε . Il suffit de poser

$$w(z, t) = \int_0^t (f(z, s, 0) - \bar{f}(z)) ds$$

(noter que w est bien de T -périodique) pour obtenir

$$\frac{dz}{dt} = \varepsilon \bar{f}(z) + O(\varepsilon^2)$$

Si cette approximation n'est pas suffisante, il faut prendre en compte les termes d'ordre 2 et éliminer leur dépendance en temps par un changement de variable du type $x = z + \varepsilon w_1(z, t) + \varepsilon^2 w_2(z, t)$ avec w_1 et w_2 T -périodique.

C.3 Un exemple classique

Soit l'équation du second ordre suivante :

$$\frac{d^2\theta}{dt^2} = -\theta + \varepsilon(1 - \theta^2) \frac{d\theta}{dt}.$$

C'est l'équation d'un pendule pour lequel on a rajouté un petit frottement positif pour les grandes amplitudes ($\theta > 1$) et négatif pour les petites ($\theta < 1$). Mettons d'abord ce système sous la forme standard

$$\frac{dx}{dt} = \varepsilon f(x, t, \varepsilon)$$

Le terme oscillant vient du système non perturbé

$$\frac{d^2\theta}{dt^2} = -\theta$$

dont les orbites sont des cercles. Les phénomènes lents (échelle de temps $1/\varepsilon$) sont clairement relatifs aux rayons de ces cercles (i.e. les amplitudes des oscillations). C'est pourquoi il convient de passer en coordonnées polaires en posant $\theta = r \cos(\psi)$ et $\frac{d}{dt}\theta = r \sin(\psi)$. Dans ces coordonnées, le système perturbé s'écrit

$$\begin{cases} \frac{dr}{dt} &= \varepsilon[1 - r^2 \cos^2(\psi)] \sin^2(\psi) \\ \frac{d\psi}{dt} &= -1 + \varepsilon \sin(\psi) \cos(\psi)[1 - r^2 \cos^2(\psi)] \end{cases}$$

ψ est quasiment égal, à une constante près, au temps $-t$. On peut écrire

$$\frac{dr}{d\psi} = \frac{dr}{dt} \frac{dt}{d\psi}.$$

Ainsi, on se ramène à la forme standard en prenant ψ comme variable de temps

$$\frac{dr}{d\psi} = \varepsilon \frac{[1 - r^2 \cos^2(\psi)] \sin^2(\psi)}{-1 + \varepsilon \sin(\psi) \cos(\psi) [1 - r^2 \cos^2(\psi)]} = \varepsilon f(r, \psi, \varepsilon)$$

Le système moyenné est alors

$$\frac{du}{d\psi} = -\frac{\varepsilon}{8} u(4 - u^2)$$

$\bar{u} = 2$ est un *point d'équilibre hyperbolique* attracteur pour $\psi \rightarrow -\infty$, i.e. $t \rightarrow +\infty$. Donc pour ε suffisamment petit, l'équation perturbée possède un cycle limite hyperbolique attracteur donc l'équation est approximativement $\theta^2 + \frac{d}{dt}\theta^2 = 4 + O(\varepsilon)$.

L'inconvénient principal de la théorie des perturbations est qu'il faut, dès le départ, avoir une idée assez précise de ce que l'on cherche : il convient de trouver un petit paramètre ε et d'isoler la partie rapide du système. À ce niveau l'intuition physique joue un rôle important.

C.4 Recherche d'extremum (extremum seeking)

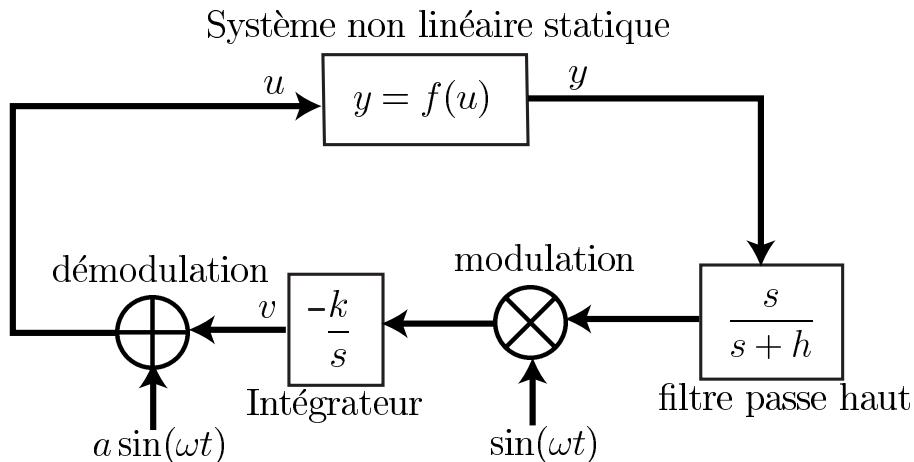


FIGURE C.1 – Boucle "extremum seeking" ; pour le système statique $y = f(u)$ où f n'est pas connue mais uniquement supposée KC^2 , u converge vers un minimum local de f lorsque la pulsation ω est choisie assez grande et l'amplitude a assez petite ((k, h) sont deux autres paramètres > 0).

Pour conduire une optimisation en temps-réel, en ligne et sans modèle, les ingénieurs ont souvent recours à une boucle de feedback décrite sur la figure C.1. Ce type de boucle est aussi couramment utilisé en spectroscopie pour ajuster, par exemple, la fréquence u d'un laser sur une fréquence atomique en cherchant le maximum d'un signal de fluorescence $y = f(u)$, la fonction f ayant la forme d'une Lorentzienne dont on cherche le maximum : $f(u) = \frac{p_1}{(u - \bar{u})^2 + p_2} + p_3$ avec $p_1, p_2, p_3 > 0$ paramètres inconnus et \bar{u} , l'argument du maximum de f , étant aussi inconnu.

Les équations associées au schéma bloc de la figure C.1 sont

$$\frac{dv}{dt} = -k \sin(\omega t) (f(v + a \sin(\omega t)) - \xi), \quad \frac{d\xi}{dt} = h(f(v + a \sin(\omega t)) - \xi).$$

Supposons que f admette un minimum local en $u = \bar{u}$, avec l'approximation $f(u) \approx \bar{f} + \frac{\bar{f}''}{2}(u - \bar{u})^2$ où $\bar{f} = f(\bar{u})$ et $\bar{f}'' = f''(\bar{u}) > 0$. Pour v autour de \bar{u} on a le système approché (à assez petit) :

$$\begin{aligned}\frac{d}{dt}v &= -k \sin(\omega t) \left(\bar{f} + \frac{\bar{f}''}{2}(v + a \sin(\omega t) - \bar{u})^2 - \xi \right) \\ \frac{d}{dt}\xi &= h \left(\bar{f} + \frac{\bar{f}''}{2}(v + a \sin(\omega t) - \bar{u})^2 - \xi \right)\end{aligned}$$

Alors si les gains (k, a, h, ω) sont tels que

$$ak\bar{f}'' \ll \omega, \quad h \ll \omega$$

on peut utiliser l'approximation séculaire et prendre la moyenne sur une période en temps. Le système moyen est alors simplement

$$\frac{d}{dt}v = \frac{ak\bar{f}''}{2}(\bar{u} - v), \quad \frac{d}{dt}\xi = h \left(\bar{f} + \frac{\bar{f}''}{2} \left((\bar{u} - v)^2 + \frac{a}{2} \right) - \xi \right).$$

Ce système triangulaire converge alors vers l'équilibre hyperbolique $v = \bar{u}$ et $\xi = \frac{a\bar{f}''}{4}$. Comme $u = v + a \sin(\omega t)$ on voit que u converge en moyenne vers \bar{u} .

C.5 Boucle à verrouillage de phase PLL

On dispose d'un signal d'entrée $v(t)$ bruité dont on souhaite estimer en temps réel la fréquence : cette dernière se situe autour d'une fréquence de référence notée $\omega_0 > 0$. Ainsi on a $v(t) = a(t) \cos(\theta(t)) + w(t)$ avec $|\dot{\theta} - \omega_0| \ll \omega_0$, $a(t) > 0$ avec $|\dot{a}| \ll \omega_0 a$ et $w(t)$ un bruit. L'écart type de w peut être bien plus grand que l'amplitude a : le rapport signal sur bruit peut être très défavorable. Pour ce faire, on dispose de circuits électroniques analogiques ou de systèmes digitaux, dits de type PLL¹, qui donnent en temps-réel une estimation non bruitée de la fréquence d'entrée $\frac{d}{dt}\theta$: ce type de circuits spécialisés se retrouve dans quasiment tous les systèmes électroniques de télé-communication (téléphone portable, carte WiFi, fibre optique, laser, horloges atomiques, GPS,...). Les PLL sont aussi utilisés par les physiciens pour mesurer très précisément les constantes fondamentales². Le principe de fonctionnement des PLL s'analyse très bien avec la moyennisation à une fréquence.

La figure C.2 donne un schéma bloc simplifié et de principe : chaque bloc correspond soit à un circuit spécifique dans le cas d'une PLL analogique, soit à une étape de l'algorithme pour une PLL digitale. La dynamique est régie par

$$\frac{d}{dt}x = \epsilon k_f \omega_0(v(t) \sin \phi - x), \quad \frac{d}{dt}\phi = \omega_0(1 - \epsilon kx)$$

où ϵ est un petit paramètre positif, k_f et k deux gains positifs. On pose $v(t) = a \cos \theta$ avec $\frac{d}{dt}\theta = \omega_0(1 + \epsilon p)$ où $a > 0$ et p sont des paramètres inconnus mais constants. Comme $2 \cos \theta \sin \phi = \sin(\phi - \theta) + \sin(\phi + \theta)$, avec $\Delta = \phi - \theta$ et $\sigma = \phi + \theta$, le système

$$\frac{d}{dt}x = \epsilon k_f \omega_0(a \cos \theta \sin \phi - x), \quad \frac{d}{dt}\phi = \omega_0(1 - \epsilon kx), \quad \frac{d}{dt}\theta = \omega_0(1 + \epsilon p)$$

1. PLL pour Phase-Locked Loop.

2. A.L. Schawlow (physicien, prix Nobel 1981 pour ses travaux sur la spectroscopie laser) disait que pour avoir une mesure très précise, il faut mesurer une fréquence.

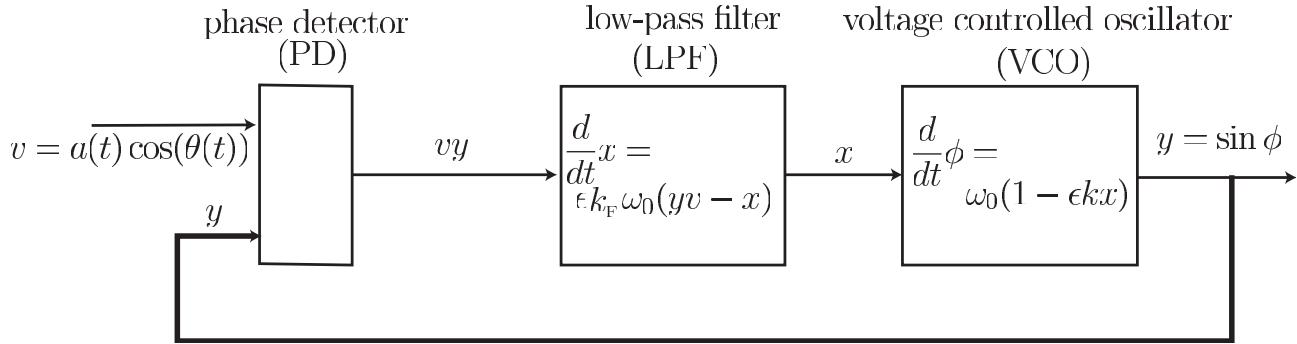


FIGURE C.2 – Le schéma bloc d'une PLL : la quantité $\omega_0(1 - \epsilon kx)$ est une estimation filtrée de la fréquence $\frac{d}{dt}\theta$ du signal d'entrée v qui peut être très fortement bruité et dont l'amplitude a n'est pas connue.

devient dans l'échelle de temps σ ($\frac{d}{dt}\sigma = \omega_0(2 + \epsilon(p - kx))$)

$$\frac{d}{d\sigma}x = \epsilon k_f \frac{\frac{a}{2}\sin\Delta + \frac{a}{2}\sin\sigma - x}{2 + \epsilon(p - kx)}, \quad \frac{d}{d\sigma}\Delta = -\epsilon \frac{p + kx}{2 + \epsilon(p - kx)}.$$

Ainsi, on est sous la forme standard du théorème 38 avec σ à la place de t . Le système moyen est

$$\frac{d}{d\sigma}x = \epsilon k_f \frac{\frac{a}{2}\sin\Delta - x}{2 + \epsilon(p - kx)}, \quad \frac{d}{d\sigma}\Delta = -\epsilon \frac{p + kx}{2 + \epsilon(p - kx)}.$$

On néglige des termes d'ordre 2 en ϵ et on prend comme système moyen :

$$\frac{d}{d\sigma}x = \frac{\epsilon k_f}{2} \left(\frac{a}{2}\sin\Delta - x \right), \quad \frac{d}{d\sigma}\Delta = -\frac{\epsilon}{2}(p + kx).$$

Ce système s'écrit aussi sous la forme d'une seule équation du second ordre avec $\sigma/\varsigma = \epsilon\sqrt{k_f ka}/8$

$$\frac{d^2}{d\varsigma^2}\Delta = -\sqrt{\frac{2k_f}{ka}} \frac{d}{d\varsigma}\Delta - \left(\sin\Delta - \frac{2p}{ak} \right).$$

On choisit le gain k assez grand pour que $|\frac{2p}{ak}| < 1$. Ainsi on pose $\sin\bar{\Delta} = \frac{2p}{ak}$ avec $\bar{\Delta} \in]-\frac{\pi}{2}, \frac{\pi}{2}[$. Ce système admet donc deux points d'équilibre (angle défini à 2π près) :

- $\Delta = \pi - \bar{\Delta}$ est un col (deux valeurs propres réelles de signes opposés)
- $\Delta = \bar{\Delta}$ est localement asymptotiquement stable (deux valeurs propres à partie réelle strictement négative).

On reprend en partie les arguments utilisés pour le PI avec anti-emballement (systèmes autonomes dans la plan, section 1.3.5). Ici, il faut faire un peu attention car l'espace des phases est un cylindre et non un plan. Une étude complète est faite au chapitre 7 de [6]. Nous en résumons ici les grandes lignes :

- Le calcul de la divergence du champ de vecteur dans les coordonnées $(\Delta, \Omega = \frac{d\Delta}{d\varsigma})$ donne $-\sqrt{\frac{2k_f}{ka}} < 0$. Les trajectoires sont bornées dans le cylindre (la vitesse est bornée).
- Ainsi, il ne peut y avoir au plus qu'une seule orbite périodique et de plus elle fait un tour autour du cylindre.

- Pour k et k_f assez grands, on a deux points d'équilibre et on n'a pas d'orbite périodique (bifurcation globale fondée sur l'espace rentrant du col $\pi - \bar{\Delta}$).

Pour k et k_f assez grands $\Delta = \phi - \theta$ converge donc vers une constante. C'est le verrouillage de phase : la différence Δ entre la phase ϕ de la PLL et la phase θ du signal d'entrée à analyser converge vers une valeur constante définie à 2π près : $\phi = \theta + \text{cte}$.

Ainsi $\frac{d}{dt}\phi$ converge vers $\frac{d}{dt}\theta$, comme $\frac{d}{dt}\phi = \omega_0(1 - \epsilon k x)$, on voit que l'on a une estimation peu bruitée de la fréquence d'entrée $\frac{d}{dt}\theta$ avec $\omega_0(1 - \epsilon k x)$. Pour cela, nous n'avons pas eu besoin de connaître précisément l'amplitude a . Noter que la même PLL donne, lorsque la fréquence reste fixe, les modulations de phase.

Annexe D

Automatique en temps discret

La représentation des systèmes dynamiques sous forme discrète est souvent nécessaire lorsqu'on doit réaliser de manière pratique le contrôle avec un système numérique d'acquisition, de traitement des données et des algorithmes ainsi que leur application. Alors que jusque dans les années 1960, les contrôleurs utilisaient seulement des composants analogiques, avec des circuits électriques assurant les tâches de mesure, de calcul et de transmission des ordres vers les actionneurs, aujourd'hui la majorité des systèmes de contrôle utilisent des *convertisseurs analogiques-numériques*, des microprocesseurs, et des *convertisseurs numériques-analogiques*.

Dans un tel schéma, les algorithmes de contrôle sont exécutés en temps discret, avec une périodicité (constante) correspondant à une fréquence choisie en fonction de la réactivité naturelle du système à contrôler. On peut considérer qu'en aéronautique, les boucles de guidage-pilotage sont exécutées tous les quelques (10) millisecondes, alors que dans le génie des procédés de la chimie lourde, les périodes sont plutôt de l'ordre de la minute.

Ainsi, dans la chaîne d'action entourant le système physique, on a des signaux numérisés utilisables par des microprocesseurs fonctionnant suivant des cycles d'horloge cadencés à une *fréquence d'échantillonage* $1/T_e > 0$. En outre, on doit souvent considérer plusieurs horloges, dont les fréquences n'ont pas nécessairement de multiple commun. Nous laissons volontairement de côté ces problèmes difficiles dits de multi-échantillonnage (multirate [3]), même s'ils ont une importance pratique certaine.

Dans la suite de ce chapitre, les mesures de signal sont supposées être effectuées à des instants nT_e où $n \in \mathbb{N}$. Elles vont être utilisées par le contrôleur prenant la forme d'un algorithme exécuté à chaque pas de temps $0, T_e, 2T_e, \dots$ supposé aboutir à un résultat de calcul en un temps inférieur¹ à T_e . Le résultat de cet algorithme est ensuite utilisé pour actionner un organe de commande (comme un moteur, une gouverne, etc...).

La description précédente correspond à la Figure D.1. Dans ce schéma, représentant une utilisation d'un algorithme de contrôle en temps discret à la fréquence $1/T_e$, deux éléments n'ont pas encore été décrits.

Le premier bloc est le CNA (*convertisseur numérique-analogique*), il s'agit d'un sous-système prenant en entrée des grandeurs discrètes et les transformant en signal continu. Cette conversion peut se faire suivant la méthode du bloqueur d'ordre zéro

$$y_b(t) = y(nT_e) = y[n], nT_e \leq t < (n+1)T_e$$

où y_b est le signal de sortie en temps continu, y est le signal discret d'entrée et n est la partie entière

1. la question des problèmes engendrés par des temps de calculs excédant T_e est discutée dans le cadre d'algorithmes numériquement intensifs comme la Model Predictive Control, MPC [58]

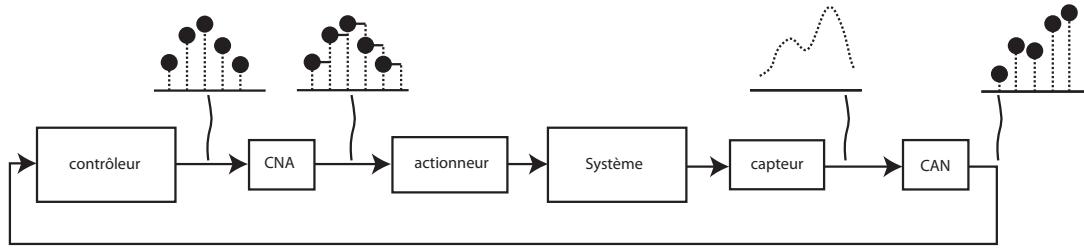


FIGURE D.1 – Schéma d’implémentation numérique d’un algorithme de contrôle.

de t/T_e . On peut utiliser d’autres types de bloqueur pour rendre le signal y_b continu, différentiable, ou plus régulier encore. Ainsi le bloqueur d’ordre 1 est

$$y_b(t) = y[n] + \frac{t - nT_e}{T_e}(y[n] - y[n-1]), nT_e \leq t < (n+1)T_e$$

Le second bloc est le CAN (*convertisseur analogique-numérique*). Il s’agit d’un système ayant en entrée un signal en temps continu (comme y_b précédemment) et en sortie un signal en temps discret (comme y précédemment). Un autre problème important est que la représentation du signal y est faite en utilisant un nombre fini de bits composant la représentation numérique binaire du signal échantillonné. On pourra se reporter à [35] pour une présentation des différentes technologies de conversion binaire. Laissant volontairement de côté les problèmes de représentation binaire de grandeurs à valeur réelle, il reste à se soucier des problèmes liés à l’échantillonnage à une fréquence $1/T_e$.

Le théorème de Shannon précise les liens entre un signal en temps continu et sa représentation échantillonnée.

Théorème 39 (Théorème de Shannon)

Si le signal en temps continu $y(t)$, $t \in \mathbb{R}$ ne contient pas de fréquence supérieure à B Hertz, alors il peut être reconstruit de manière exacte à partir d’échantillons régulièrement espacés de $1/(2B)$ secondes (ou moins).

Ce théorème fondamental en théorie de l’information, voir [76], doit être compris en détails pour en saisir la porté, et les limitations. Le signal $y(t)$ est supposé être défini pour tout $t \in \mathbb{R}$, et les échantillons $y[n]$ sont supposés être connus pour tout $n \in \mathbb{Z}$. L’hypothèse principale est que $y(t)$ est un signal à contenu fréquentiel borné, c.-à-d. que sa *transformée de Fourier*²

$$F(t \mapsto y(t), N) = \int_{-\infty}^{+\infty} y(t) \exp(-2i\pi Nt) dt$$

est à support borné inclus dans $] -B, B[$ (c.-à-d. qu’elle est nulle en dehors de cet intervalle).

L’énoncé ne s’applique pas à des signaux en temps continu défini seulement sur un intervalle de temps borné, car ces signaux ne sont en général pas à contenu fréquentiel à support borné³. L’hypothèse de contenu fréquentiel nul en dehors du domaine $] -B, B[$ peut en pratique être relâchée, au

2. on notera la notation très proche de celle du cours [54] avec la correspondance $F(t \mapsto y(t), N) = (\mathcal{F}y)(2\pi N)$.

3. Si y est à support compact, alors sa transformée de Fourier est une fonction analytique, et donc ne peut pas être à support compact car sinon elle est nulle, voir [26][§31.5.4]

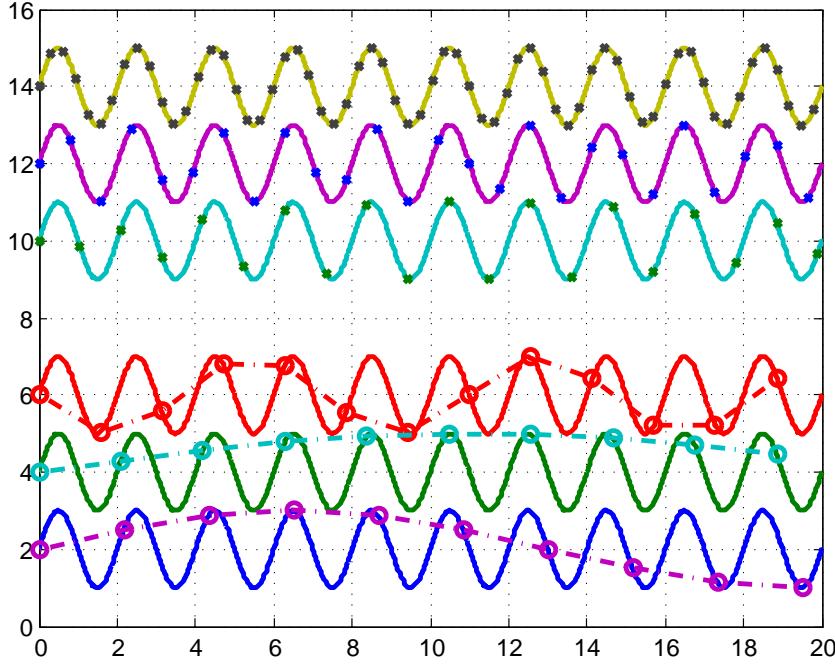


FIGURE D.2 – Apparition du phénomène de repliement spectral lors de l'échantillonage d'un signal sinusoïdal. Des basses fréquences apparaissent lorsque la fréquence d'échantillonage choisie viole l'hypothèse du Théorème 39 de Shannon (courbes du bas).

prix d'une reconstruction inexacte due au phénomène de repliement spectral (aussi appelé *aliasing*) exposé ci-dessous.

Le phénomène d'*aliasing* peut-être mis en évidence très simplement sur un signal sinusoïdal de fréquence supérieure à $1/(2T_e)$. Considérons la figure D.2

Si le signal à échantillonner est préalablement filtré par un filtre *anti-aliasing* ayant pour fonction d'atténuer complètement toute fréquence supérieure à $1/(2T_e)$ (c.-à-d. d'avoir un gain nul à toutes les fréquences au dessus de $1/(2T_e)$), ce signal filtré peut être en théorie parfaitement reconstruit. Des filtres de dimension finie peuvent approcher ce filtre *anti-aliasing* idéal. On peut ainsi utiliser les filtres de Butterworth, ou de Bessel réglés à la bonne fréquence [35].

La formule de reconstruction est la formule de Shannon-Whittaker

$$y(t) = \sum_{n=-\infty}^{+\infty} y[n] \text{sinc} \frac{(t - nT_e)}{T_e}$$

où *sinc* est la fonction sinus-cardinal $\text{sinc}(t) = \frac{\sin(t)}{t}$. Cette formule de reconstruction est équivalente au filtrage d'un signal constitué d'impulsions de Dirac δ pondérées par les échantillons $y[n]$ avec un filtre passe bas idéal (ayant une fonction de transfert rectangle dans le domaine fréquentiel) à la fréquence $1/(2T_e)$.

La formule de reconstruction précédente est convergente sous la condition suffisante que la suite $(y[n])_{n \in \mathbb{Z}}$ appartient à l'espace $\ell^p(\mathbb{Z}, \mathbb{R})$ c.-à-d. $\sum_{n \in \mathbb{Z}} (y[n])^p < \infty$, avec $1 < p < \infty$. Elle est aussi convergente si la suite est constante.

Pour démontrer le théorème de Shannon, on utilise un résultat intermédiaire utilisant le lemme suivant (formule de Poisson).

Lemme 4 (Formule de Poisson). *Soit y une fonction à valeur réelle (ou dans \mathbb{C}) de classe C^2 telle que f, \dot{f}, \ddot{f} sont sommables et à décroissance $|y(x)| \leq \frac{K}{1+x^2}$, $K > 0$. On note $F(t \mapsto y(t), N)$ sa transformée de Fourier. Alors, on a la formule de sommation de Poisson*

$$\sum_{n=-\infty}^{+\infty} y[n] = \sum_{k=-\infty}^{+\infty} F(t \mapsto y(t), k)$$

Ce résultat est démontré dans [72], voir aussi [54] pour un énoncé dans l'espace de Schwartz, et [26] pour un énoncé concernant les fonctions de classe $L^1(\mathbb{R})$.

On utilise ce lemme pour donner une expression simple à la fonction suivante, nommée *sommation périodique* de la transformée de Fourier de la fonction y

$$\phi(N) = \sum_{k=-\infty}^{+\infty} F(t \mapsto y(t), N - k/T_e) \quad (\text{D.1})$$

En utilisant les propriétés usuelles de la transformée de Fourier (modulation et translation en fréquence, voir [54]), on obtient

$$F(t \mapsto y(t), N - k/T_e) = T_e F(t \mapsto \exp(2i\pi t NT_e) y(-tT_e), k)$$

En utilisant la formule de Poisson du lemme 4, il vient

$$\begin{aligned} \phi(N) &= T_e \sum_{n=-\infty}^{+\infty} \exp(2i\pi n NT_e) y(-nT_e) \\ &= T_e \sum_{n=-\infty}^{+\infty} y[n] \exp(-2i\pi n NT_e) \end{aligned} \quad (\text{D.2})$$

Lorsque $F(t \mapsto y(t), N)$ est nulle en dehors de $]-\frac{1}{2T_e}, \frac{1}{2T_e}[$, alors, sous l'hypothèse que $B < \frac{1}{2T_e}$, la formule (D.1) s'interprète comme une somme (infinie) de fonctions à supports disjoints. On peut isoler chaque terme de la somme en procédant ainsi

$$F(t \mapsto y(t), N) = \phi(N) \times \begin{cases} 1 & \text{si } |N| < B \\ 0 & \text{sinon} \end{cases} = \phi(N) \times \text{rect}\left(\frac{N}{B}\right)$$

où rect est la fonction nulle partout sauf sur $[-1, 1]$ où elle vaut 1. En utilisant (D.2), on a alors

$$F(t \mapsto y(t), N) = T_e \sum_{n=-\infty}^{+\infty} \exp(-2i\pi n NT_e) y[n] \text{rect}(N/B)$$

or

$$\exp(-2i\pi n NT_e) \text{rect}(N/B) = F(t \mapsto \frac{1}{T_e} \text{sinc}\left(\frac{t - nT_e}{T_e}\right), N)$$

comme cela est aisément montré par le calcul.

Ainsi, on a

$$F(t \mapsto y(t), N) = F(t \mapsto \sum_{n=-\infty}^{+\infty} \text{sinc}\left(\frac{t - nT_e}{T_e}\right) y[n], N)$$

d'où l'identification

$$y(t) = \sum_{n=-\infty}^{+\infty} \text{sinc}\left(\frac{t - nT_e}{T_e}\right) y[n]$$

Ce résultat ne peut plus être établi lorsqu'on a recouvrement des spectres dans la sommation de Poisson. Des termes additionnels viennent s'ajouter au signal $y(t)$, ce phénomène est désigné *aliasing* [76].

En pratique, les principaux problèmes liés à l'échantillonnage sont liés à la violation des hypothèses du théorème de Shannon. Pour limiter l'effet d'aliasing, il est nécessaire de filtrer les signaux avant leur échantillonnage. Néanmoins, même si le signal considéré $y(t)$ satisfait l'hypothèse, il advenit souvent que le signal effectivement accessible à la mesure, c.-à-d. à la conversion CAN, viole cette hypothèse. En effet, les bruits sont souvent à spectre haute fréquence et viennent donc se replier en signaux basse-fréquence par le phénomène d'aliasing décrit précédemment. Ce phénomène renforce la nécessité de filtrer avec l'échantillonnage.

D.1 Représentation externe et interne

Dans le domaine discret, les équations de récurrence (appelées également équations aux différences) remplacent les équations différentielles.

Comme dans le cas du temps continu, les relations entre les entrées et les sorties peuvent être obtenues soient par une identification entrée-sortie, on parle alors de *représentation externe*, soit en modélisant avec des états les équations du fonctionnement du système, c'est la *représentation interne*.

Le passage de la représentation *interne à externe* est direct, à travers le calcul algébrique de la transformée en z . Ainsi, au système à 1 seule entrée, et une seule sortie défini par les équations matricielles de récurrence

$$x[k+1] = A_d x[k] + B_d u[k] \quad (\text{D.3})$$

$$y[k] = C x[k] + D u[k] \quad (\text{D.4})$$

où A_d et B_d sont, respectivement, une matrice $n \times n$, $n > 0$ et un vecteur $n \times 1$, il correspond la *fonction de transfert discrète*

$$H(z) = C(zI - A_d)^{-1} B_d + D$$

On peut généraliser ce calcul aux systèmes à plusieurs entrées et plusieurs sorties. Dans ce cas, $H(z)$ est une matrice de fonctions de transfert discrètes.

Cette fonction de transfert est calculée à partir de la transformée en z des équations de la description interne (D.3) (D.4). Cette transformée est présentée dans ce qui suit.

Les matrices A_d , B_d peuvent avoir un lien avec des équations différentielles $\dot{X} = AX + BU$, $Y = CX + DU$ qu'on aura discrétisées à la période d'échantillonnage T_e . Ainsi une manière de les discréteriser de manière dite “*exacte*” (au sens du bloqueur d'ordre 0) est d'utiliser les matrices

$$A_d = \exp(AT_e), \quad B_d = \int_0^{T_e} \exp(At) B dt$$

Dans la suite, on simplifiera la notation A_d , B_d pour n'utiliser que A et B dans un cadre discret.

D.1.1 Transformée en z

Considérons un signal en temps continu $y(t)$, pour lequel on s'intéresse aux valeurs $y(nT_e) = y[n]$ aux instants d'échantillonnage nT_e , $n \in \mathbb{N}$. On appelle *transformée en z* du signal y la fonction

$$\mathcal{Z}(t \mapsto y(t), z) = \sum_{n=0}^{+\infty} y(nT_e)z^{-n} = \sum_{n=0}^{+\infty} y[n]z^{-n} \quad (\text{D.5})$$

La dernière écriture sera préférée dans la suite. Cette fonction de la variable z dépend de la période d'échantillonnage T_e . La transformation en z d'un signal n'est pas une bijection (même si on n'a pas encore précisé d'espace d'arrivée et d'espace de départ) : étant donnée un signal y , il existe une unique transformée en z correspondante, mais une infinité de signaux continus y ont pour image une transformée en z donnée. En effet, seules les valeurs aux instants d'échantillonnage déterminent le signal continu antécédent d'une transformée en z donnée.

Une transformée en z définie par (D.5) est une série. Nous laissons de côté les problèmes de convergence, comme nous l'avons déjà fait pour la transformée de Laplace, le lecteur intéressé pourra se reporter à [20][Chap. XVI]. Dans le domaine des signaux échantillonnés, les propriétés de la transformée en z en font l'équivalent de la transformée de Laplace pour les signaux en temps continu. Ses propriétés principales sont : la linéarité (D.6), la translation par un réel (pour les systèmes à retard (D.7) ou avance (D.8)), l'homothétie exponentielle (D.9), la dérivation (D.10) et la convolution

$$\mathcal{Z}(a_1y_1(\cdot) + a_2y_2(\cdot), z) = a_1\mathcal{Z}(y_1(\cdot), z) + a_2\mathcal{Z}(y_2(\cdot), z) \quad (\text{D.6})$$

$$\mathcal{Z}(t \mapsto y(t - nT_e), z) = z^{-n}\mathcal{Z}(t \mapsto y(t), z), \quad n \in \mathbb{N} \quad (\text{D.7})$$

$$\mathcal{Z}(t \mapsto y(t + nT_e), z) = z^n \left(\mathcal{Z}(t \mapsto y(t), z) - \sum_{i=0}^{n-1} y[i]z^{-i} \right), \quad n \in \mathbb{N} \quad (\text{D.8})$$

$$\mathcal{Z}(t \mapsto \exp(-at)y(t), z) = \mathcal{Z}(t \mapsto y(t), \exp(aT_e)z) \quad (\text{D.9})$$

$$\mathcal{Z}(t \mapsto ty(t), z) = -T_ez\mathcal{Z}(t \mapsto y(t), z) \quad (\text{D.10})$$

Enfin, on notera que la transformée en z est une isométrie car elle préserve l'énergie d'un signal au sens de l'égalité de Parseval

$$\sum_{n=0}^{+\infty} y^2[n] = \frac{1}{2\pi i} \oint_{\mathcal{C}} z^{-1} \mathcal{Z}(t \mapsto y(t), z) \mathcal{Z}(t \mapsto y(t), z^{-1}) dz$$

où \mathcal{C} est un contour entourant l'origine dans le plan complexe.

D.1.2 Réalisation canonique d'une fonction de transfert discrète

Une réalisation de l'équation aux différences

$$y[k] + a_1y[k-1] + \dots + a_ny[k-n] = b_1u[k-1] + \dots + b_nu[k-n]$$

correspondant à la fonction de transfert⁴

$$G(z) = \frac{b_1z^{-1} + \dots + b_nz^{-n}}{1 + a_1z^{-1} + \dots + a_nz^{-n}}$$

4. dont certains coefficients peuvent être nuls sans perte de généralité

est

$$A_d = \begin{pmatrix} 0 & 1 & & 0 \\ \vdots & \cdots & \cdots & \\ 0 & \dots & 0 & 1 \\ -a_n & -a_{n-1} & \dots & -a_1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

$$C_d = (b_n, b_{n-1}, \dots, b_2, b_1)$$

D'autres réalisations canoniques peuvent être considérées. De manière générale, soit T une matrice (inversible) de changement de base, le triplet $\{TA_dT^{-1}, TB_d, C_dT\}$ est une réalisation au même titre que $\{A_d, B_d, C_d\}$.

D.2 Analyse de la stabilité

Considérons la fonction de transfert discrète

$$G(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_n z^{-n}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} = \frac{N(z)}{D(z)}$$

En soumettant le signal à des signaux d'entrée (y compris des conditions initiales par exemple), on engendre des signaux de sortie qui peuvent se calculer explicitement à travers une décomposition en éléments simples de la fraction rationnelle obtenue et identification de la transformée inverse de chaque terme, pour obtenir une description complète de la sortie au cours du temps (continu). Le comportement asymptotique de la solution, c.-à-d. lorsque le temps tend vers l'infini, est défini comme suit.

Définition 24. *Un point d'équilibre \bar{x} de l'équation aux différences $x[k+1] = Ax[k]$ est asymptotiquement stable si, pour toute condition initiale x_0 , la solution $x[k]$ tend vers \bar{x} lorsque $k \rightarrow +\infty$. Le point est instable s'il existe une condition initiale telle que $\|x[k]\| \rightarrow +\infty$ lorsque $k \rightarrow +\infty$. Il est stable si $x[k]$ reste borné.*

Il apparaît la condition nécessaire et suffisante suivante.

Théorème 40 (Stabilité en temps discret [52])

Une condition nécessaire et suffisante pour qu'un point d'équilibre de l'équation aux différences $x[k+1] = Ax[k]$ soit asymptotiquement stable est que les valeurs propres de A soient toutes à l'intérieur (strict) du cercle de rayon 1 ayant l'origine pour centre. Si une valeur propre est en dehors de ce cercle, alors le système est instable.

Comme dans le cas des systèmes en temps continu, on dispose d'un résultat permettant de tester le fait que les racines d'un polynôme satisfont les propriétés requises pour les conditions nécessaires de stabilité asymptotique.

Théorème 41 (Critère de Jury,[16])

Soit un polynôme $P(z) = a_0z^n + a_1z^{n-1} + \dots + a_n$ avec $a_0 > 0$. On forme le tableau

Ligne 1	a_0	a_1	\dots	\dots	a_n
Ligne 2	a_n	a_{n-1}	\dots	\dots	a_0
Ligne 3	b_0	b_1	\dots	b_{n-1}	0
	$\left\{ \begin{array}{l} = \text{Ligne 1} - \frac{a_n}{a_0} \text{Ligne 2} \\ b_{n-1} \quad b_{n-2} \quad \dots \quad b_0 \quad 0 \end{array} \right.$				
Ligne 4	b_{n-1}	b_{n-2}	\dots	b_0	0
Ligne 5	c_0	c_1	\dots	c_{n-2}	0 0
Ligne 6	$\left\{ \begin{array}{l} = \text{Ligne 3} - \frac{b_{n-1}}{b_0} \text{Ligne 4} \\ \dots \end{array} \right.$				
\vdots					
Ligne $2n+1$	w_0	0	\dots	0	

Une condition nécessaire et suffisante pour que les racines de $P(z)$ soient toutes (strictement) à l'intérieur du cercle de rayon 1 ayant l'origine comme centre est que tous les éléments de (a_0, b_0, \dots, w_0) soient strictement positifs.

D.3 Commandabilité en temps discret

Considérons maintenant l'équation aux différences où intervient la commande

$$x[k+1] = Ax[k] + Bu[k] \quad (\text{D.11})$$

où A est une matrice de dimension $n \times n$ et B est une matrice de dimension $n \times m$. Comme on va le voir, il faut distinguer, sauf dans le cas où A est inversible, la notion de commandabilité (possibilité d'aller de n'importe quel état à 0) de l'atteignabilité (possibilité d'aller de 0 à n'importe quel état).

Définition 25 (Commandabilité en temps discret). *Le système en temps discret(D.11) est dit commandable, si quel que soit l'état initial x^0 , il existe une séquence de commande $u[0], u[1], \dots, u[n-1]$ qui permet d'atteindre l'état 0.*

Définition 26 (Atteignabilité en temps discret). *Le système en temps discret(D.11) est dit atteignable, si quel que soit l'état final désiré x^f , il existe une séquence de commande $u[0], u[1], \dots, u[n-1]$ amenant le système initialisé en 0 à l'état x^f .*

Par un simple calcul, on établit la relation générale entre un état initial x^0 , une séquence de commande $u[0], u[1], \dots, u[n-1]$ et le point atteint $x[k]$ au bout de k itérations.

$$\begin{aligned} x[n] &= A^n x[0] + [B, AB, \dots, A^{n-1}B] [u[n-1], \dots, u[0]]^T \\ &= A^n x[0] + \mathcal{C} [u[n-1], \dots, u[0]]^T \end{aligned} \quad (\text{D.12})$$

où \mathcal{C} est la *matrice de commandabilité* de la paire (A, B) déjà rencontrée à la Proposition 8. En prenant comme inconnue la séquence de commande dans la dernière équation, on établit directement le résultat suivant (voir [39]).

Proposition 11 (Atteignabilité et commande en boucle ouverte atteignant un point arbitraire en temps fini). Soit un système discret $x[k+1] = Ax[k] + Bu[k]$ avec $\dim x = n$. Ce système est atteignable si et seulement si la matrice $\mathcal{C} = [B, AB, \dots, A^{n-1}B]$ est de rang plein. Soit x^f un point arbitraire qu'on souhaite atteindre. La commande en boucle ouverte

$$[u[n-1], \dots, u[0]]^T = \mathcal{C}^{-1} \times (x^f - A^n x[0])$$

permet de faire transiter en exactement n itérations le système de $x[0]$ à x^f .

Si on souhaite que le système atteigne l'état 0, la condition que la matrice de commandabilité \mathcal{C} est de rang plein est une condition suffisante mais pas nécessaire. Ainsi \mathcal{C} peut être singulière si A est nilpotente, et dans ce cas, la commande $u = 0$ suffit à atteindre 0. Ainsi la propriété de commandabilité dans le cas discret n'est pas équivalente au fait que la matrice \mathcal{C} est de rang plein.

La condition nécessaire de commandabilité est que $A^n x^0$ doit appartenir à l'espace engendré par les vecteurs $B, AB, \dots, A^{n-1}B$ qui est l'espace accessible à la commande. Si A est inversible, on peut ré-écrire (D.12) sous la forme

$$A^{-n} x[n] = x[0] + A^{-n} \mathcal{C} [u[n-1], \dots, u[0]]^T$$

En considérant la séquence de commande comme inconnue dans cette dernière équation, on voit que dans ce cas, la commandabilité est équivalente au fait que la matrice $A^{-n} \mathcal{C}$ soit de rang plein, ce qui est équivalent à ce que \mathcal{C} soit de rang plein.

D.3.1 Placement de pôles

Considérons l'équation aux différences où intervient la commande

$$x[k+1] = Ax[k] + Bu[k]$$

où A est une matrice de dimension $n \times n$ et B est un vecteur de dimension n . Comme dans le cas continu, il est possible lorsque le système est commandable, de choisir le polynôme caractéristique $\det(zI - A + BK)$ de $A - BK$ en choisissant égal à un polynôme $P(z)$ librement choisi (de premier terme normalisé z^n). Ceci est réalisé en choisissant le gain K d'après la formule d'Ackermann (qui sert également dans le cas continu)

$$K = [0, \dots, 0, 1] \times \mathcal{C}^{-1} \times P(A)$$

D.3.2 Rendre une matrice nilpotente par feedback

Pour stabiliser asymptotiquement le système (commandable) il est suffisant de rendre la matrice $A - BK$ nilpotente. Ainsi, quelle que soit la condition initiale, le système atteint le point 0 en au plus n itérations.

Pour réaliser ceci, il suffit d'utiliser la formule d'Ackermann avec $P(z) = z^n$. Ainsi un gain qui rend la matrice $A - BK$ nilpotente est

$$K = [0, \dots, 0, 1] \times \mathcal{C}^{-1} \times A^n = [0, \dots, 0, 1] \times [A^{-n}B, \dots, A^{-1}B]^{-1}$$

après simplification.

D.3.3 Synthèse d'une commande pour aller en temps fini à un point d'équilibre arbitraire

En utilisant le résultat précédent, il est aisément de calculer une loi de feedback permettant d'atteindre en un nombre fini d'itérations un point d'équilibre x^f arbitrairement choisi. Ceci résoud le problème de *planification de trajectoires*. A ce point d'équilibre il correspond une commande d'équilibre u^f solution de l'équation

$$x^f = Ax^f + Bu^f$$

En utilisant le contrôleur K précédent, on forme la commande

$$u = -K(x - x^f) + u^f \quad (\text{D.13})$$

Ce qui donne l'équation aux différences

$$x[k+1] = Ax[k] - BK(x[k] - x^f) + Bu^f$$

d'où, après factorisation, et en utilisant (D.13),

$$(x[k+1] - x^f) = (A - BK)(x[k] - x^f)$$

La matrice $(A - BK)$ étant nilpotente, grâce au choix de K , on atteint donc x^f en au plus n itérations, quelle que soit la condition initiale. Ceci établit le résultat suivant.

Proposition 12 (Commande en feedback atteignant un point d'équilibre en temps fini). *Soit un système discret $x[k+1] = Ax[k] + Bu[k]$ atteignable avec $\dim x = n$. Soit x^f un point d'équilibre qu'on souhaite atteindre, et soit u^f la commande d'équilibre correspondante. La commande en boucle fermée $u = -K(x - x^f) + u^f$ où $K = [0, \dots, 0, 1] \times [A^{-n}B, \dots, A^{-1}B]^{-1}$ permet d'atteindre le point x^f en au plus n itérations.*

D.3.4 Commande linéaire quadratique LQR en temps discret

Tout comme dans le cas continu, il est intéressant de planifier des transitoires optimaux entre deux points de l'espace d'état, ainsi que de calculer des régulateurs stabilisants optimaux au sens d'un critère quadratique. On se reportera à § 3.4 pour des explications générales sur la formulation du critère et l'obtention des équations de stationnarité qui permettent de caractériser la solution. Nous donnons ici la solution dans le cas du temps discret.

Théorème 42 (Commande quadratique en temps discret)

Soit un système discret $x[k+1] = Ax[k] + Bu[k]$, $\dim x = n$, $\dim u = 1$, et sa condition initiale x^0 . Soit le critère à minimiser

$$J_N = x^T[N]S_f x[N] + \sum_{i=1}^{N-1} x^T[i]Rx[i] + \sum_{i=0}^{N-1} u^T[i]Qu[i]$$

avec $N > 0$, S_f et R matrices symétriques positives, Q matrice symétrique définie positive. La loi de commande minimisant ce critère est la loi de feedback

$$u[i] = -K[i]^{-1}k[i]x[i]$$

avec

$$\begin{aligned} k[i-1] &= B^T S[i] A, \\ K[i] &= Q + B^T S[i] B \\ S[i-1] &= A^T S[i] A + R - k[i-1]^T K[i]^{-1} k[i-1], \quad i = N, \dots, 0 \\ S[N] &= S_f \end{aligned}$$

et le coût optimal associé est $(x^0)^T S[0]x^0$.

Les formules permettant de calculer le gain optimal et la commande en boucle fermée sont, comme dans le cas du temps continu, des formules en temps rétrograde.

Le passage à la limite lorsque N , l'horizon temporel sur lequel est formulé le critère, tend vers l'infini est traité par le résultat suivant

Théorème 43 (Régulateur LQR en temps discret)

Soit un système discret $x[k+1] = Ax[k] + Bu[k]$, $\dim x = n$, $\dim u = 1$, et sa condition initiale x^0 . Soit le critère à minimiser

$$J = \sum_{i=1}^{+\infty} x^T[i]Rx[i] + \sum_{i=0}^{+\infty} u^T[i]Qu[i]$$

avec $N > 0$, P_f et R matrices symétriques positives, Q matrice symétrique définie positive. La loi de commande minimisant ce critère est la loi de feedback

$$u[i] = -(B^T S B + Q)^{-1} B^T S A \times x[i]$$

où S est l'unique solution stabilisante de

$$0 = S - A^T S A + A^T S B (B^T S B + Q)^{-1} B^T S A - R$$

(équation algébrique de Riccati en temps discret pour laquelle l'existence et l'unicité d'une solution stabilisante est garantie sous des hypothèses de commandabilité et d'écriture du critère semblables au cas en temps continu du Théorème 26, i.e. $(A, R^{1/2})$ observable)) et le coût (optimal) associé est $(x^0)^T S[0] x^0$.

D.4 Observabilité, reconstructibilité et filtrage

D.4.1 Observabilité en temps discret

De même qu'en ce qui concerne la commande, certaines subtilités apparaissent dans le domaine de la reconstruction d'état à partir des mesures lorsqu'on considère un système sous sa forme en temps discret. Ainsi, on doit distinguer la propriété d'observabilité et celle de reconstructibilité d'un système (voir ci-dessous).

Considérons l'équation aux différences, où intervient la mesure y de l'état x

$$\begin{aligned} x[k+1] &= Ax[k] + Bu[k] \\ y[k] &= Cx[k] \end{aligned} \tag{D.14}$$

où A est une matrice de dimension $n \times n$, B est une matrice de dimension $n \times m$, et C une matrice de dimension $p \times n$.

Définition 27 (Observabilité en temps discret). *Le système en temps discret (D.14) est dit observable, si, pour tout k , on peut déterminer de manière unique l'état $x[k]$ d'après les mesures futures $y[k], \dots, y[k+n-1]$ et les commandes $u[k], \dots, u[k+n-2]$ futures.*

Définition 28 (Reconstructibilité en temps discret). *Le système en temps discret (D.14) est dit reconstructible, si, pour tout k , on peut déterminer de manière unique l'état $x[k]$ d'après les mesures passées $y[k-n+1], \dots, y[k]$ et les commandes $u[k-n+1], \dots, u[k-1]$ passées.*

La constructibilité n'implique pas l'observabilité, on peut trouver des contre-exemples simples dans [39].

Considérons les équations suivantes, obtenues par un calcul direct depuis un indice k quelconque.

$$\begin{aligned} y[k] &= Cx[k] \\ y[k+1] &= C(Ax[k] + Bu[k]) \\ y[k+2] &= C(A^2x[k] + ABu[k] + Bu[k+1]) \\ &\vdots \\ y[k+n-1] &= C(A^{n-1}x[k] + A^{n-2}Bu[k] + \dots + Bu[k+n-2]) \end{aligned}$$

Il vient, en regroupant, sous la forme de vecteurs :

$$\begin{pmatrix} y[k] \\ \vdots \\ y[k+n-1] \end{pmatrix} = \mathcal{O}x[k] + \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \ddots & & \\ CA^{n-2}B & & & CB \end{pmatrix} \begin{pmatrix} u[k] \\ \vdots \\ u[k+n-2] \end{pmatrix}$$

où \mathcal{O} est la *matrice d'observabilité* de la paire (A, C) déjà rencontrée dans le cas du temps continu au Théorème 28. Dans cette dernière équation, il apparaît directement le résultat suivant

Proposition 13 (Observabilité et reconstructibilité). *Soit un système $x[k+1] = Ax[k] + Bu[k]$, $y[k] = Cx[k]$ avec $\dim x = n$. Ce système est observable si et seulement si la matrice d'observabilité $\mathcal{O} = [C; CA; \dots; CA^{n-1}]$ est de rang plein. Ce système est reconstructible si la matrice d'observabilité est de rang plein.*

Comme ceci est précisé dans [39], les propriétés de commandabilité et de reconstructibilité sont duales, de même que celle d'atteignabilité et d'observabilité.

D.4.2 Filtre de Kalman en temps discret

Lorsqu'on cherche effectivement à reconstruire l'état à partir de la connaissances des entrées et des sorties du système, on peut le faire par un observateur asymptotique, en utilisant, par dualité avec les problèmes de commande, les théorèmes de placement de pôles, où invoquer un principe d'optimalité pour prendre en compte au mieux les incertitudes sur le modèle, les entrées et sur les mesures. Comme dans le cas du temps continu, l'outil permettant ce filtrage optimal au sens stochastique est le filtre de Kalman qui est implémenté, dans le cas considéré, sous forme d'équations aux différences.

Théorème 44 (Filtre de Kalman en temps discret)

Soit un système $x[k+1] = Ax[k] + Bu[k] + D\xi[k]$, $y[k] = Cx[k] + \rho[k]$, initialisé en une variable aléatoire x^0 , où ξ est un bruit gaussien centré de matrice de covariance $\text{cov}(\xi[k], \xi[j]) = \delta_k^j M_\xi$, et ρ est un bruit gaussien centré indépendant de ξ de matrice de covariance $\text{cov}(\rho[k], \rho[j]) = \delta_k^j M_\rho$. Le filtre de Kalman reconstruit, à travers les équations de *propagation* et de *recalage* qui suivent une estimation \hat{x} de l'état x . Notons $\hat{x}_p[k]$ l'état des équations de propagation, et $\hat{x}_r[k]$ l'état des équations de recalage, ainsi que $\Sigma_p[k]$ et $\Sigma_r[k]$ les matrices de covariance pour les étapes de propagation et de recalage, respectivement. Au cours des itérations, l'estimation recherchée de $x[k]$ est simplement $\hat{x}_r[k]$.

$$\text{Initialisation} \begin{cases} \hat{x}_p[0] = \hat{x}_r[0] = E(x^0) \\ \Sigma_p[0] = \Sigma_r[0] = \text{var}(x^0) \end{cases}$$

$$\text{Propagation} \begin{cases} \hat{x}_p[k+1] = A\hat{x}_r[k] + Bu[k] \\ \Sigma_p[k+1] = A\Sigma_r[k]A^T + DM_\xi D^T \end{cases}$$

$$\text{Recalage} \begin{cases} K = \Sigma_p[k+1]C^T (C\Sigma_p[k+1]C^T + M_\rho)^{-1} \\ \hat{x}_r[k+1] = \hat{x}_p[k+1] + K(y[k+1] - C\hat{x}_p[k+1]) \\ \Sigma_r[k+1] = (\Sigma_p[k+1]^{-1} + C^T M_\rho^{-1} C)^{-1} \end{cases}$$

Ces formules sont celles usuellement considérées (voir [73]). Avantageusement, on pourra remplacer la dernière formule dans l'équation de recalage par l'une ou l'autre des deux équations suivantes (*formule de Joseph* [14, 23]) obtenue par la *formule de Sherman-Morrison-Woodbury* appliquée à l'inverse de la matrice $\Sigma_p[k+1]^{-1} + C^T M_\rho^{-1} C$

$$\Sigma_r[k+1] = \Sigma_p[k+1] - \Sigma_p[k+1]C^T (C\Sigma_p[k+1]C^T + M_\rho)^{-1} C\Sigma_p[k+1]$$

qui se réécrit encore

$$\Sigma_r[k+1] = (I_n - KC)\Sigma_p[k+1](I_n - KC)^T + KM_\rho K^T$$

qui garantit que la mise à jour $\Sigma_r[k+1]$ est une matrice symétrique définie positive dès lors que $\Sigma_p[k+1]$ l'est.

Bibliographie

- [1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.
- [2] M. J. Abzug and E. Eugene Larrabee. *Airplane stability and control*. Cambridge Aerospace Series. Cambridge University Press, 2002.
- [3] Hamed M. Al-Rahmani and Gene F. Franklin. Techniques in multirate digital control. In C.T. Leondes, editor, *Digital Control Systems Implementation Techniques*, volume 70 of *Control and Dynamic Systems*, pages 1 – 24. Academic Press, 1995.
- [4] V. Alexeev, E. Galeev, and V. Tikhomirov. *Receuil de Problèmes d'Optimisation*. Editions Mir, Moscou, 1987.
- [5] B. D. O. Anderson, R. R. Bitmead, C. R. Johnson, P. V. Kokotovic, R. L. Kosut, I. M. Y. Mareels, L. Praly, and B. D. Riedle. *Stability of adaptive systems : Passivity and averaging analysis*. MIT Press, 1986.
- [6] A. Andronov, S. Khaikin, and A. Vitt. *Theory of Oscillators*. Dover, 1987. English Translation.
- [7] A. Angot. *Compléments de mathématiques*. Editions de la revue d'optique, Paris, third edition, 1957.
- [8] W. F. III Arnold and A. J. Laub. Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proc. IEEE*, 72 :1746–1754, 1984.
- [9] M. Athans. *The Control Handbook*, chapter Kalman filtering, pages 589–594. CRC Press and IEEE Press, 1996.
- [10] R. Bellman and K. L. Cooke. *Modern elementary differential equations*. Addison-Wesley Publishing Company, 1971.
- [11] J. T. Betts. *Practical Methods for optimal control using nonlinear programming*. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.
- [12] F. Brauer and J. A. Nohel. *The qualitative theory of ordinary differential equations*. W. A. Benjamin, Inc., New York, 1969.
- [13] A. E. Bryson and Y. C. Ho. *Applied Optimal Control*. Ginn and Company, 1969.
- [14] R. S. Bucy and P. D. Joseph. *Filtering for stochastic processes with applications to guidance*. Interscience, 1968.
- [15] H. Cartan. *Théorie élémentaire des fonctions analytiques*. Hermann, 1961.
- [16] C.-T. Chen. *Signals and Systems*. Oxford University Press, 2004.
- [17] K. L. Chien, J. A. Hrones, and J. B. Reswick. On the automatic control of generalized passive systems. *Transactions ASME*, 74 :175–185, 1952.

- [18] G. H. Cohen and G. A. Coon. Theoretical consideration of retarded control. *Trans. A.S.M.E.*, Vol. 75(No. 1) :pp. 827–834, 1953.
- [19] R. F. Curtain and H. J. Zwart. *An Introduction to infinite-Dimensional Linear Systems Theory*. Text in Applied Mathematics, 21. Springer-Verlag, 1995.
- [20] R. Dautray and J.-L. Lions. *Analyse mathématique et calcul numérique pour les sciences et les techniques*, volume 7. Masson, 1988.
- [21] M. Demazure. *Bifurcations and Catastrophes*. Universitext. Springer, 2000.
- [22] V. Ditkin and A. P. Prudnikov. *Formulaire pour le calcul opérationnel*. Masson et Cie., 1967.
- [23] P. Faurre. *Navigation inertielle et filtrage stochastique*. Méthodes mathématiques de l'information. Dunod, 1971.
- [24] F.R. Gantmacher. *Théorie des Matrices : tome 1*. Dunod, Paris, 1966.
- [25] F.R. Gantmacher. *Théorie des Matrices : tome 2*. Dunod, Paris, 1966.
- [26] C. Gasquet and P. Witomski. *Analyse de Fourier et applications*. Masson, 1990.
- [27] J.P. Gauthier and I. Kupka. *Deterministic Observation Theory and Applications*. Cambridge University Press, 2001.
- [28] C. Godbillon. *Géométrie différentielle et mécanique analytique*. Hermann, Paris, 1969.
- [29] K. Gu, V. L. Kharitonov, and J. Chen. *Stability of time-delay systems*. Birkhäuser, 2003.
- [30] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. Springer, New York, 1983.
- [31] J. K. Hale and S. M. Verduyn Lunel. *Introduction to functional differential equations*. Springer-Verlag, 1993.
- [32] S. Haroche and J.M. Raimond. *Exploring the Quantum : Atoms, Cavities and Photons*. Oxford University Press, 2006.
- [33] P. Hartman. *Ordinary Differential Equations*. Birkhäuser, Boston, 1982.
- [34] M.W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press : New-York, 1974.
- [35] P. Horowitz and H. Winfield. *The art of electronics*. Cambridge University Press, 2 edition, 1989.
- [36] W. Hurewicz. *Lectures on ordinary differential equations*. Dover Phoenix Editions, 1990.
- [37] J. Istas. *Introduction aux modélisations mathématiques pour les sciences du vivant*. Number 34 in Mathématiques & Applications. Springer, 2000.
- [38] M.R. James and J.E. Gough. Quantum dissipative systems and feedback control design by interconnection. *Automatic Control, IEEE Transactions on*, 55(8) :1806–1821, 2010.
- [39] T. Kailath. *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1980.
- [40] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction problems. *ASME Journal of Basic Engineering*, pages 95–108, 1961.
- [41] J. Kerckhoff, H. Nurdin, D. Pavlichin, and H. Mabuchi. Designing quantum memories with embedded control : photonic circuits for autonomous quantum error correction. *Phys. Rev. Lett.*, 105 :040502, 2010.
- [42] H.K. Khalil. *Nonlinear Systems*. MacMillan, 1992.

- [43] M. Krstić, I. Kanellakopoulos, and P.V. Kokotovic. *Nonlinear and Adaptive Control Design*. J. Wiley, New-York, 1995.
- [44] O. Lafitte. Distributions et applications. Cours et exercices, École Nationale Supérieure des Mines de Paris, 2005.
- [45] C. Lanczos. *Linear Differential Operators*. Dover Publications, 1997.
- [46] L. Landau and E. Lifshitz. *Mécanique*. Mir, Moscou, 4th edition, 1982.
- [47] Y. Lee, S. Park, M. Lee, and C. Brosilow. PID controller for desired closed-loop responses for SI/SO systems. *AICHE Journal*, Vol. 44(No. 1) :pp. 106–115, 1998.
- [48] W. Leonhard. *Control of Electrical Drives*. Elsevier, 1985.
- [49] W. S. Levine. *The Control Handbook*. CRC Press and IEEE Press, 1996.
- [50] J. Ligou. *Introduction au génie nucléaire*. Presses Polytechniques et Universitaires Romandes, 2 edition, 1997.
- [51] W. Lohmiller and J.J.E. Slotine. On metric analysis and observers for nonlinear systems. *Automatica*, 34(6) :683–696, 1998.
- [52] D. G. Luenberger. *Introduction to Dynamic Systems : Theory, Models, and Applications*. John Wiley & Sons, 2 edition, 1979.
- [53] F. Maisonneuve. Calcul différentiel. Cours et exercices, École Nationale Supérieure des Mines de Paris, 2006.
- [54] F. Maisonneuve. Calcul intégral. Cours et exercices, École Nationale Supérieure des Mines de Paris, 2006.
- [55] F. Maisonneuve. Fonctions d'une variable complexe. Cours et exercices, École Nationale Supérieure des Mines de Paris, 2006.
- [56] I. Mareels, S. Van Gils, J. W. Polderman, and A. Ilchmann. Asymptotic dynamics in adaptive gain control. In P. M. Frank, editor, *Advances in Control, Highlights of ECC'99*, pages 391–449. Springer, 1999.
- [57] Ph. Martin, R. Murray, and P. Rouchon. Flat systems, equivalence and trajectory generation, 2003. Technical Report <http://www.cds.caltech.edu/reports/>.
- [58] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control : stability and optimality. *Automatica*, 36 :789–814, 2000.
- [59] M. B. Milam. *Real-time optimal trajectory generation for constrained systems*. PhD thesis, California Institute of Technology, 2003.
- [60] R. Murray, Z. Li, and S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1993.
- [61] N. E. Nahi. *Estimation theory and applications*. John Wiley & Sons, Inc., 1969.
- [62] Z. J. Palmor. Stability properties of Smith dead-time compensator controller. *Int. J. Control.*, 32(6) :937–949, 1980.
- [63] N. Petit. Optimisation. Notes de cours, École Nationale Supérieure des Mines de Paris, 2004.
- [64] J. F. Poyatos, J. I. Cirac, and P. Zoller. Quantum reservoir engineering with laser cooled trapped ions. *Phys. Rev. Lett.*, 77(23) :4728–4731, December 1996.
- [65] W. J. Rugh and J. S. Shamma. Research on gain scheduling. *Automatica*, 36 :1401–1425, 2000.

- [66] C. Sayrin, I. Dotsenko, X. Zhou, B. Peaudecerf, Th. Rybarczyk, S. Gleyzes, P. Rouchon, M. Mirrahimi, H. Amini, M. Brune, J.M. Raimond, and S. Haroche. Real-time quantum feedback prepares and stabilizes photon number states. *Nature*, 477 :73–77, 2011. doi :10.1038/nature10376.
- [67] R. Sepulchre, M. Janković, and P. Kokotović. *Constructive nonlinear control*. Springer Verlag, 1997.
- [68] G. J. Silva, A. Datta, and S. P. Bhattacharyya. *PID controllers for time-delay systems*. Birkhäuser, 2005.
- [69] L. Sinègre. *Etude des instabilités dans les puits activés par gas-lift*. PhD thesis, École des Mines de Paris, 2006.
- [70] O. J. M. Smith. Closer control of loops with dead time. *Chemical Engineering Progress*, 53(5) :217–219, 1958.
- [71] E. Sontag. *Mathematical Control Theory*. Springer Verlag, 1990.
- [72] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, 1971.
- [73] R. F. Stengel. *Optimal control and estimation*. Dover Publications, 1994.
- [74] K. J. Åström and T. Hägglund. *PID Controllers : Theory, Design, and Tuning*. Instrument Society of America, 1995.
- [75] A. Tikhonov, A. Vasil'eva, and A. Sveshnikov. *Differential Equations*. Springer, New York, 1980.
- [76] M. Unser. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88(4) :569 –587, 2000.
- [77] H.M. Wiseman and G.J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, 2009.
- [78] K. Yosida. *Operational Calculus*. Springer, Berlin, 1999.
- [79] J. G. Ziegler and N. B. Nichols. Optimum settings for automatic controllers. *Trans. A.S.M.E.*, Vol. 64 :pp. 759–765, 1942. Available from www.driedger.ca.
- [80] J. G. Ziegler and N. B. Nichols. Process lags in automatic-control circuits. *Transp. A.S.M.E.*, Vol. 65 :pp. 433–444, 1943.

Index

- anti-windup, 47, 49
- approximation adiabatique, 39
- approximation quasi-statique, 39
- approximation séculaire, 39
- atome à trois niveaux, 101
- atteignabilité en temps discret, 192
- attracteur, 29
- backstepping, 137
- bassin d'attraction, 42, 173
- bouclage dynamique, 51
- bouclage statique régulier, 109, 110
- boucle fermée, 51
- boucle ouverte, 51
- calcul différentiel intrinsèque, 134
- capteurs logiciels, 139
- centre, 20
- champ de vecteurs, 34, 133
- CNS de stabilité asymptotique d'un système linéaire stationnaire, 17
- CNS de stabilité d'un système linéaire stationnaire, 18
- col, 20
- commandabilité, 97, 105, 115
- commandabilité en temps discret, 192
- commandabilité linéaire, 107
- commandabilité non linéaire, 104
- commande, 7
- commande adaptative, 25, 160, 161
- commande en boucle fermée, 100
- commande en boucle ouverte, 105
- commande linéaire quadratique, 97, 115, 117
- commande linéaire quadratique en temps discret, 195
- commande LQR, 115, 123
- commande LQR en temps discret, 196
- commande modale, 150
- consigne, 47
- contrôle en boucle fermée, 100
- contrôle en boucle ouverte, 99
- contrôleur P, 81
- contrôleur PI, 45, 71, 79
- contraction, 162
- contrainte finale, 123
- contrôle hiérarchisé, 61
- convergence exponentielle, 18, 127
- convertisseur analogique-numérique, 185, 186
- convertisseur numérique-analogique, 185
- critère d'atteignabilité en temps discret, 193
- critère d'observabilité, 147
- critère d'observabilité en temps discret, 197
- critère de Bendixon, 34
- critère de commandabilité, 110
- critère de Jury, 192
- critère de Nyquist, 76, 81
- critère de Routh, 23
- critère du revers, 79
- crochet de Lie, 134
- cycle limite, 33, 35
- dépendance de la solution en la condition initiale, 11
- distinguabilité, 144
- ensemble invariant de LaSalle, 29
- ensemble positivement invariant, 28, 53, 172, 174
- entrée, 7
- entrelacement, 22
- équation algébrique de Riccati en temps discret, 196
- équation de Lyapounov, 25, 155
- équation de Riccati algébrique, 126, 127, 129, 130
- équation de sortie, 7
- équation différentielle de Riccati, 122–124, 127
- équivalence statique, 132
- estimation, 141, 146
- état, 7, 51
- état adjoint, 118, 121
- excitation sinusoïdale, 68

- existence d'une solution au problème de Cauchy, 169
 existence et unicité des solutions d'une équation différentielle, 11
 exponentielle de matrice, 16
 exposants caractéristiques, 33
 facteur d'amortissement, 60
 feedback, 8, 47, 97, 99
 feedback optimal, 123
 feedforward, 47, 58–60, 113, 124
 filtrage, 141
 filtrage de consigne, 59
 filtre de Kalman, 139
 filtre de Kalman en temps discret, 198
 fonction Lipschitz, 166
 fonction Lipschitz vectorielle, 166
 fonction de transfert, 65
 fonction de Lyapounov, 173
 fonction de Lyapounov, 28
 fonction de transfert, 66, 70
 fonction méromorphe, 77
 fonction non bornée radialement, 28, 29, 173
 fonctions méromorphes, 76
 forme d'état, 51
 forme canonique d'un système linéaire, 21
 forme d'état, 7, 70
 forme d'état canonique, 72
 forme d'état canonique minimale, 71
 forme de Brunovsky, 97, 110, 114, 115, 133, 150
 forme de Jordan, 16
 forme normale, 110
 forme standard du théorème de Tikhonov, 42
 formule de Joseph, 198
 formule de Poisson, 187
 formule de reconstruction de Shannon-Whittaker, 187
 formule de Sherman-Morrison-Woodbury, 198
 foyer instable, 20, 37
 foyer stable, 20
 fraction rationnelle causale, 70
 fraction rationnelle propre, 70
 fréquence d'échantillonage, 185
 gain d'anti-emballement, 52
 gain intégral, 48
 identification, 141
 indices de commandabilité, 109, 110
 intégrale première, 105–108
 intégrale première triviale, 106
 invariance de LaSalle, 29, 174
 invariant chimique, 106
 Lagrangien, 118, 119
 lieu de Nyquist, 81, 83, 86, 87, 131
 linéarisation par bouclage dynamique, 136
 linéarisation par injection de sortie, 161
 loi d'Arrhenius, 106
 marge de gain, 85
 marge de phase, 82, 84, 85
 marges de robustesse, 65, 76
 marges de stabilité du régulateur LQR, 130
 matrice d'observabilité, 147, 197
 matrice de commandabilité, 107, 108, 110, 192
 matrice de transfert, 70
 matrice de transition, 155
 matrice Hurwitz, 22, 25, 42, 114
 mesure, 7
 moteur à courant continu, 140
 moyennisation, 39, 178
 multiplicateurs de Lagrange, 117
 nœud instable, 20, 37
 nœud stable, 20
 observabilité, 140
 observabilité en temps discret, 196
 observabilité globale, 144
 observabilité locale, 144, 145
 observable, 126
 observateur, 8, 146
 observateur asymptotique, 140, 141, 149
 observateur réduit de Luenberger, 149
 observateur-contrôleur, 103, 140, 150
 orbite hétérocline, 33
 orbite homocline, 33
 orbite périodique, 35
 oscillateur de Van der Pol, 37
 oscillateur harmonique, 14
 oscillateurs, 98
 période d'échantillonnage, 48
 perturbation, 51
 perturbations singulières, 39
 phénomène d'aliasing, 187

- pic de résonance, 68
placement de pôles, 97, 114, 127, 149
placement de pôles en temps discret, 193
placement de pôles pour l'observateur, 142
plan de phases, 19, 53
planification de trajectoire en temps discret, 194
planification de trajectoires, 97–99, 104, 105, 114, 123, 125
point d'équilibre, 8
point d'équilibre asymptotiquement stable, 65
point d'équilibre globalement asymptotiquement stable, 14, 17, 30, 173
point d'équilibre hyperbolique, 26, 35, 44, 178, 180
point d'équilibre instable, 14
point d'équilibre localement asymptotiquement stable, 14, 27
point d'équilibre stable (au sens de Lyapounov), 14
point stationnaire, 8
pôle, 67
pôle d'une fonction méromorphe, 77
pôle instable, 77
pôles d'une fonction de transfert, 70
pôles dominants, 87
polynôme caractéristique, 16, 21
polynôme Hurwitz, 22
portrait de phases, 19–21
pré-compensation, 47, 58
prédicteur de Smith, 92, 93
principe de séparation, 139, 151
principe de superposition, 9
problème aux deux bouts, 121
problème de Cauchy, 10, 11, 105, 121, 165, 168, 169
pulsation de coupure, 60
quasi-polynômes, 89
réalisation d'une fonction de transfert discrète, 190
réacteur exothermique, 105
réalisation, 71
réalisation d'un transfert rationnel, 71
reconstructibilité en temps discret, 196
réduction de Jordan, 16
réentrée atmosphérique, 116
régime asymptotique forcé, 68
réglages de Ziegler-Nichols, 90, 91
régulateur LQR, 125, 126
régulateur PI, 47, 48, 66, 76, 83, 88
régulateur PI avec anticipation, 59
régulateur PID, 88
rejet de l'erreur statique, 71
réponse à un échelon, 90
réponse à une entrée sinusoïdale, 71
réponse inverse, 94
résolution de l'équation de Riccati algébrique, 129
résonance, 68
retard critique, 76, 82, 85
retour d'état, 114
retour de sortie, 8
retour dynamique de sortie, 74
rétro-action, 8
robuste, 32
robustesse, 65, 66, 93
robustesse au retard, 88
robustesse par rapport aux dynamiques négligées, 45
robustesse paramétrique, 32, 46
schéma blocs, 73
simplifications pôles-zéros, 68, 69, 71
solution approchante, 165, 167
sortie, 7
sortie de Brunovsky, 97, 110, 112
sortie de Brunovsky non linéaire, 132
sous-variété invariante, 42
stabilisation, 97, 99, 100, 114
stabilité, 13, 14
stabilité asymptotique, 14, 27, 70, 171, 173
stabilité EBSB, 70
stabilité en temps discret, 191
suivi asymptotique, 115
suivi asymptotique de trajectoire, 99
suivi de trajectoire, 97, 99, 100, 115
système nominal, 65
système à non minimum de phase, 94
système autonome, 34, 55
système commandé, 97
système commandable, 105
système discrétisé exact, 189
système instationnaire, 8
système libre, 8
système linéaire à coût quadratique, 121
système linéaire commandable, 114

système linéaire instationnaire, 115, 123
système linéaire stationnaire, 141
système linéarisé tangent, 8, 15, 32, 65, 75
système linéarisable par bouclage statique, 132,
 133
système perturbé, 65
système plan, 33
système sous-actionné, 104
système sous-déterminé, 97
système stationnaire, 8, 47
systèmes de Liénard, 36
systèmes plats, 136
système du second ordre, 61

terme intégral, 48
théorème d'invariance de LaSalle, 29
théorème de Cauchy, 78
théorème de Cauchy-Lipschitz, 11
théorème de Hermite-Biehler, 22
théorème de Poincaré, 33
théorème de Poincaré-Bendixon, 35
théorème de Shannon, 186
théorème de Sylvester, 25
théorème de Tikhonov, 42
théorie des bifurcations et des catastrophes, 65
trajectoire, 105, 114
trajectoire d'un système commandé, 105
trajectoire de référence, 115
trajectoire en boucle ouverte, 114
transformée de Fourier, 186
transformée de Laplace, 65
transformée en z , 190

unicité de la solution au problème de Cauchy, 168
utilisation pratique de la commande LQR, 129

zéro d'une fonction méromorphe, 77
zéros d'une fonction de transfert, 70