

Identification des systèmes linéaires

Amine HADJ ABDELKADER
m_hadjabdelkader@mail.univ-tlemcen.dz

Document de cours pour les étudiants de M1 en ‘Contrôle des Processus’

Ce document est basé principalement sur le cours de Y. TANGUY, Supelec – Paris ©.
Merci d’envoyer vos remarques, commentaires ou questions sur l’e-mail de l’auteur.

Identification des Systèmes Linéaires

Amine, HADJ ABDELKADER

Octobre 2007

Table des matières

1 Aspect généraux de l'identification	5
1.1 Présentation du problème	5
1.1.1 Essai de définition de l'identification expérimentale	5
1.1.2 Utilité d'un modèle	6
1.1.3 Obtention d'un modèle de représentation	7
1.2 Systèmes dynamiques "objet" et "modèle" (aspect structurel)	7
1.2.1 Description de l'objet en vue d'une modélisation	7
1.2.2 Les représentations mathématiques dont on dispose	8
1.2.3 Modèles de conduite et de simulation	11
1.3 Critère d'équivalence "objet-modèle" (aspect paramétrique)	14
1.3.1 Définition de l'erreur	14
1.3.2 Interprétation déterministe de l'erreur de sortie	15
1.3.3 Validité d'un modèle - Nature des signaux d'entrée	17
1.3.4 L'initialisation du modèle	21
1.3.5 L'objet est inséré dans une boucle de régulation	21
1.4 Algorithme d'identification	23
2 Méthodes directes	25
2.1 Analyse harmonique	25
2.2 Analyse indicielle	27
2.3 Analyse impulsionnelle	28
2.4 Approximation par un modèle paramétrique	28
2.4.1 A partir de la réponse harmonique	28
2.4.2 A partir de la réponse indicielle - Méthode de Strejc	29
2.4.3 A partir de la réponse indicielle - Méthode de Broïda	33
2.4.4 Processus Intégrateur	33
2.5 Obtention d'un modèle non paramétrique	35
2.5.1 Problème préliminaire : Passage d'une réponse impulsionnelle à une réponse harmonique	35
2.5.2 Principe du traitement numérique des données par corrélation	37
2.5.3 Utilisation des séquences binaires pseudo aléatoires	40
2.5.4 Méthode des moindres carrés appliquée à l'obtention d'une séquence de pondération (déconvolution numérique)	45

TABLE DES MATIÈRES

2.6	Obtention d'un modèle paramétrique	47
2.6.1	Notion d'ordre minimal	47
2.6.2	Recherche de l'ordre minimal par factorisation de la matrice de Hankel	50
2.6.3	Recherche de l'ordre par essais successifs	51
2.6.4	Méthode des moindres carrés	51
3	Méthodes Itératives	55
3.1	Méthodes stochastiques	55
3.1.1	Principe	55
3.1.2	Méthode du maximum de vraisemblance (Astrom et Bohlin 1965)	57
3.1.3	Méthode des moindres carrés généralisés (Clarcke 1967)	61
3.1.4	Conclusion	63
3.2	Méthodes du modèle (Méthode déterministe)	63
3.2.1	Méthode du modèle appliquée à la détermination d'un modèle paramétrique	64
3.2.2	Méthode du modèle appliquée à la détermination d'une séquence de pondération	69
4	Filtre de Kalman	71
4.1	Principe	71
4.2	L'algorithme discret du filtre de Kalman	74
4.3	Schéma structurel du Filtre de Kalman	74
4.4	Filtre de Kalman sous-optimal du régime permanent	76

Chapitre 1

Aspect généraux de l'identification

1.1 Présentation du problème

1.1.1 Essai de définition de l'identification expérimentale

Identifier un système dynamique réel (appelé **objet**) c'est caractériser un autre système (appelé **modèle**), à partir de la connaissance expérimentale des entrées et sorties de manière à obtenir identité de comportement.

Le modèle peut être un système physique (au sens de simulateur analogique ou numérique et de modèle réduit), ou bien un système abstrait (modèle mathématique, i.e. système d'équations algébriques ou différentielles).

Dans la suite du cours, nous rechercherons en fin de compte un modèle mathématique, recherche qui peut cependant comporter comme étape une simulation (i.e. le passage par un modèle physique).

Exemple 1 Soit le moteur d'asservissement donné par la figure 4.2. Un modèle

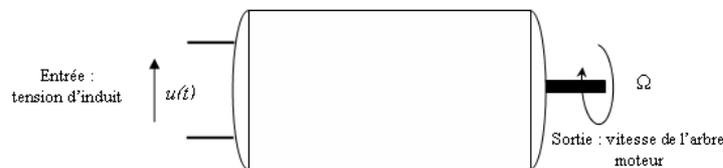


FIG. 1.1 – Moteur d'asservissement à courant continu

1.1. PRÉSENTATION DU PROBLÈME

physique pour cet objet peut être donné par la figure 1.2.

Et un modèle mathématique pour ce même objet est donné par l'équation 3.3.

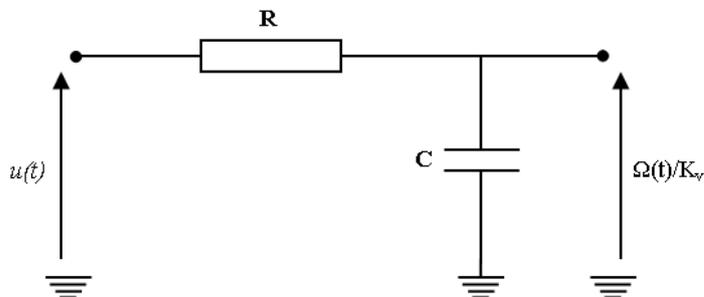


FIG. 1.2 – Modèle physique

$$\Omega(t) + \tau \frac{d\Omega(t)}{dt} = K_v u(t) \quad (1.1)$$

1.1.2 Utilité d'un modèle

Un modèle a pour but :

- d'accroître la connaissance d'un système physique sous l'aspect interne ou microscopique. Ce modèle sera alors dit de **connaissance**,
- de représenter le comportement global (ou macroscopique) d'un système physique. Le modèle sera dit de **représentation**.

Le premier modèle est nécessaire à la construction et au dimensionnement (i.e. détermination des dimensions) d'un système industriel (exemple. fabrication d'un véhicule, d'une chaudière, d'une unité de production,...). Dans ce modèle les coefficients caractérisant la structure de l'objet sont explicités.

C'est le modèle de représentation qui va nous intéresser dans la suite du cours. La nécessité d'un tel modèle apparaît dès que l'on désire :

- commander un système dynamique (réaliser une commande optimale ou seulement régler un régulateur à trois actions). On a recours alors à un modèle dit de **conduite**,
- surveiller le comportement d'un système afin de déceler les anomalies par rapport à un modèle de référence.

Prenons à nouveau le moteur d'asservissement comme exemple :

Un modèle de connaissance est :

$$\Omega(t) + \tau_1 \frac{d\Omega(t)}{dt} = K_v u(t) \quad \text{avec} \quad \begin{cases} \tau_1 = \frac{R \cdot J}{aR + \phi_0^2} \\ K_v = \phi_0 \frac{1}{aR + \phi_0^2} \end{cases}$$

Un modèle de représentation est :

$$\Omega(t) + \tau_1 \frac{d\Omega(t)}{dt} = K_v u(t) \quad \text{avec} \quad \begin{cases} \tau_1 = 20 \text{ ms} \\ K_v = 50 \text{ (rd/s) / v} \end{cases}$$

1.1.3 Obtention d'un modèle de représentation

Un modèle mathématique de représentation est défini par :

- la nature des équations (par exemple : équations différentielles d'ordre n , à coefficients constants) caractérisant la **structure** du modèle,
- un ensemble de **valeurs numériques** intervenant dans cette structure.

Une méthode classique d'obtention du modèle consiste à mettre le système en "équation" à partir de lois physiques connues. Mais le modèle global est en général trop complexe (pour un but de représentation). Des simplifications (obtenues en linéarisant, en négligeant certains termes) peuvent alors conduire à la définition de la structure d'un modèle de représentation.

Les valeurs numériques caractérisant cette structure sont ensuite déterminées (à partir d'essais expérimentaux) par des algorithmes que nous nous proposons de décrire dans la suite.

Une autre démarche revient à considérer le système globalement comme "boite noire" et à rechercher expérimentalement les relations "entrées-sorties". Mais cette méthode peut conduire à un échec si les informations dont on dispose a priori ne sont pas suffisantes.

1.2 Systèmes dynamiques "objet" et "modèle" (aspect structurel)

1.2.1 Description de l'objet en vue d'une modélisation

d'un point de vue externe, un objet se présente comme une "boite noire" admettant des signaux d'entrée (ou de commande) et des signaux de sortie (ou d'observation). Ces signaux (continus ou bien discrets) sont accessibles à la mesure (cf. figure 1.3).

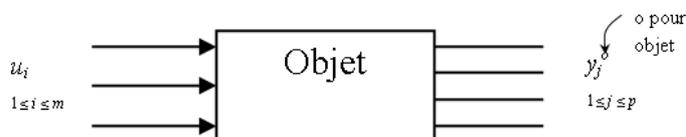


FIG. 1.3 – L'objet

1.2. SYSTÈMES DYNAMIQUES "OBJET" ET "MODÈLE" (ASPECT STRUCTUREL)

Une analyse plus fine permet de voir que les sorties y_j^o ne sont pas les effets des seuls signaux u_i . D'où le nouveau schéma sur la figure 1.4.

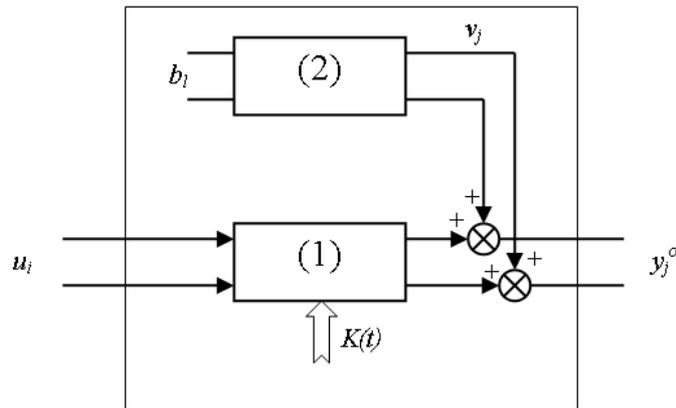


FIG. 1.4 – Détails de l'objet

- La "boite" notée (1) caractérise l'action des grandeurs de commande observables sur les sorties y_j^o . Cette action n'est pas invariante dans le temps et la structure de (1) est susceptible d'évoluer en fonction de perturbations symbolisées par $K(t)$.
- La boite (2) caractérise l'action perturbatrice sur l'état de (1). L'effet de ces perturbations, dans un but de simplification, a été globalement ramené en sortie par addition des v_j . Les causes de ces perturbations notées b_l sont le plus souvent fictives et non mesurables (on note d'ailleurs que si b_l était mesurable, b_l serait une entrée).

Dans la suite nous utiliserons, pour décrire l'objet, le symbole donné par la figure 1.5.

1.2.2 Les représentations mathématiques dont on dispose

Dans cette section, on va rappeler les différentes représentations des systèmes dynamiques linéaires et stationnaires susceptibles de décrire le comportement d'un objet.

De telles représentations nous limitent donc d'emblée à la description dynamique

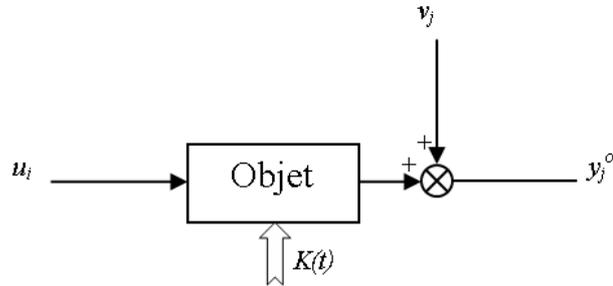


FIG. 1.5 – Symbole de l'objet à identifier

des objets quasi-linéaires (ou linéarisés autour d'un point de fonctionnement) et quasi-stationnaires (les perturbations de structure $K(t)$ sont à variations lentes).

Représentation des systèmes déterministes linéaires, stationnaires et causaux

Dans les tableaux suivants, nous allons décrire seulement les systèmes à une entrée et une sortie.

Représentations non paramétriques

Cas continu	Cas discret
Réponse impulsionnelle	
$h(\theta)$	$a(k)$
$y(t) = \int_0^{\infty} h(\theta)u(t - \theta)d\theta$	$y(k) = \sum_{l=0}^{\infty} a(l)u(k - l)$
Réponse indicielle	
$\gamma(t) = \int_0^t h(\theta)d\theta$	$\gamma(k) = \sum_{l=0}^k a(l)$
Réponse harmonique	
$H(f) = \frac{Y(f)}{U(f)}$	$A(\nu) = \frac{Y(\nu)}{U(\nu)}$
$H(f) = \int_0^{+\infty} h(\theta) \exp(-2\pi j f \theta) d\theta$	$A(\nu) = \sum_{l=0}^{\infty} a(l) \exp(-2\pi j l \nu)$

Les relations entrée/sortie précédentes n'ont en fait de sens que lorsque la réponse impulsionnelle est intégrable (hypothèse de stabilité entrée-bornée, sortie-bornée). Dans ce cas, $h(\theta)$ (ou $a(k)$) tend vers zéro lorsque $\theta \rightarrow \infty$ ($k \rightarrow \infty$).

L'instant T_0 (ou N) à partir duquel $h(\theta)$ (ou $a(k)$) est négligeable est appelé mémoire du système.

Représentations paramétriques

Cas continu	Cas discret
Equations différentielles, aux différences	
$y(t) + b_1 y'(t) + \dots + b_n y^{(n)}(t)$ $= a_0 u(t) + \dots + a_n u^{(n)}(t)$	$y(k) + \beta_1 y(k-1) + \dots + \beta_n y(k-n)$ $= \alpha_0 u(k) + \dots + \alpha_n u(k-n)$
Fonctions de transfert	
$H(p) = \frac{\sum_{i=0}^n \alpha_i p^i}{1 + \sum_{j=1}^n b_j p^j}$	$A(z) = \frac{\sum_{i=0}^n \alpha_i z^{-i}}{1 + \sum_{j=1}^n \beta_j z^{-j}}$
Equations d'état	
$\dot{x} = \mathbf{A}x + \mathbf{B}u$ $y = \mathbf{C}x$	$x(k+1) = \mathbf{F}x(k) + \mathbf{G}u(k)$ $y(k) = \mathbf{H}x(k)$
\mathbf{A}, \mathbf{F} : matrices ($n \times n$); \mathbf{B}, \mathbf{G} : matrices ($n \times 1$); \mathbf{C}, \mathbf{H} : matrice ($1 \times n$)	

Représentations stochastiques

Les équations précédentes ne peuvent décrire le comportement d'un objet que lorsque celui-ci est faiblement perturbé. Si ce n'est pas le cas, une représentation stochastique peut s'avérer nécessaire. Pour un système discret, on écrira de manière très générale :

$$\begin{cases} x(k+1) = \mathbf{F}x(k) + \mathbf{G}u(k) + b(k) \\ y(k) = \mathbf{H}x(k) + w(k) \end{cases} \quad (1.2)$$

où $b(k)$ et $w(k)$ sont des bruits blancs discrets gaussiens de variance inconnue à priori.

Afin de réduire le nombre de paramètres à identifier la théorie du filtrage de Kalman fournit une représentation simplifiée du système abstrait précédent 1.2 :

$$\begin{cases} \hat{x}(k+1) = \mathbf{F}\hat{x}(k) + \mathbf{G}u(k) + \mathbf{k}\nu(k) \\ y(k) = \mathbf{H}\hat{x}(k) + \nu(k) \end{cases} \quad (1.3)$$

avec $\hat{x}(k)$ l'estimée de l'état $x(k)$ et $\nu(k)$ une séquence de bruit blanc gaussien (Innovation).

L'écriture en variables d'état se transpose en l'équations aux différences suivantes :

$$y(k) + \beta_1 y(k-1) + \dots + \beta_n y(k-n) = \alpha_0 u(k) + \dots + \alpha_n u(k-n) + \nu(k) + c_1 \nu(k-1) + \dots + c_n \nu(k-n)$$

soit :

$$\left(1 + \sum_{i=0}^n \beta_i z^{-i}\right) Y(z) = \left(\sum_{j=0}^n \alpha_j z^{-j}\right) U(z) + \left(1 + \sum_{l=1}^n c_l z^{-l}\right) \nu(z)$$

$$Y(z) = \left(\frac{\sum_{j=0}^n \alpha_j z^{-j}}{1 + \sum_{i=0}^n \beta_i z^{-i}} \right) U(z) + \left(\frac{1 + \sum_{l=1}^n c_l z^{-l}}{1 + \sum_{i=0}^n \beta_i z^{-i}} \right) \nu(z)$$

Un schéma synoptique de cette représentation peut être donnée par la figure 1.6.

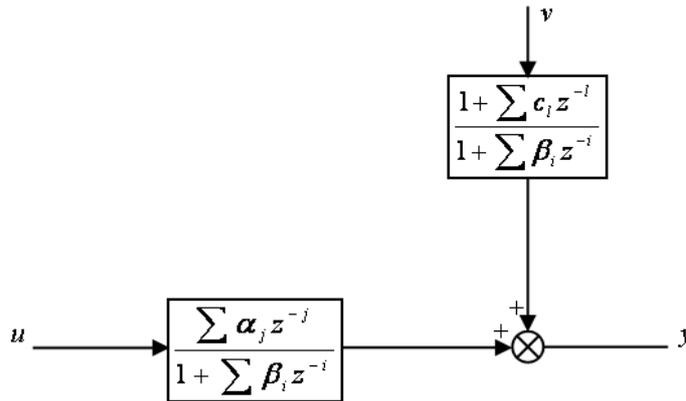


FIG. 1.6 – Représentation stochastique

1.2.3 Modèles de conduite et de simulation

Ces modèles, étant eux-mêmes des systèmes dynamiques (réels ou abstraits), admettent une représentation mathématique du type de celles décrites précédemment. Un certain nombre de questions se pose alors :

Le modèle est continu ou bien discret ? L'objet est en général continu. Il semble naturel d'en établir un modèle lui aussi continu. En pratique, il en est parfois autrement. En effet :

Dans le cas d'un modèle de simulation :

- Le simulateur est analogique et un modèle continu s'impose.
- Le simulateur est numérique et les deux types de modèles sont envisageable en considérant le schéma d'équivalence sur la figure 1.7.

Dans le cas d'un modèle de conduite :

- Le régulateur est analogique et un modèle continu s'impose.
- Le régulateur est numérique et un modèle discret est nécessaire, si toutefois on désire optimiser l'encombrement des voies de transmissions et la charge du calculateur (c'est-à-dire ne pas transposer systématiquement les

1.2. SYSTÈMES DYNAMIQUES "OBJET" ET "MODÈLE" (ASPECT STRUCTUREL)

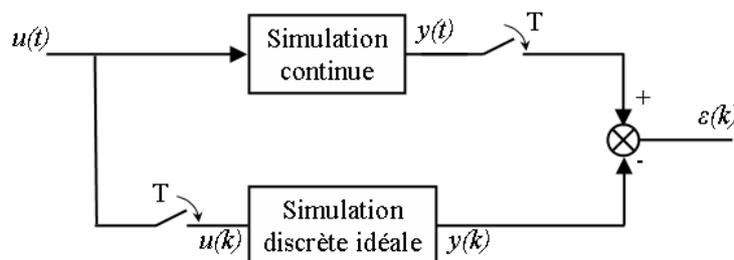


FIG. 1.7 – Schéma d'équivalence

régulateurs continus en discrets, transposition justifiée uniquement *pour un pas d'échantillonnage très fin*). La période d'échantillonnage choisie est naturellement la même en identification qu'en commande ; et tout se passe comme si on disposait de l'objet discret de la figure 1.8.

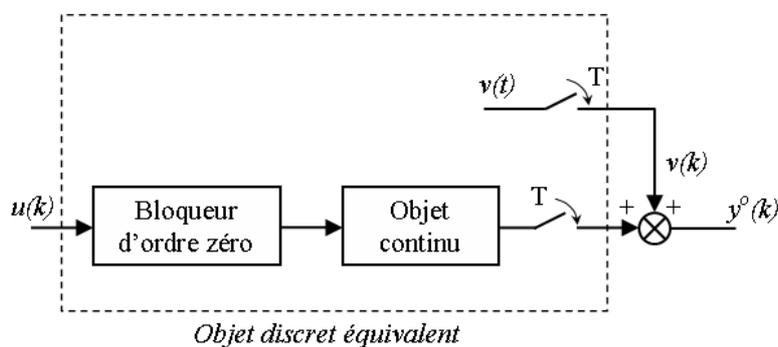


FIG. 1.8 – Objet discret équivalent

Notons toutefois qu'au cours de l'établissement d'un modèle continu, le recours à un calculateur numérique peut s'avérer nécessaire (algorithmes de programmation non linéaire P.N.L, transformée de Fourier,...). Dans ce cas, la période d'échantillonnage est choisie suffisamment fine de manière à éviter tout recouvrement de spectre et demeurer muette vis à vis de l'exploitation des résultats.

Le modèle est paramétrique ou non paramétrique ? D'une manière générale, les modèles non paramétriques exigent peu de connaissance a priori de

la structure de l'objet, sont applicables à la représentation de tous les objets linéaires et quasi-stationnaires et en particulier aux processus physiques dits à "paramètres répartis".

Les modèles paramétriques nécessitent la donnée de l'ordre n (ordre fini) et ne sont justifiés rigoureusement que dans le cas de processus dits "à constantes localisées".

Cependant, tout modèle non paramétrique n'est pas simulable directement à partir des relations entrée/sortie données dans 1.2.2. Par exemple, la représentation par une séquence de pondération $a(k)$ (à priori infinie) ne conduit à un modèle de simulation qu'après troncature de la séquence $a(k)$ aux N premiers termes où N est la mémoire approximative du système ($a(k) = 0$ pour $k > N$)

De plus, la synthèse d'une loi de commande n'est pas facilitée par de tels modèles (les théories utilisent le plus souvent un modèle paramétrique).

Nous verrons par la suite qu'un modèle non paramétrique conduit à des algorithmes d'identification très simples et son établissement sera souvent considéré comme une étape permettant de concentrer les résultats d'une expérience.

Le modèle est déterministe ou stochastique Dans la plupart des cas pratiques la modélisation qui nous intéresse est celle de la relation déterministe liant $y^o - v$ à u . Les perturbations d'état v sont alors considérées comme une gêne importante vis à vis de l'identification de cette relation. Le recours à une représentation stochastique a finalement pour but d'éliminer le biais causé par les perturbations dans l'estimation du modèle déterministe.

Dans la suite, le modèle déterministe est symbolisé par la figure 1.9, et la représentation stochastique par la figure 1.10.

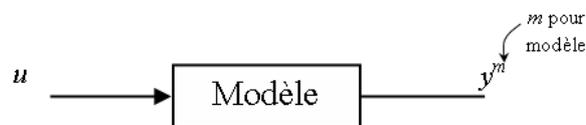


FIG. 1.9 – Représentation déterministe

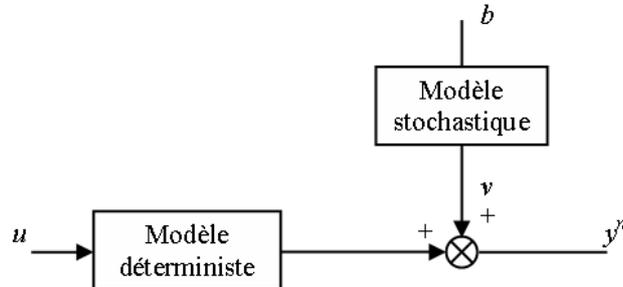


FIG. 1.10 – Représentation stochastique

1.3 Critère d'équivalence "objet-modèle" (aspect paramétrique)

lorsque la structure du modèle a été déterminée, il reste à fixer les valeurs numériques des coefficients intervenant dans cette structure. L'ensemble de ces coefficients est noté vecteur paramètre \mathbf{P}^m que le modèle soit paramétrique ou non.

Exemple 2 Soit un système linéaire discret (objet) de mémoire approximative N . Un modèle "non paramétrique", par séquence de pondération, est le suivant :

$$y^m(k) = \sum_{i=0}^{N-1} a^m(i)u(k-i)$$

Le vecteur paramètre est $\mathbf{P}^m = \begin{bmatrix} a^m(0) \\ a^m(1) \\ \vdots \\ a^m(N-1) \end{bmatrix}$

1.3.1 Définition de l'erreur

Dans le but de définir des algorithmes d'estimation du vecteur paramètre, il est courant de formuler une équivalence "objet-modèle" pour une fonctionnelle de type

$$\|\epsilon\| = \int_0^{t_0} \epsilon^2(t)dt \quad \left(\sum_1^M \epsilon^2(k) \right)$$

où ϵ est l'erreur de sortie donné par $\epsilon = y^m - y^o$. (cf. figure 1.11)

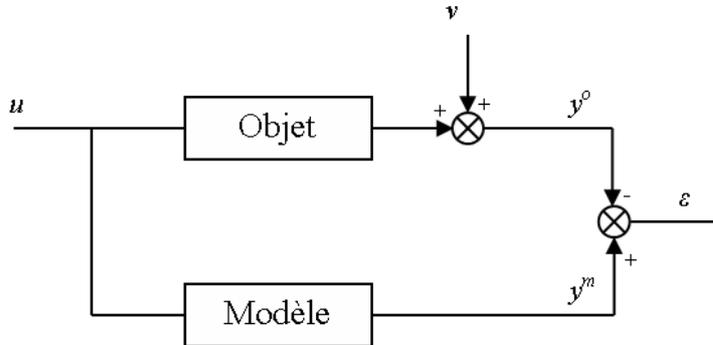


FIG. 1.11 – L'erreur de sortie

Dans une interprétation déterministe, l'erreur de sortie caractérise la différence de comportement (d'un point de vue externe) entre l'objet et le modèle.

1.3.2 Interprétation déterministe de l'erreur de sortie

Supposons (problème abstrait) que le modèle est l'objet aient une structure identique. (Ce qui permet de définir un vecteur "paramètre" objet \mathbf{P}^o). (cf. figure 1.12).

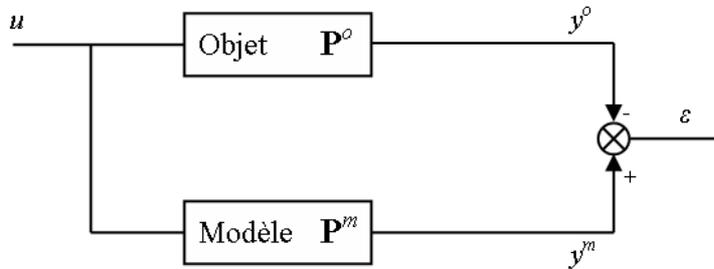


FIG. 1.12 – Erreur de sortie (Interprétation)

On peut alors définir une distance objet -modèle par :

$$D(O, M) = \sum_1^N (P_i^m - P_i^o)^2 = \sum_1^N (\Delta P_i)^2$$

1.3. CRITÈRE D'ÉQUIVALENCE "OBJET-MODÈLE" (ASPECT PARAMÉTRIQUE)

Cette distance ne peut évidemment pas être explicitée car \mathbf{P}^o est inconnu (**distance dite de structure**). (cf. figure 1.13)

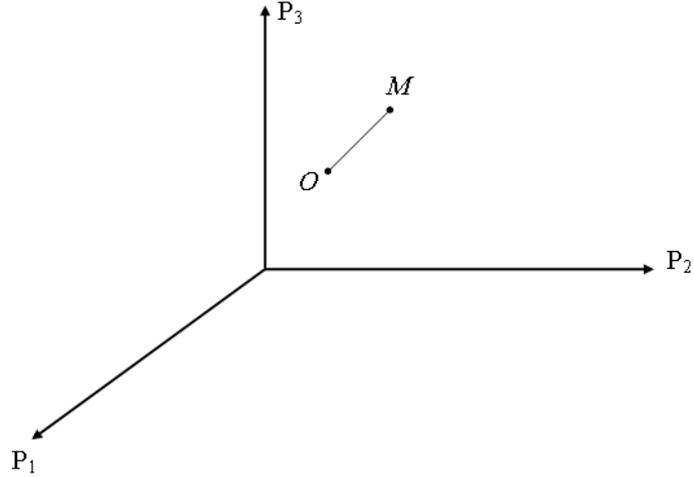


FIG. 1.13 – Distance de structure

La question suivante se pose : l'annulation de la norme quadratique de l'erreur de sortie entraîne-t-elle $D(O, M) = 0$?

En toute généralité $\|\varepsilon\|$ n'induit pas une distance paramétrique mais on peut, sous certaines conditions, établir une équivalence en se plaçant au voisinage de l'optimum :

Développons $\|\varepsilon\|$ au second ordre par rapport aux P_i^m au voisinage de $P_i^m = P_i^o$:

$$\|\varepsilon\|_{\mathbf{P}^o + \Delta \mathbf{P}} \simeq \|\varepsilon\|_{\mathbf{P}^o} + \sum_{i=1}^N \left(\frac{\partial \|\varepsilon\|}{\partial P_i^m} \right)_{P_i^m = P_i^o} \Delta P_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{\partial^2 \|\varepsilon\|}{\partial P_i^m \partial P_j^m} \right)_{\mathbf{P}^m = \mathbf{P}^o} \Delta P_i \Delta P_j \quad (1.4)$$

en posant :

$$(\nabla \|\varepsilon\|)_{\mathbf{P}^o}^T = \left[\dots, \dots, \left(\frac{\partial \|\varepsilon\|}{\partial P_i^m} \right)_{P_i^m = P_i^o}, \dots, \dots \right]$$

et :

$$H = [h_{ij}]_{\mathbf{P}^o} \quad \text{avec} \quad h_{ij} = \left(\frac{\partial^2 \|\varepsilon\|}{\partial P_i^m \partial P_j^m} \right)_{\mathbf{P}^m = \mathbf{P}^o}$$

(H est appelé le Hessien, ∇ le gradient).

d'où :

$$\|\varepsilon\|_{\mathbf{P}^o + \Delta\mathbf{P}} \simeq \|\varepsilon\|_{\mathbf{P}^o} + (\nabla \|\varepsilon\|)_{\mathbf{P}^o}^T \cdot \Delta\mathbf{P} + \frac{1}{2} (\Delta\mathbf{P})^T \cdot H \cdot (\Delta\mathbf{P}) \quad (1.5)$$

$$\begin{aligned} & - \|\varepsilon\|_{\mathbf{P}^o} = 0 \\ & - \left(\frac{\partial \|\varepsilon\|}{\partial P_i^m} \right)_{\mathbf{P}^m = \mathbf{P}^o} = 2 \int_0^{t_0} (y^m - y^o) \frac{\partial y^m}{\partial P_i^m} dt = 0 \\ & \quad \text{car } y^m - y^o = 0 \text{ pour } \mathbf{P}^m = \mathbf{P}^o \text{ (l'objet et le modèle ont la même structure)} \\ & - \left(\frac{\partial^2 \|\varepsilon\|}{\partial P_i^m \partial P_j^m} \right) = 2 \int_0^{t_0} (y^m - y^o) \frac{\partial^2 y^m}{\partial P_i^m \partial P_j^m} dt + 2 \int_0^{t_0} \frac{\partial y^m}{\partial P_i^m} \frac{\partial y^m}{\partial P_j^m} dt = 2 \int_0^{t_0} \frac{\partial y^m}{\partial P_i^m} \frac{\partial y^m}{\partial P_j^m} dt \end{aligned}$$

soit :

$$\|\varepsilon\|_{\mathbf{P}^o + \Delta\mathbf{P}} \simeq \frac{1}{2} (\Delta\mathbf{P})^T H (\Delta\mathbf{P}) \quad (1.6)$$

$\|\varepsilon\|$ peut être donc approximée, au voisinage de l'optimum, par une forme quadratique en $\Delta\mathbf{P}$. Il s'agit d'établir qu'elle est définie positive :

$$H = 2 \int_0^{t_0} \begin{bmatrix} \vdots \\ \sigma_i \\ \vdots \end{bmatrix} \begin{bmatrix} \dots & \sigma_j & \dots \end{bmatrix} dt = 2 \int_0^{t_0} \begin{bmatrix} \dots & \sigma_i \sigma_j \\ \sigma_i \sigma_j & \dots \end{bmatrix} dt$$

en posant : $\sigma_i(t) = \frac{\partial y^m}{\partial P_i^m}(\sigma_i(t))$: fonction de sensibilité) et $\boldsymbol{\sigma}^T = [\sigma_1 \quad \sigma_i \quad \sigma_N]$

$$H = 2 \int_0^{t_0} \boldsymbol{\sigma} \boldsymbol{\sigma}^T dt$$

soit :

$$\begin{aligned} \|\varepsilon\|_{\mathbf{P}^o + \Delta\mathbf{P}} & \simeq \int_0^{t_0} (\Delta\mathbf{P})^T \boldsymbol{\sigma} \boldsymbol{\sigma}^T (\Delta\mathbf{P}) dt = \int_0^{t_0} (\boldsymbol{\sigma}^T \Delta\mathbf{P})^T (\boldsymbol{\sigma}^T \Delta\mathbf{P}) dt \\ & = \int_0^{t_0} (\boldsymbol{\sigma}^T \Delta\mathbf{P})^2 dt \end{aligned}$$

car $(\boldsymbol{\sigma}^T \Delta\mathbf{P})$ est un scalaire.

Il apparaît donc, sous condition d'indépendance des fonctions de sensibilité sur l'horizon $(0, t_0)$, que la forme quadratique $\frac{1}{2} (\Delta\mathbf{P})^T H (\Delta\mathbf{P})$ est définie positive et par la même, l'équivalence entre la norme $\|\varepsilon\|$ et $D(O, M)$.

1.3.3 Validité d'un modèle - Nature des signaux d'entrée

La plupart des méthodes d'identification visent à ajuster le comportement temporel du modèle avec celui de l'objet, pour une classe de signaux d'entrée fixée, en général en fonction du but de l'opération.

1.3. CRITÈRE D'ÉQUIVALENCE "OBJET-MODÈLE" (ASPECT PARAMÉTRIQUE)

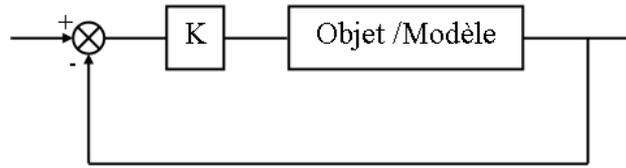


FIG. 1.14 – Système

Exemple 3 Soit à identifier un objet dans le but de définir l'action proportionnelle nécessaire à assurer la stabilité du système bouclé (cf. figure 1.14). On sait que le comportement en boucle fermée est déduit du comportement fréquentiel en boucle ouverte au voisinage des fréquences où le déphasage est de -135° . Il suffit donc d'avoir coïncidence des réponses fréquentielles du modèle et de l'objet dans ce voisinage pour obtenir une prédétermination correcte de la boucle fermée (cf. figure 1.15). Cependant, les comportements temporels de l'objet et du modèle en

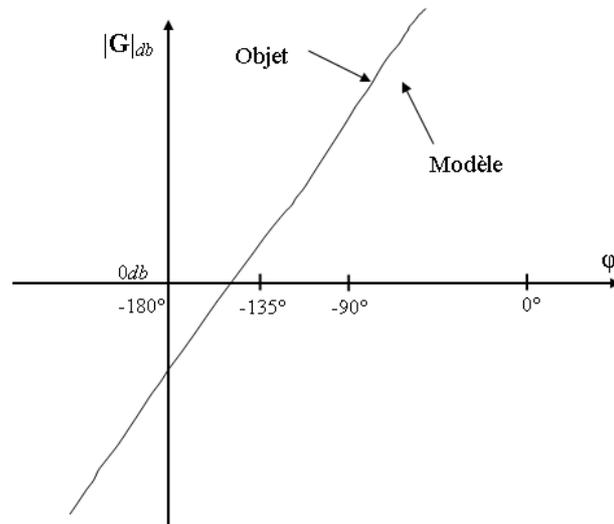


FIG. 1.15 – Réponses harmoniques en boucle ouverte (Plan de Black)

boucle ouverte, en réponse à un échelon, peuvent être très différents. (cf. figures 1.16 et 1.17)

Cependant, les "bons" signaux d'excitation dépendent aussi des caractéristiques (à priori inconnues) de l'objet. Ainsi, on conçoit aisément que pour ajuster

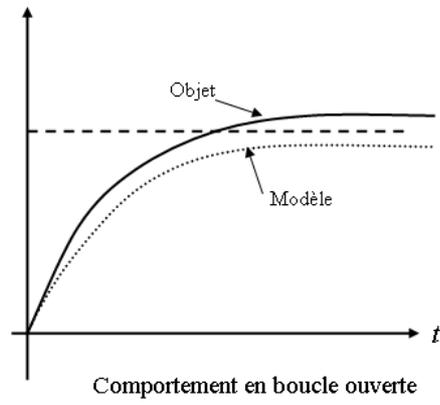


FIG. 1.16 – Réponse temporelle en boucle ouverte

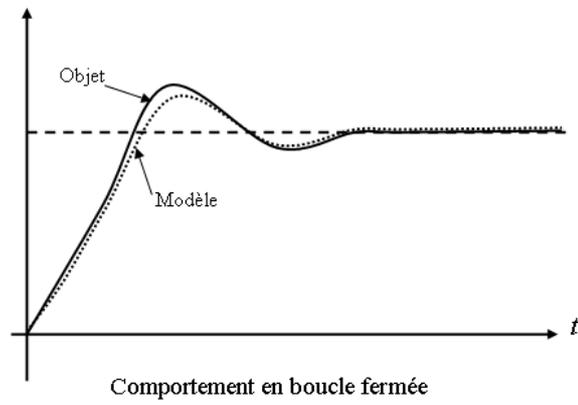


FIG. 1.17 – Réponse temporelle en boucle fermée

1.3. CRITÈRE D'ÉQUIVALENCE "OBJET-MODÈLE" (ASPECT PARAMÉTRIQUE)

un paramètre du modèle avec une bonne précision, toute modification de celui-ci doit entraîner une modification sensible de y^m . Autrement dit, les coefficients de sensibilité $\sigma_i(t) = \frac{\partial y^m}{\partial P_i^m}$ ne doivent pas être négligeables d'une part et être indépendants (voir la conclusion de la section 1.3.2) d'autre part.

d'où :

- Il faut choisir une structure du modèle "minimale" de manière à ne prendre en compte que les paramètres qui agissent efficacement sur la sortie.
- Il faut disposer de signaux d'entrée, suffisamment riches en information dans le domaine temporel (ou fréquentiel) que l'on désire explorer. Par exemple : échelon, séquences binaires aléatoires, signaux impulsionnels, ...

Exemple 4 Soit à identifier le système donné par la figure 1.18. Hypothèses :

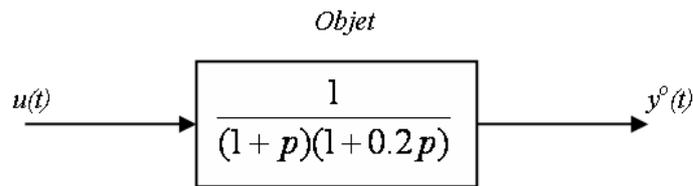


FIG. 1.18 – Objet à identifier

- Caractérisation idéale du modèle sous la forme :

$$\frac{K}{(1 + T_1 p)(1 + T_2 p)}$$

- Absence de bruit de mesure et de perturbations.
- $u(t)$ est un signal harmonique pur de pulsation 1rd/s .

Conclusion 5 Il existe une infinité de K, T_1, T_2 qui annulent l'erreur de sortie, car le signal d'entrée étant harmonique, les trois fonctions de sensibilité sont aussi harmoniques donc dépendantes.

Deux solutions sont alors envisageables :

- Réduction du modèle (par exemple sous la forme $\frac{\lambda_1}{p(1+\tau_1 p)}$)
- Choix d'un signal d'excitation plus riche en information (échelon, combinaison de plusieurs pulsations, ...)

1.3.4 L'initialisation du modèle

Au départ d'une expérience sur l'objet, les conditions initiales de celui-ci sont inconnues. En conséquence, $y^o - v$ est l'effet non seulement de la commande mesurée depuis le début de l'expérience mais aussi de tout le passé (résumé dans l'état initial). D'où la nécessité d'une phase d'initialisation : cette phase consiste généralement à n'exploiter la sortie y^o qu'au bout d'un temps T_0 (ou N) correspondant à la mémoire estimée de l'objet (cf. figure 1.19).

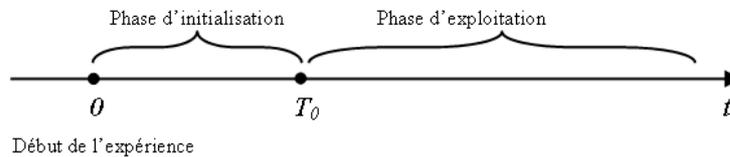


FIG. 1.19 – Déroulement de l'expérience

En effet, pour $t > T_0$ ($k > N$) l'évolution du système ne dépend pratiquement plus des conditions initiales.

1.3.5 L'objet est inséré dans une boucle de régulation

On considère le schéma de la figure 1.20. La consigne a ici pour rôle de fixer le point de fonctionnement autour duquel on désire obtenir un modèle dynamique de l'objet (donc, d'un point de vue dynamique, on peut considérer que $C = 0$). L'expérience peut être menée de deux manières :

- On ajoute un signal d'analyse w et l'on observe w , u et y^o que l'on exploite par la suite.
- On observe simplement u et y^o (donc avec $w = 0$ - le signal de commande u n'est pas constant car il dépend des perturbations v). Dans ce cas, lorsque le régulateur R est linéaire, on montre que toute tentative d'identification du comportement dynamique de l'objet, à partir des données u et y^o , par des méthodes qui n'exigent pas explicitement une structure causale du modèle (par exemple les méthodes fondées sur l'analyse spectrale), conduit à identifier l'inverse du régulateur et non l'objet (cf. figure 1.21).

1.3. CRITÈRE D'ÉQUIVALENCE "OBJET-MODÈLE" (ASPECT PARAMÉTRIQUE)

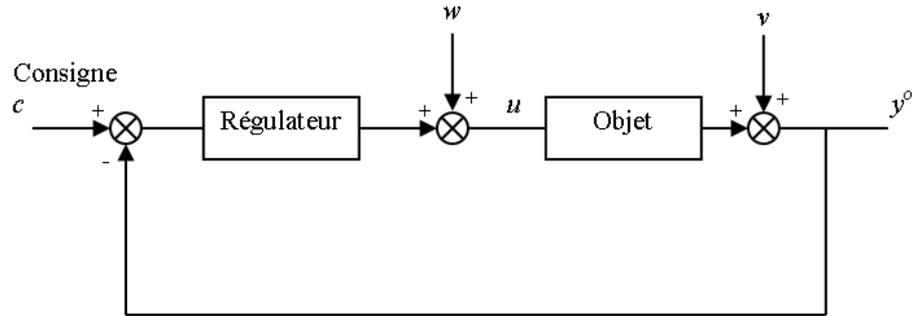


FIG. 1.20 – Objet inséré dans une boucle de régulation

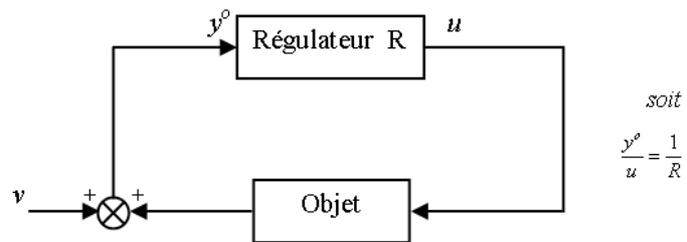


FIG. 1.21 – Erreur : identification de l'inverse du régulateur

1.4 Algorithme d'identification

L'identification d'un objet va comporter les étapes suivantes :

- exploitation de connaissances à priori, l'établissement de modèles analytiques,
- définition des conditions d'expériences,
- choix d'une structure de modèle, et de signaux d'analyse (étape de caractérisation),
- expérimentation et enregistrement de données,
- application d'algorithmes d'identification,
- vérification du modèle obtenu, et retour éventuel à la première étape.

Dans la suite, nous allons nous limiter à la description d'algorithmes, ce qui ne correspond en fait qu'à une faible partie du travail global d'identification.

Ces algorithmes reviennent à déterminer un vecteur paramètre en minimisant de manière plus ou moins explicite un critère d'erreur. Selon la manière dont la minimisation est conduite, on distingue :

- *Les méthodes directes* : dans lesquelles le modèle (ou bien tout paramètre le caractérisant), est relié de manière explicite aux observations. (solution de système d'équations linéaires, relevés de caractéristiques graphiques, ...)
- *Les méthodes itératives* : dans lesquelles tout paramètre est modifié à chaque cycle selon une loi de type :

$$\dot{P}(k+1) = P(k) - \lambda(k)\nabla_{P(k)}$$

où k est le numéro de cycle, ∇ est le plus souvent le gradient et $\lambda(k)$ un paramètre constant ou variable.

Le choix entre ces deux types de méthode est dicté par la manière dont le problème est formulé.

En effet, si le vecteur paramètre est solution d'un système d'équation linéaire, une méthode directe est possible, sinon on a recours à une méthode itérative.

Ce concept de linéarité en paramètre, **fondamental** pour les algorithmes d'identification, est sans rapport avec la linéarité dynamique de la théorie des systèmes comme le montrent les exemples suivants :

Exemple 6 Soit à identifier un objet discret par le modèle paramétrique suivant :

$$y^m(k) = \beta y^m(k-1) + u(k)$$

L'inconnue β peut être estimée à partir des observations $u(k)$ et $y^o(k)$, en ré-

1.4. ALGORITHME D'IDENTIFICATION

solvant le système d'équation suivant :

$$\begin{cases} y^0(0) = u(0) \\ y^0(1) = \beta y^0(0) + u(1) \\ y^0(2) = \beta y^0(1) + u(2) \\ \vdots \end{cases}$$

Ce système d'équation (surrabondant) est linéaire par rapport à β .

Un modèle non paramétrique équivalent au précédent est :

$$y^m(k) = \sum_{i=0}^k a_i^m u(k-i) \quad \text{avec} \quad a^m(i) = \beta^i$$

d'où le système d'équations, non linéaire en β :

$$\begin{cases} y^0(0) = u(0) \\ y^0(1) = \beta u(0) + u(1) \\ y^0(2) = \beta^2 u(0) + \beta u(1) + u(2) \\ y^0(3) = \beta^3 u(0) + \beta^2 u(1) + \beta u(0) + u(3) \\ \vdots \end{cases}$$

Exemple 7 Soit un objet discret que l'on modélise par l'équation suivante :

$$y^m(k+1) = \beta (y^m(k))^2 + u(k)$$

Ce modèle est celui d'un système dynamique non linéaire, et pourtant le paramètre β peut être obtenu à partir de la résolution d'un système d'équations linéaires :

$$\begin{cases} y^0(1) = \beta (y^0(0))^2 + u(0) \\ y^0(2) = \beta (y^0(1))^2 + u(1) \\ \vdots \end{cases}$$

En conclusion, on retiendra que la linéarité en comportement dynamique et la linéarité en paramètre sont des choses distinctes et qu'il est parfois possible de transformer un système d'équation non linéaire en un problème linéaire (et inversement).

Chapitre 2

Méthodes directes

Nous allons distinguer dans ce chapitre les méthodes dites graphiques (c'est-à-dire - simples du point de vue théorique et expérimental - ne réclamant aucun calcul complexe) de méthodes plus évoluées nécessitant l'utilisation de calculateurs numériques.

Il est à noter que ces méthodes graphiques, bien que n'étant pas toujours les plus performantes, sont le plus couramment utilisées et appréciées dans le domaine industriel.

2.1 Analyse harmonique

Le modèle adopté est ici la réponse fréquentielle (modèle non paramétrique). Notons que cette réponse est directement exploitable pour la synthèse de correcteurs et l'étude de la stabilité en théorie des asservissements.

Principe

L'objet est soumis à une excitation harmonique du type $u(t) = u_k \sin(w_k t)$ et l'on procède après extinction du régime transitoire, à l'enregistrement des signaux d'entrée et de sortie (cf. figure 2.1).

Cette expérience est réalisée pour différentes pulsations. On relève ensuite, sur les enregistrements, le gain et le déphasage de l'objet aux diverses pulsations. Ces points, reportés dans le plan de Bode ou de Nyquist, après interpolation permettent d'obtenir un tracé continu constituant un modèle "non paramétrique" de l'objet.

Mise en oeuvre

Pour un processus suffisamment rapide (du type servomécanisme) on peut se dispenser de réaliser des enregistrements en disposant d'un générateur harmo-

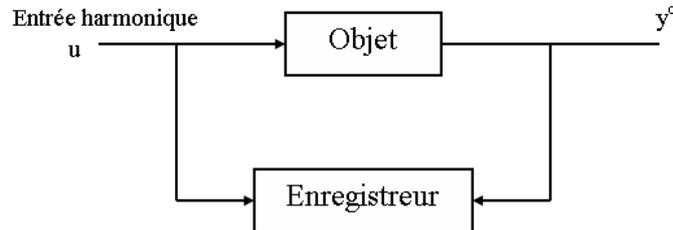


FIG. 2.1 – Analyse harmonique

nique à phase variable. La mesure de phase et de gain se fait alors directement à l'oscilloscope.

Pour les processus lents et perturbés, un enregistrement est le plus souvent indispensable. Notons qu'il existe des appareils spécialisés appelés "transféro-mètres" qui effectuent automatiquement une telle analyse.

Domaine de validité - Intérêt

Une caractéristique importante des méthodes harmoniques est liée au fait que le signal sinusoïdal n'est pas riche en information. Le nombre de fréquences d'essai doit être donc élevé. En pratique, il est suffisant de couvrir une à deux décades, avec 3 à 10 points par décade.

Il est à remarquer que le signal harmonique n'est pas un signal habituel de commande d'un système réel. L'application de cette méthode nécessite donc l'immobilisation de l'objet.

Pour un processus lent (temps de réponse supérieur à 10 secondes) on conçoit aisément que cette procédure soit fastidieuse.

Cette méthode, lorsqu'elle ne conduit pas à un temps d'expérimentation trop lent, est cependant excessivement intéressante car :

- Elle nécessite peu de connaissances à priori de l'objet (il est suffisant de connaître grossièrement la gamme d'analyse en fréquence et en amplitude).
- Elle est simple à mettre en oeuvre.
- Elle permet une linéarisation autour d'un point de fonctionnement en ne considérant que le premier harmonique de sortie.
- Elle peut être appliquée en boucle fermée, lorsque les perturbations ne sont pas elles aussi harmoniques à la même pulsation que l'entrée, et cela sans réduction de sensibilité.

- Elle est particulièrement bien adaptée à l'étude de la stabilité.
- Elle permet de déduire rapidement le fonctionnement en boucle fermée à partir d'une analyse restreinte au voisinage de la fréquence de coupure.

2.2 Analyse indicielle

Le modèle adopté est celui par réponse indicielle et c'est encore un modèle non paramétrique. Cette représentation peut être directement exploitable pour le réglage de régulateurs **PI** ou **P.I.D** (méthode de Ziegler Nichols), mais le plus souvent on cherche à paramétrer cette réponse (voir la section 2.4).

Principe

L'objet est soumis à une variation brusque de commande, et l'on enregistre l'effet de sortie (cf. figure 2.2).

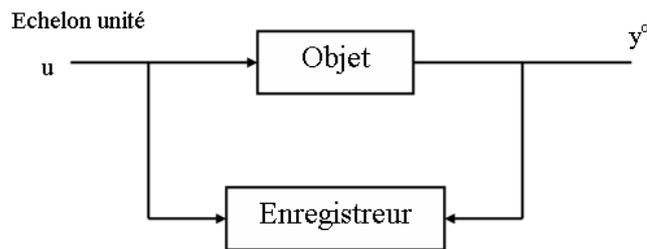


FIG. 2.2 – Analyse indicielle

La réponse à l'échelon unité, obtenue simplement par mise à l'échelle de la sortie, constitue un modèle non paramétrique de l'objet.

Intérêt, domaine d'application

L'échelon est très riche en information, et à priori un seul enregistrement suffit. En fait il faut faire extrêmement attention aux conditions d'expérience car celle-ci ne peuvent être validées que si le système est initialement au repos et si les perturbations intervenant en cours d'essai sont négligeables.

Notons que l'échelon est un signal de commande naturel d'un processus, et en ce sens l'analyse indicielle fournit des renseignements extrêmement utiles à la synthèse d'une commande.

Retenons toutefois que l'objet ne peut être inséré dans une boucle de régulation (la commande de l'objet ne serait pas un échelon).

2.3 Analyse impulsionnelle

L'impulsion idéale $\delta(t)$ ne peut être simulée. On peut cependant espérer obtenir une réponse proche de la réponse impulsionnelle en appliquant à l'objet le signal donné par la figure 2.3.

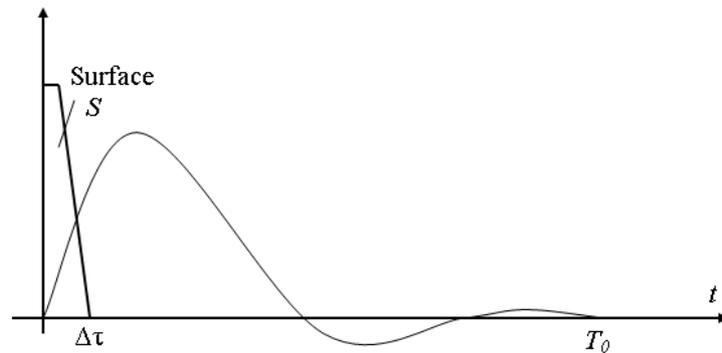


FIG. 2.3 – Signal d'excitation - réponse impulsionnelle

Ce signal est une impulsion au sens électronique du terme, de largeur τ et de surface S . La réponse de l'objet est proche de $Sh(t)$ si $\Delta\tau \ll T_0$ (où T_0 est le temps de réponse, et $h(t)$ la réponse impulsionnelle exacte du système).

En pratique cette méthode est rarement appliquée car l'impulsion est de faible énergie (la largeur $\Delta\tau$ est telle que $\Delta\tau \ll T_0$ et la hauteur est limitée par les saturations).

2.4 Approximation par un modèle paramétrique

2.4.1 A partir de la réponse harmonique

La méthode la plus simple consiste à comparer la réponse harmonique expérimentale à diverses réponses typiques (1^{er} ordre, 2nd ordre, voir l'annexe 2)

Lorsque la réponse expérimentale est plus complexe on peut tenter *la mise en série* de différents modèles classiques, en approximant la réponse globale par un jeu d'asymptotes dans le plan de Bode.

L'approximation asymptotiques est largement facilitée par confrontation des courbes de gain et de phase et en notant que cette dernière est plus sensible à la position des fréquences de brisure que ne l'est la courbe de gain.

La connaissance à priori de la structure d'un modèle paramétrique peut être extrêmement utile dans cette recherche (surtout lorsque cette structure est série).

2.4.2 A partir de la réponse indicielle - Méthode de Strejc

Ici encore on a intérêt à consulter un catalogue de réponses typiques. L'examen de cette réponse permet :

- d'estimer la mémoire approximative de l'objet,
- de déceler la présence d'un retard pur, d'une intégration, d'un bloc à déphasage non minimal,
- de préciser le degré du numérateur de la fonction de transfert par rapport au dénominateur en observant la tangente à l'origine (si cette tangente est nulle le degré du dénominateur est supérieur à celui du numérateur d'au moins deux unités).
- de déceler la présence de zéros, ou de pôles complexes conjugués lorsqu'il y a dépassement.

Le modèle Cette méthode peut s'appliquer aux systèmes dont la réponse indicielle ne présente pas de dépassement. On identifie à une fonction de la forme :

$$T(p) = \frac{K.e^{-\tau.p}}{(1 + T.p)^n} \quad (2.1)$$

Les paramètres à identifier sont donc :

- le gain statique K ,
- le retard τ ,
- la constante de temps T
- et l'ordre n .

La figure 2.4 représente les réponses indiciaires pour plusieurs jeux de paramètres.

La méthode Pour identifier le système, la méthode peut se décomposer en :

- Le gain statique est mesuré directement par la valeur finale de la sortie. Celle-ci vaut $K.E_0$ où E_0 est l'amplitude de l'échelon d'entrée.
- On trace la tangente au point d'inflexion I pour déterminer deux valeurs : T_1 et T_2 . Voir figure 2.5 pour la mesure de ces deux temps.
- Relever T_1 et T_2 , en déduire l'ordre n en utilisant le tableau 1. Entre deux lignes du tableau, on choisit la valeur de n la plus petite.
- Déterminer la constante de temps à partir de $\frac{T_2}{T}$ du tableau.

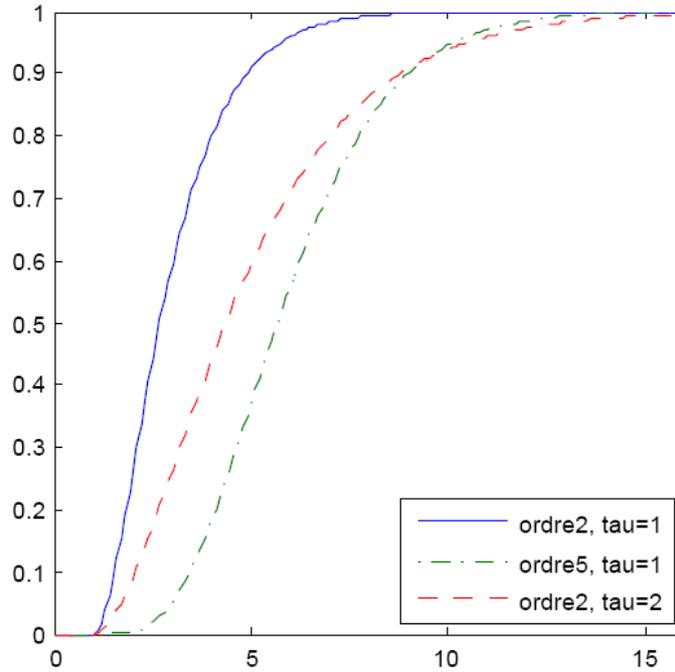


FIG. 2.4 – Réponses de modèles de Strejc pour $K = 1$, $T = 1$

- Déterminer le retard τ quand il existe à partir de la différence entre la valeur de T_1 mesurée et celle donnée par la colonne $\frac{T_1}{T_2}$ du tableau.

n	$\frac{T_1}{T}$	$\frac{T_2}{T}$	$\frac{T_1}{T_2}$
1	0	1	0
2	0.28	2.72	0.1
3	0.8	3.7	0.22
4	1.42	4.46	0.32
5	2.10	5.12	0.41
6	2.81	5.70	0.49

Exemple Pour tester cette méthode, nous partons d'un système dont la fonction de transfert est :

$$T(p) = \frac{100}{(p+4)(p+5)(p+1)}$$

Sa réponse indicielle est sur la figure 2.6 en trait plein.

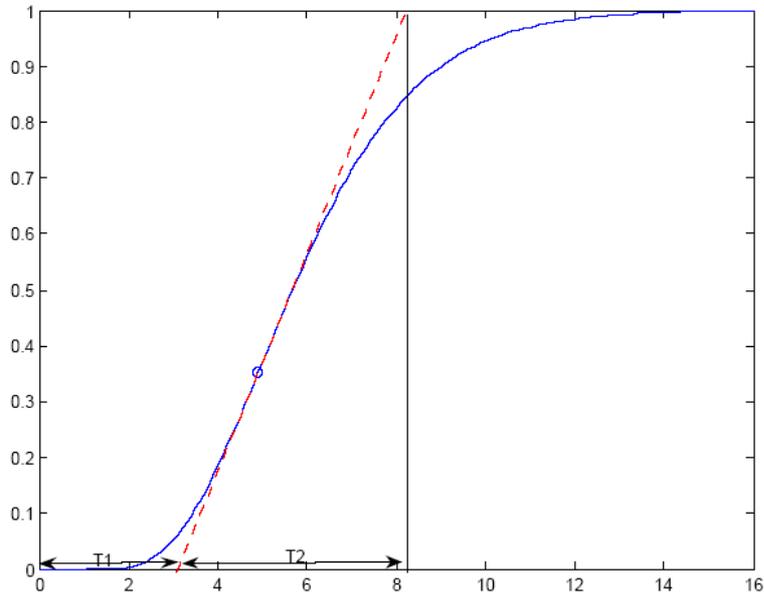


FIG. 2.5 – Méthode pour obtenir T_1 et T_2

- Le gain statique est mesuré directement par la valeur finale de la sortie : $K = 5$
- On trace la tangente au point d'inflexion I et on mesure : $T_1 = 0,27$ et $T_2 = 1,76$
- D'après le tableau, avec $\frac{T_1}{T_2} = 0,15$, un ordre $n = 2$ semble convenir.
- La constante de temps est évaluée à partir de $\frac{T_2}{T} = 2,72$ au tableau. Cela donne $T = 0,65s$.
- D'après le tableau, $\frac{T_1}{T} = 0,28$, ce qui donnerait une valeur de $T_1 = 0,18$. Or on mesure $T_1 = 0,27$. On peut en déduire un retard $\tau = 0,09$

La méthode identifie la réponse indicielle comme étant proche de celle du système suivant :

$$\hat{T}(p) = \frac{5.e^{-0.09p}}{(1 + 0.65p)^2}$$

La réponse de ce système est tracée dans la figure 2.6 en trait pointillé. On peut noter la grande ressemblance avec celle du système de départ alors qu'on a identifié un deuxième ordre avec retard au lieu d'un troisième ordre.

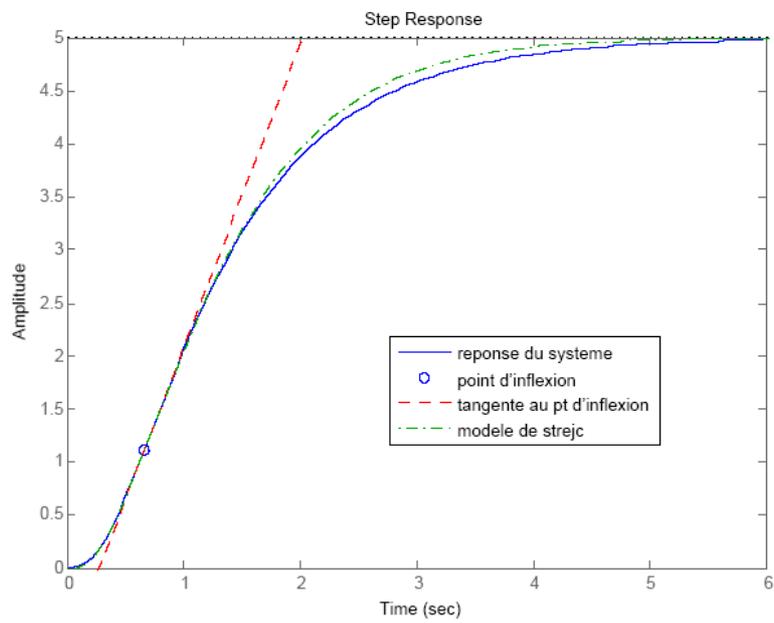


FIG. 2.6 – Réponses du système de départ et du système identifié

2.4.3 A partir de la réponse indicielle - Méthode de Broïda

Le modèle proposé pour approcher le comportement du système est un premier ordre avec un retard pur. Sa fonction de transfert est :

$$T(p) = \frac{K.e^{-\tau p}}{1 + Tp} \quad (2.2)$$

Le principe n'est pas de faire coïncider la tangente au point d'inflexion (souvent imprécis) mais d'ajuster les paramètres T et τ pour que les courbes de réponse du modèle et du processus aient deux points communs judicieusement choisis. Les points communs C_1 et C_2 habituellement utilisés correspondent respectivement à 28% et 40% de la valeur finale. Le modèle de Broïda donne les points C_1 et C_2 pour les dates suivantes :

$$\begin{aligned} - \frac{s(t)}{K.E_0} = 0.28 &\implies \frac{t - \tau}{T} = 0.328 \\ - \frac{s(t)}{K.E_0} = 0.40 &\implies \frac{t - \tau}{T} = 0.510 \end{aligned}$$

La méthode d'identification s'appuie sur les résultats précédents. Soient t_1 et t_2 les temps au bout desquels la réponse expérimentale atteint respectivement 28% et 40% de la valeur finale. On va simplement résoudre le système donné par :

$$\frac{t - \tau}{T} = 0.328 \implies t - \tau = 0.328T \quad (2.3)$$

$$\frac{t - \tau}{T} = 0.510 \implies t - \tau = 0.510T \quad (2.4)$$

La résolution de ces équations donne :

$$T = 5.5(t_2 - t_1) \quad (2.5)$$

$$\tau = 2.8t_1 - 1.8t_2 \quad (2.6)$$

Le gain K est déterminé comme dans la méthode de Strejc avec la valeur finale de la sortie.

Pour l'exemple précédent, la méthode de Broïda donne le modèle suivant :

$$T(p) = \frac{5.e^{-0.375p}}{(1 + 1.12p)}$$

La figure 2.7 donne les courbes de réponse du système réel et du modèle de Broïda. La concordance des deux points C_1 et C_2 est bien vérifiée.

2.4.4 Processus Intégrateur

Les systèmes contenant un intégrateur ont une réponse indicielle en rampe, en régime permanent. L'asymptote de cette réponse est une droite d'équation $y = a(t - t_1)$ de pente a et qui coupe l'axe des abscisses pour $t = t_1$ (voir

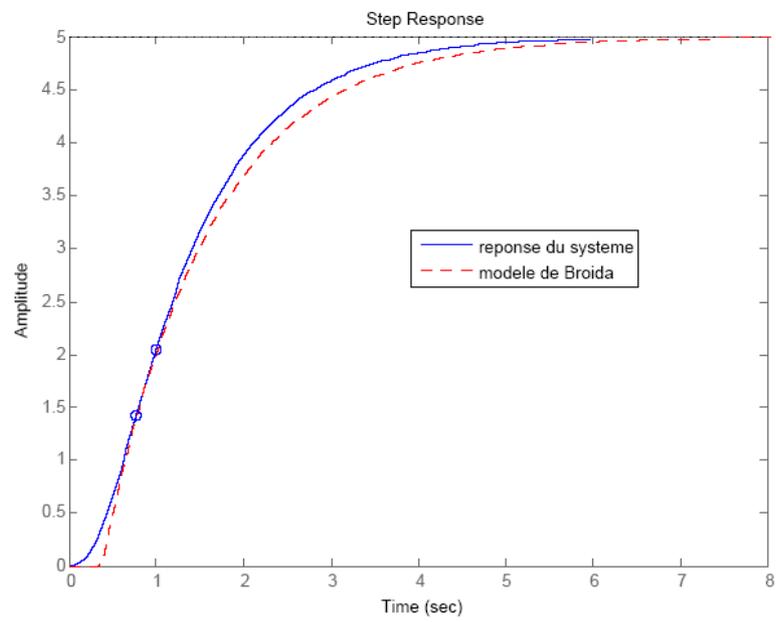


FIG. 2.7 – Courbe réelle approchée par un modèle de Broïda

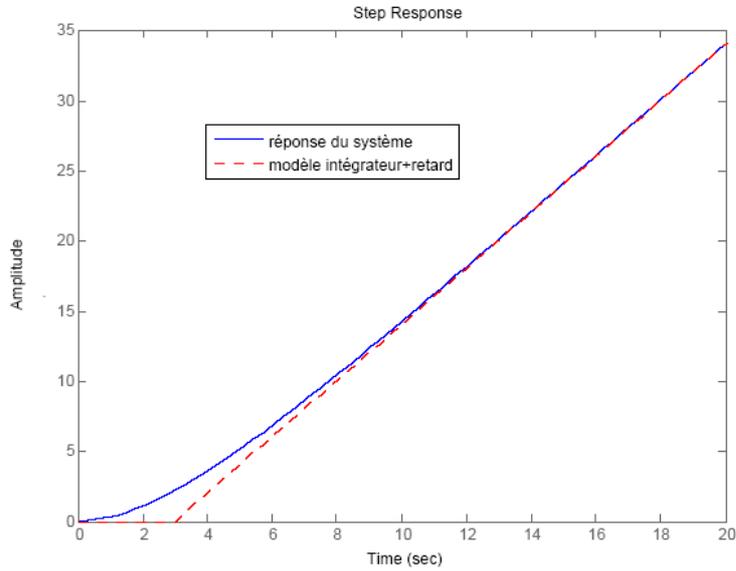


FIG. 2.8 – Courbe réelle approchée par un intégrateur retardé

figure 2.8). On identifie la réponse du système réel à la réponse d'un système intégrateur pur avec retard c'est à dire avec la fonction de transfert suivante :

$$T(p) = \frac{K.e^{-\tau p}}{p} \quad (2.7)$$

Les paramètres de ce système sont donnés par :

$$K = \frac{a}{E_0} \quad \tau = t_1$$

où E_0 est l'amplitude de l'échelon appliqué en entrée.

2.5 Obtention d'un modèle non paramétrique

Le but de ces méthodes est de calculer la réponse impulsionnelle ou fréquentielle dans des conditions de fonctionnement de l'objet plus complexes que celles imposées par des entrées harmoniques ou en échelon.

2.5.1 Problème préliminaire : Passage d'une réponse impulsionnelle à une réponse harmonique

Supposons disposer d'un modèle sous forme d'une réponse impulsionnelle relative à l'objet continu :

Le problème posé est le suivant :

- Quelles zones de la réponse fréquentielle est-il nécessaire d'obtenir avec précision dans un but de commande ?
- Quels sont les calculs mis en oeuvre pour qu'il en soit ainsi ?

La réponse à la première question a déjà été donnée partiellement dans le chapitre 1. Précisons cette idée : si notre but est de réaliser une commande en boucle fermée (commande du type PID), le comportement du système bouclé dépend essentiellement de l'allure de la réponse fréquentielle de l'objet au voisinage de la fréquence de coupure du système corrigé. Le problème peut paraître ici insoluble car la correction dépendra précisément de l'identification de l'objet. En fait, la réalité n'est pas exactement ainsi car les performances attendues en boucle fermée (du type temps du premier maximum t_m en réponse à un échelon, assorti d'un amortissement convenable) fixent ce voisinage fréquentiel à partir de la relation approximative $w_c t_m \simeq \pi$.

Il est bien évident, par ailleurs, que le temps t_m est choisi à partir d'une connaissance à priori de l'objet, et il est raisonnable de prendre t_m voisin de $\lambda.t_0$ où t_0 est le temps de réponse à 5% de l'objet, et λ compris entre 1 et $1/3$ (par exemple, on choisit $\lambda = 1/3$ lorsque l'objet semble d'ordre faible et sans retard pur).

Exemple Soit un objet présentant la réponse indicielle donnée par la figure 2.9.

L'objet semble d'ordre peu élevé, et l'on choisira $\lambda = \frac{1}{3}$ d'où :

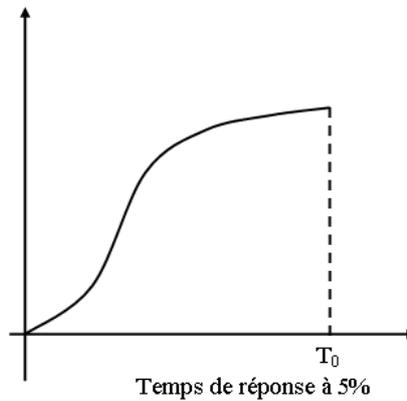


FIG. 2.9 – Réponse indicielle

$$t_m = \frac{t_0}{3} \implies f_c \simeq \frac{3}{2t_0}$$

En conséquence, on explorera avec attention une gamme fréquentielle d'au moins une décade centrée sur cette fréquence, soit $\frac{1}{2t_0} \leq f \leq \frac{10}{2t_0}$.

En cas d'utilisation d'un programme de transformée de fourier rapide (programme FFT) on pourra par exemple :

- Calculer la réponse impulsionnelle sur l'horizon $(0 - t_0)$ (où t_0 est le temps de réponse à 5% environ) en 64 points équidistants d'un intervalle T .
- Compléter cette réponse à 128 ou 256 points par des zéros (soit h_n cette séquence de M points)
- Appliquer l'algorithme FFT c'est-à-dire faire le calcul de la transformée de Fourier discrète de la séquence h_n soit :

$$H_k = \sum_{n=0}^{M-1} h_n \exp\left(\frac{-2\pi knj}{M}\right) \quad 0 \leq k \leq M-1 \quad (2.8)$$

où H_k est (au recouvrement de spectre près) la valeur de la réponse fréquentielle en $k \times \frac{1}{MT}$, soit $H_k \simeq \frac{1}{T} H(2\pi jf)$ pour $f = k \cdot \frac{1}{MT}$.

Cette mise en oeuvre n'est donnée qu'à titre d'exemple.

On retiendra cependant que seule la partie significative de la réponse impulsionnelle doit être calculée (c'est-à-dire sur un horizon $0 - t_0$) mais par contre, dans le cas de l'utilisation de l'algorithme FFT, cette réponse doit être complétée par des zéros afin d'obtenir une meilleure définition de la réponse fréquentielle (cette finesse d'analyse est égale à l'inverse de l'horizon de la réponse temporelle soit $\frac{1}{MT}$).

2.5.2 Principe du traitement numérique des données par corrélation

Dans le but d'éliminer l'effet des perturbations, sans avoir recours à un modèle stochastique, on peut envisager le traitement par corrélation suivant (figure 2.10) :

Données

- L'objet est supposé discret et caractérisé par une séquence de pondération $a^o(i)$.
- $w(k)$ est une séquence de nombres (séquence d'analyse) simulant une réalisation de processus aléatoire stationnaire d'ordre deux et ergodique.
- $w(k)$ et $u(k)$ sont corrélés : $\phi_{wu} \neq 0$
- $w(k)$ et $v(k)$ ne sont pas corrélés : $\phi_{wv} = 0$

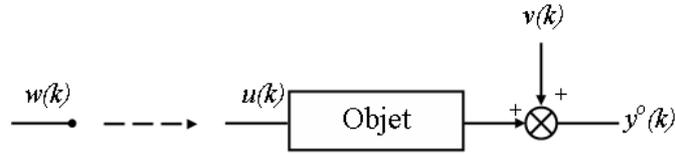


FIG. 2.10 – Schéma du traitement par corrélation

Problème Soit à calculer $\phi_{wy^o}(k)$.

$$y^o(k) = \sum_{l=0}^{\infty} a^o(l)u(k-l) + v(k) \quad (2.9)$$

$$\phi_{wy^o}(k) = E\{w(i)y^o(i+k)\} = E\left\{w(i) \sum_{l=0}^{\infty} a^o(l)u(k+i-l)\right\} + \underbrace{E\{w(i)v(k+i)\}}_{\phi_{wv}=0}$$

$$\phi_{wy^o}(k) = \sum_{l=0}^{\infty} a^o(l)E\{w(i)u(k+i-l)\} + 0 \quad (2.10)$$

$$\phi_{wy^o}(k) = \sum_{l=0}^{\infty} a^o(l)\phi_{wu}(k-l) \quad (2.11)$$

On aboutit au schéma équivalent donné par la figure 2.11.

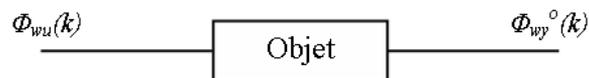


FIG. 2.11 – Schéma équivalent

où $\phi_{wy^o}(k)$ est la sortie fictive de l'objet lorsque $\phi_{wu}(k)$ est son entrée.

Le prétraitement des mesures w , u et y^o par la méthode de corrélation permet donc, en toute théorie, d'éliminer l'effet des perturbations.

Les aspects pratiques du calcul des fonctions ϕ_{wu} et ϕ_{wy^o} seront envisagés dans la suite pour le cas particulier où $w(k)$ est une séquence binaire pseudo aléatoire.

Application 1 Soit le schéma de la figure 2.12, où $u(k) = w(k) + x(k)$ avec $\phi_{wx} = 0$

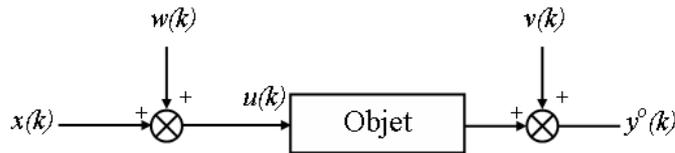


FIG. 2.12 – Application 1

Dans ce cas :

$$\phi_{wu}(k) = E \{w(i).u(i+k)\} \quad (2.12)$$

$$= E \{w(i).w(i+k) + w(i).x(i+k)\}$$

$$= E \{w(i).w(i+k)\}$$

$$= \phi_{ww}(k) \quad (2.13)$$

d'où :

$$\phi_{wy^o}(k) = \sum_{l=0}^{\infty} a^o(l)\phi_{ww}(k-l) \quad (2.14)$$

soit donc le schéma de la figure 2.13.

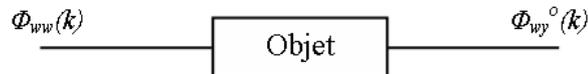


FIG. 2.13 – Application 1 : schéma équivalent

Ce prétraitement peut d'ailleurs conduire directement à l'obtention d'un modèle par séquence de pondération si l'on choisit le signal d'analyse $w(k)$ de telle façon

2.5. OBTENTION D'UN MODÈLE NON PARAMÉTRIQUE

que $\phi_{ww}(k) = \lambda^2 \delta(k)$. ($w(k)$ est alors une réalisation d'un bruit blanc discret).

Dans ces conditions, il est clair que :

$$\phi_{wy^o}(k) = \lambda^2 a^o(k) \quad (2.15)$$

Application 2 L'objet est inséré dans une boucle de régulation (cf. figure 2.14).

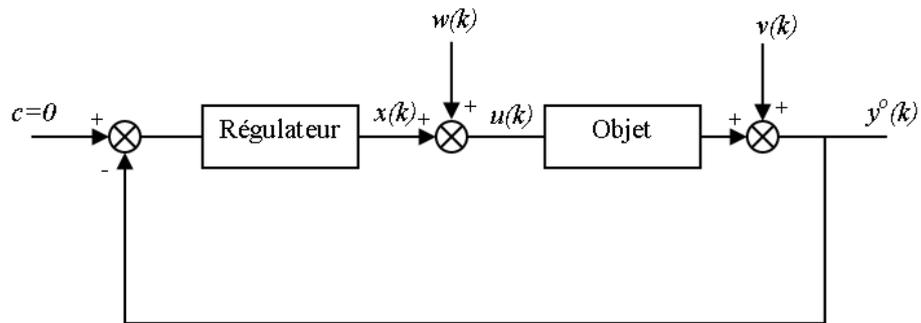


FIG. 2.14 – Application 2 : Objet inséré dans une boucle de régulation

Le prétraitement par corrélation conduit au schéma équivalent donnée par la figure 2.15, où l'objet n'est plus inséré dans une boucle.

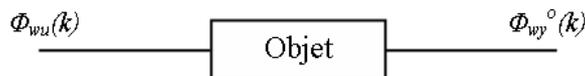


FIG. 2.15 – Application 2 : schéma équivalent

2.5.3 Utilisation des séquences binaires pseudo aléatoires

En pratique, $w(k)$ est générée par un calculateur numérique.

$w(k)$ est une séquence binaire de nombres valant $\pm A$, périodique, de période N , et dont la fonction d'autocorrélation est représentée sur la figure 2.16.

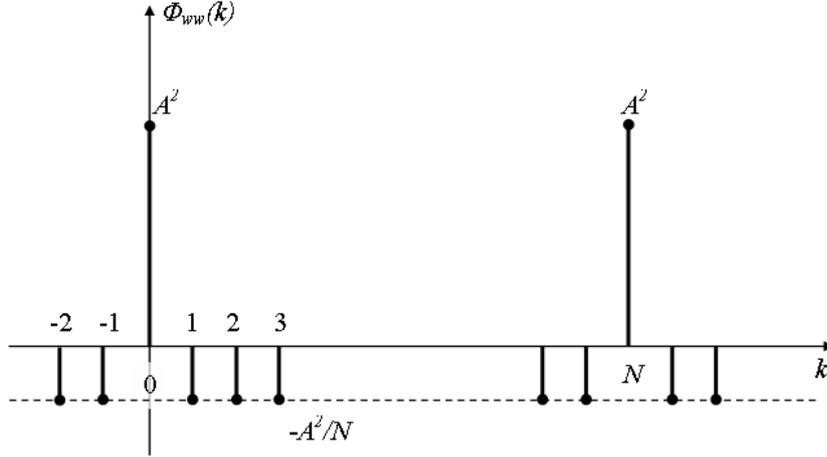


FIG. 2.16 – Fonction d'autocorrélation d'une SBPA

La mise en oeuvre sur ordinateur suppose le schéma suivant (cf. figure 2.17).

La tâche du calculateur consiste en :

- la génération de la séquence $w(k)$,
- l'élaboration (ou l'acquisition) de la grandeur de commande $u(k) = x(k) + w(k)$
- le traitement des signaux $w(k)$, $u(k)$ et $y^o(k)$ par corrélation.

Les fonctions d'intercorrélation sont calculées par les formules suivantes, compte tenu du caractère périodique de $w(k)$:

$$\begin{cases} \phi_{wy^o}(k) = \frac{1}{KN} \sum_{i=0}^{KN-1} y^o(i) \cdot w(i-k) \\ \phi_{wu}(k) = \frac{1}{KN} \sum_{i=0}^{KN-1} u(i) \cdot w(i-k) \end{cases} \quad \text{avec} \quad 0 \leq k \leq N-1 \quad (2.16)$$

où K est un nombre entier très grand à priori ; et N la longueur de la séquence $w(k)$.

Le nombre K peut être réduit à l'unité lorsque les signaux $y^o(k)$ et u sont périodiques de période N (Ceci suppose en particulier que les perturbations soient nulles).

L'intégration sur un horizon KN ayant pour but d'éliminer ces bruits, KN doit être d'autant plus élevé que leur niveau est important.

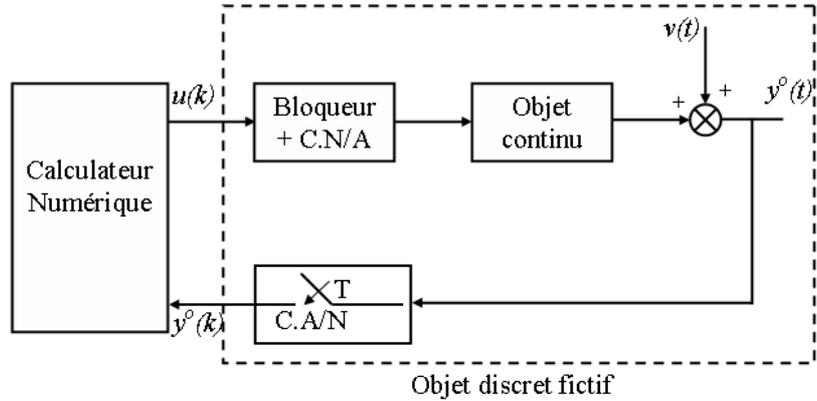


FIG. 2.17 – Mise en oeuvre sur calculateur numérique

Quand au nombre N , il doit être choisit de telle sorte que N soit supérieur ou égal à la mémoire de l'objet discret. Le choix de N est donc lié au choix de la période d'échantillonnage dans le cas d'un objet continu ($NT \geq T_0$) :

- s'il s'agit d'estimer un modèle continu on a intérêt à choisir un pas d'échantillonnage fin, donc un nombre N élevé ($N = 63, 127, 255, \dots$),
- s'il s'agit d'établir un modèle numérique de conduite, un échantillonnage conduisant à $N = 15, 31$ ou 63 est le plus souvent suffisant (en respectant toutefois $NT \geq T_0$).

L'estimation des fonctions de corrélation par les formules précédentes supposent la stationnarité des signaux u , y^o et w . Cette stationnarité peut être obtenue pratiquement en appliquant depuis l'instant $-N$ la séquence $w(k)$ (cf. figure 2.18).

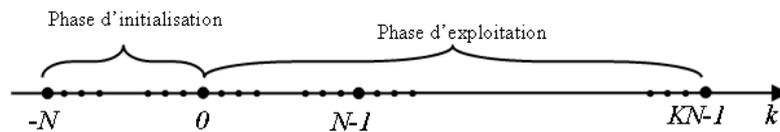


FIG. 2.18 – Déroulement de l'expérience

En effet, N étant supérieur à la mémoire de l'objet, tout se passe à l'instant 0 comme si la séquence $w(k)$ existait depuis $-\infty$.

L'interprétation des résultats conduit à distinguer les deux applications envisagées.

Application 1 Le calcul de $\phi_{wy^o}(k)$ suffit car $\phi_{ww}(k)$ est connue.

Si les conditions précédentes relatives à K et N sont respectées, nous pouvons écrire (cf. figures 2.19 et 2.20).

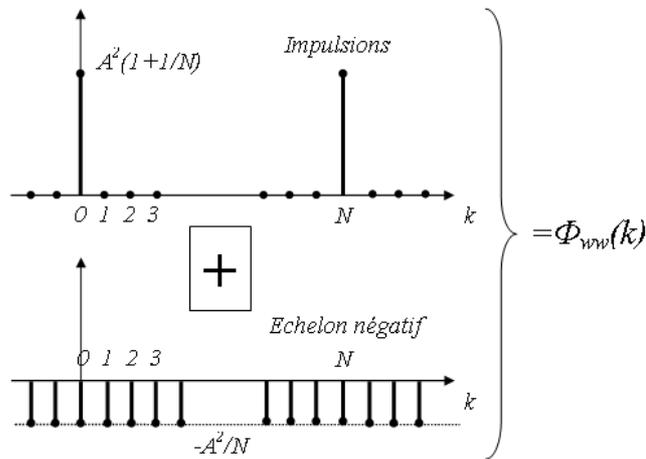


FIG. 2.19 – Fonction d'autocorrélation $\phi_{ww}(k)$

D'où, puisque $\phi_{wy^o}(k)$ est la réponse de l'objet à $\phi_{ww}(k)$:

$$\phi_{wy^o}(k) \simeq A^2 \left(1 + \frac{1}{N} \right) a^o(k) + \lambda, \quad \lambda : \text{constante}$$

Cette constante peut être aisément évaluée lorsque $a^o(0) = 0$ (ce qui est en général le cas) soit :

$$\phi_{wy^o}(k) - \phi_{wy^o}(0) = A^2 \left(1 + \frac{1}{N} \right) .a^o(k) \quad (2.17)$$

Le traitement conduit donc à la connaissance approchée de la séquence de pondération de l'objet discret schématisé par la figure 2.21.

2.5. OBTENTION D'UN MODÈLE NON PARAMÉTRIQUE

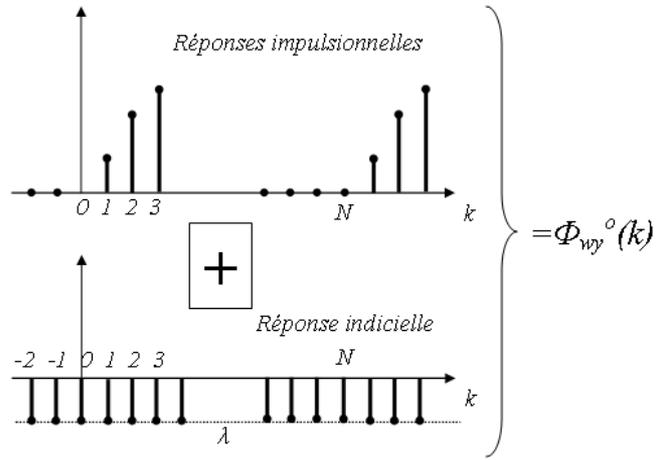


FIG. 2.20 – Fonction d'intercorrélation $\phi_{wy^o}(k)$

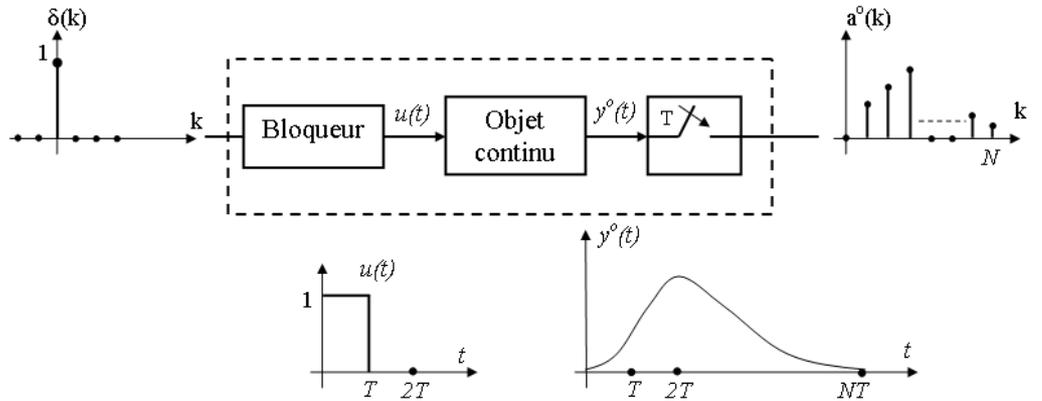


FIG. 2.21 – Application 1

On établit immédiatement la relation entre $a^o(k)$ et la réponse impulsionnelle du système continu :

$$a^o(k) = \int_{(k-1)T}^{kT} h^o(\theta) d\theta \quad (2.18)$$

et pour T suffisamment faible : $a^o(k) \simeq T.h^o(kT)$.

Application 2 Le traitement ne comporte pas de simplification et le calcul explicite de ϕ_{wu} et ϕ_{wy^o} est nécessaire.

Le problème de départ a été remplacé par un problème plus simple où les nouvelles séquences d'entrée et de sortie fictives de l'objet vont être traitées par des méthodes décrites ultérieurement (moindres carrés, méthode du modèle,...).

Conclusion Le traitement par corrélation, associé à l'analyse par signaux binaires constitue une méthode puissante d'étude des systèmes physiques réels.

2.5.4 Méthode des moindres carrés appliquée à l'obtention d'une séquence de pondération (déconvolution numérique)

Considérons ici un objet discret, de mémoire approximative N dont on cherche un modèle par séquence de pondération :

$$y^m(k) = \sum_{i=0}^{N-1} a^m(i).u(k-i) \quad \text{avec} \quad \begin{cases} a^m(i) = 0 \\ \text{pour } i \geq N \end{cases} \quad (2.19)$$

Pour M observations, l'équation précédente s'écrit sous forme matricielle ($M \geq N$) :

$$\begin{bmatrix} y^m(1) \\ y^m(2) \\ \vdots \\ y^m(M) \end{bmatrix} = \begin{bmatrix} u(1) & u(0) & \dots & u(-N+2) \\ u(2) & u(1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ u(M) & \dots & \dots & u(M-N+1) \end{bmatrix} \begin{bmatrix} a^m(0) \\ a^m(1) \\ \vdots \\ a^m(N-1) \end{bmatrix} \quad (2.20)$$

$$\mathbf{Y} = \mathbf{A} \times \mathbf{P}^m$$

Le vecteur modèle paramètre \mathbf{P}^m est obtenu en minimisant la fonction de coût :

$$Q = [\mathbf{Y}^o - \mathbf{A}\mathbf{P}^m]^T [\mathbf{Y}^o - \mathbf{A}\mathbf{P}^m] \quad (2.21)$$

où \mathbf{Y}^o est le vecteur d'observation de la sortie objet : $\mathbf{Y}^o = [y^o(1), y^o(2), \dots, y^o(M)]^T$

Pour cette structure de modèle (représentation par séquence de pondération)

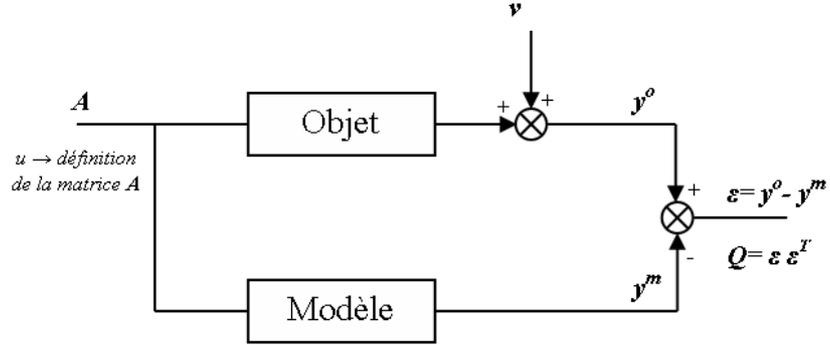


FIG. 2.22 – Méthode des moindres carrés

la méthode des moindres carrés revient bien à minimiser *l'erreur de sortie* car la matrice \mathbf{A} ne dépend que de l'entrée (cf. figure 2.22) :

La solution est alors donnée par :

$$\hat{\mathbf{P}}^m = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}^o \quad (2.22)$$

Lorsque le bruit v n'est pas corrélé avec la matrice \mathbf{A} , la méthode ne présente pas de biais systématique. On montre de plus que, si v est un bruit blanc gaussien, l'estimation est optimale.

La comparaison de cette méthode avec la méthode de corrélation fournit une interprétation intéressante :

$$\frac{\mathbf{A}^T \mathbf{A}}{M} \hat{\mathbf{P}}^m = \frac{\mathbf{A}^T \mathbf{Y}^o}{M} \quad (2.23)$$

On vérifie aisément que, pour M très élevé :

$$\frac{\mathbf{A}^T \mathbf{A}}{M} \simeq \begin{bmatrix} \phi_{uu}(0) & \phi_{uu}(1) & \dots & \phi_{uu}(N-1) \\ \phi_{uu}(1) & \phi_{uu}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \phi_{uu}(N-1) & \dots & \dots & \phi_{uu}(0) \end{bmatrix} \quad (2.24)$$

Par exemple : le terme noté $\phi_{uu}(0)$ correspond à $\frac{1}{M} [u^2(1) + u^2(2) + u^2(3) + \dots + u^2(M)]$ qui est bien une estimation de la fonction d'autocorrélation de l'entrée u pour un retard nul.

On vérifie de même que :

$$\frac{\mathbf{A}^T \mathbf{Y}^o}{M} = \begin{bmatrix} \phi_{uy^o}(0) \\ \phi_{uy^o}(1) \\ \vdots \\ \phi_{uy^o}(N-1) \end{bmatrix} \quad (2.25)$$

d'où

$$\begin{bmatrix} \phi_{uu}(0) & \phi_{uu}(1) & \dots & \phi_{uu}(N-1) \\ \phi_{uu}(1) & \phi_{uu}(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \phi_{uu}(N-1) & \dots & \dots & \phi_{uu}(0) \end{bmatrix} \begin{bmatrix} a^0(0) \\ \vdots \\ \vdots \\ a^o(N-1) \end{bmatrix} = \begin{bmatrix} \phi_{uy^o}(0) \\ \phi_{uy^o}(1) \\ \vdots \\ \phi_{uy^o}(N-1) \end{bmatrix} \quad (2.26)$$

c'est-à-dire :

$$\phi_{uy^o}(k) = \sum_{i=0}^{N-1} a^0(i) \phi_{uu}(k-i) \quad (2.27)$$

On retrouve ainsi la formule de corrélation, et lorsque u est une entrée du type bruit blanc la matrice $\mathbf{A}^T \mathbf{A}$ est diagonale, ce qui conduit à :

$$\frac{\mathbf{A}^T \mathbf{Y}^o}{M} = \mathbf{P}^m \phi_{uu}(0) \quad (2.28)$$

Cette manière de procéder évite l'inversion de la matrice $\mathbf{A}^T \mathbf{A}$.

Dans le cas général l'inversion de $\mathbf{A}^T \mathbf{A}$ peut poser quelques problèmes car son déterminant est souvent très faible et l'on pourra avoir recours à la méthode de Choleski.

La comparaison avec la méthode de corrélation nous invite cependant à choisir une séquence d'entrée riche en information, ce qui a pour effet de bien conditionner la matrice $\mathbf{A}^T \mathbf{A}$ (interprétée comme matrice de corrélation).

2.6 Obtention d'un modèle paramétrique

2.6.1 Notion d'ordre minimal

Si un système abstrait décrit par un modèle non paramétrique possède une représentation paramétrique, il en admet alors une infinité dont l'une est d'ordre minimal.

Le problème de passage d'une représentation non paramétrique à une représentation paramétrique d'ordre minimal est le suivant :

2.6. OBTENTION D'UN MODÈLE PARAMÉTRIQUE

Connaissant les coefficients $a(i)$ de la séquence de pondération, il s'agit de trouver l'équation d'état 2.29 (cas monovarié) d'ordre minimal représentant le système. L'ordre sera dit minimal si et seulement si le triplet $\{F, G, H\}$ est entièrement observable et commandable.

$$\begin{cases} \mathbf{x}(k+1) = F\mathbf{x}(k) + Gu(k) \\ y(k) = H\mathbf{x}(k) \end{cases} \quad \text{où } \mathbf{x}(k) \text{ est un vecteur d'ordre } n \quad (2.29)$$

Rappelons que :

- Un système est dit *complètement commandable* si, quelque soit l'état initial $\mathbf{x}(0)$, il existe une commande $\{u(0), u(1), \dots\}$ qui transfère en un temps fini le système en un état donné quelconque.

Soit $\mathbf{x}(k)$ l'état à atteindre, et K le nombre de coups nécessaires. On a alors :

$$\mathbf{x}(k) = F^k \mathbf{x}(0) + \sum_{i=0}^{K-1} F^{K-i-1} G u(i) \quad (2.30)$$

ou encore :

$$\mathbf{x}(k) = F^k \mathbf{x}(0) + [F^{K-1}G | F^{K-2}G | \dots | G] \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(K-1) \end{bmatrix} \quad (2.31)$$

On démontre alors que la condition nécessaire et suffisante de commandabilité complète est que la matrice de commandabilité $\mathcal{C} = [G | FG | \dots | F^{K-1}G]$ soit de rang n , égal à l'ordre de \mathbf{x} .

- Un système est dit *complètement observable* s'il est possible de retrouver l'état initial $\mathbf{x}(0)$ à partir de l'observation de la suite des sorties sur un nombre fini d'échantillons, pour une commande nulle :

$$y(k) = HF^k \mathbf{x}(0) \quad (2.32)$$

Soit encore : $[y(0), y(1), \dots, y(K-1)] = \mathbf{x}^T(0) \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{K-1} \end{bmatrix}$

On démontre que la condition nécessaire et suffisante d'observabilité complète est que la matrice d'observabilité $\mathcal{O} =$

$$\begin{bmatrix} H \\ HF \\ \vdots \\ HF^{K-1} \end{bmatrix} \text{ soit de rang } n.$$

Ces matrices de commandabilité et d'observabilité sont liées à la définition de la matrice de Hankel :

$$\mathcal{H}_r = \begin{bmatrix} a(1) & \cdots & a(n) & \cdots & a(r) \\ \vdots & \ddots & \vdots & & \vdots \\ a(n) & \cdots & a(2n-1) & & \vdots \\ \vdots & & & \ddots & \vdots \\ a(r) & \cdots & \cdots & \cdots & a(2r-1) \end{bmatrix} = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{r-1} \end{bmatrix} [G|FG|\dots|F^{r-1}G] \quad (2.33)$$

L'équation 2.33 donne la matrice de Hankel d'ordre r .

D'après l'expression de la réponse impulsionnelle à partir de l'équation d'état :

$$\begin{cases} a(i) = HF^{i-1}G \\ a(0) = 0 \end{cases} \quad (2.34)$$

On en déduit donc que l'ordre de la réalisation minimale est égal au rang de la matrice de Hankel (Résultat extensible au cas multivariable).

Dans le cas monovariante on montre que la matrice de Hankel d'ordre n est régulière.

2.6.2 Recherche de l'ordre minimal par factorisation de la matrice de Hankel

La matrice de Hankel étant symétrique, une factorisation en matrices triangulaires est possible de la manière suivante :

$$\begin{aligned}
 \begin{bmatrix} a(1) & \cdots & a(n) & \cdots & a(r) \\ \vdots & \ddots & \vdots & & \vdots \\ a(n) & \cdots & a(2n-1) & & \vdots \\ \vdots & & & \ddots & \vdots \\ a(r) & \cdots & \cdots & \cdots & a(2r-1) \end{bmatrix} &= \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ p_{21} & 1 & & \vdots & & \vdots \\ \vdots & & \ddots & \vdots & & \vdots \\ p_{n1} & \cdots & \cdots & 1 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ p_{r1} & \cdots & \cdots & \cdots & \cdots & 1 \end{bmatrix} \times \\
 &\begin{bmatrix} q_{11} & \cdots & \cdots & q_{1n} & \cdots & q_{1r} \\ 0 & q_{22} & & \vdots & & \vdots \\ \vdots & & \ddots & \vdots & & \vdots \\ 0 & \cdots & \cdots & q_{nn} & & \vdots \\ \vdots & \cdots & \cdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \ddots \end{bmatrix} \\
 \mathcal{H}_r &= \mathbf{P} \times \mathbf{Q} \tag{2.35}
 \end{aligned}$$

Par construction la matrice \mathbf{P} est de rang r , donc la matrice \mathbf{Q} est d'un rang égal à celui de la matrice de Hankel.

En conséquence le premier terme nul sur la diagonale principale de la matrice \mathbf{Q} fixe le rang de \mathcal{H}_r (c'est le premier bloc de \mathcal{H}_r qui est régulier).

On remarque que cette factorisation est simple à réaliser au point de vue numérique :

- $q_{1i} = a(i)$
- $p_{21} \times q_{11} = a(2) \implies p_{21} = \frac{a(2)}{q_{11}}$
- connaissant p_{21} on peut déterminer la seconde ligne de Q et ainsi de suite, jusqu'à ce que l'on rencontre un zéro sur la diagonale de Q .

En pratique la séquence de pondération n'est connue qu'avec une certaine erreur ϵ , ce qui conduit à définir une matrice de Hankel avec un ordre quasi infini (la séquence d'erreur additionnelle est considérée comme la séquence de pondération d'un système n'admettant pas, à priori, de représentation paramétrique). Il convient donc de rechercher la nullité des termes diagonaux de Q en tolérant une erreur.

2.6.3 Recherche de l'ordre par essais successifs

Une méthode plus simple et plus systématique consiste à rechercher l'efficacité de l'ordre d'un modèle. Pour cela :

1. On se fixe un ordre minimum a priori (par exemple $n = 1$).
2. On recherche les paramètres du modèle correspondant (d'après les méthodes paramétriques décrites dans la suite).
3. On mesure l'efficacité de l'ordre en construisant une fonction de coût normalisée du type :

$$Q = \frac{\sum_{k=1}^M \{y^o(k) - y^m(k)\}^2}{\sum_{k=1}^M \{y^o(k)\}^2} \quad (2.36)$$

où $y^o(k)$ est la sortie expérimentale de l'objet (séquence de longueur M) et $y^m(k)$ est la sortie du modèle calculée pour les paramètres déterminés précédemment, ce modèle étant soumis à la même entrée que l'objet.

4. On fait $n = n + 1$ et on reprend les opérations (2) et (3).
5. On construit la courbe d'efficacité $Q = F(n)$, donnée par la figure 2.23.

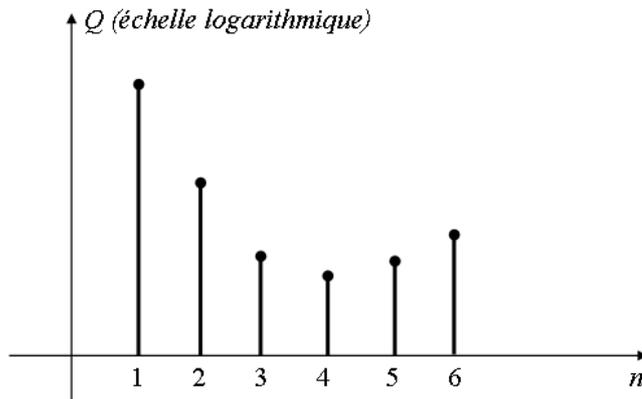


FIG. 2.23 – Courbe d'efficacité

On notera que pour le graphe précédent il est raisonnable de choisir $n = 3$.

2.6.4 Méthode des moindres carrés

Cette méthode est appliquée ici à la détermination d'un modèle par l'équation aux différences d'ordre n (n est donné).

$$y^m(k) = -\sum_{i=1}^n \beta_i^m y^m(k-i) + \sum_{j=1}^n \alpha_j^m u(k-j) \quad (2.37)$$

2.6. OBTENTION D'UN MODÈLE PARAMÉTRIQUE

On remarque que cette structure de modèle n'est pas de la forme :

$$y^m(k) = f(\mathbf{x}(k), \mathbf{P}^m) \quad (2.38)$$

(où $\mathbf{x}(k)$ est un vecteur de mesure) car $y^m(k)$ dépend de $y^m(k-i)$. Nous pouvons supprimer cette dépendance en explicitant successivement les $y^m(k-i)$ en fonction des sorties modèle précédentes et ainsi de suite, mais il apparaît alors que le système d'équations devient non linéaire en paramètre, comme le montre l'exemple suivant.

Exemple 8 Soit le système donnée par :

$$\begin{cases} y^m(k) = \beta y^m(k-1) + \alpha u(k) \\ y^m(0) = 0, \quad u(0) = 0 \end{cases}$$

On en déduit le système explicite "non linéaire en paramètre" donné par :

$$\begin{cases} y^m(1) = \alpha u(1) \\ y^m(2) = \beta(\alpha u(1)) + \alpha u(2) \\ y^m(3) = \beta[\beta\alpha u(1) + \alpha u(2)] + \alpha u(3) \\ \dots \end{cases}$$

Ce système d'équation n'admet pas de solution simple au sens des moindres carrés.

Par contre un modèle du type donné par 2.39 : (un modèle dit *de prédiction*)

$$y^m(k) = - \sum_{i=1}^n \beta_i^m y^o(k-i) + \sum_{j=1}^n \alpha_j^m u(k-j) \quad (2.39)$$

possède la structure explicite et la linéarité en paramètre. La méthode des moindres carrés vise à minimiser la fonction de coût :

$$Q = \sum_{k=1}^M \{y^o(k) - y^m(k)\}^2 \quad (2.40)$$

$$Q = \sum_{k=1}^M \left\{ y^o(k) + \sum_{i=1}^n \beta_i^m y^o(k-i) - \sum_{j=1}^n \alpha_j^m u(k-j) \right\}^2 \quad (2.41)$$

Il apparaît donc une minimisation effective de l'erreur généralisée $e(k)$ et non de l'erreur de sortie (cf. figure 2.24) ($e(k)$ est encore appelée *résidu*).

$$e(k) = y^o(k) + \sum_{i=1}^n \beta_i^m y^o(k-i) - \sum_{j=1}^n \alpha_j^m u(k-j) \quad (2.42)$$

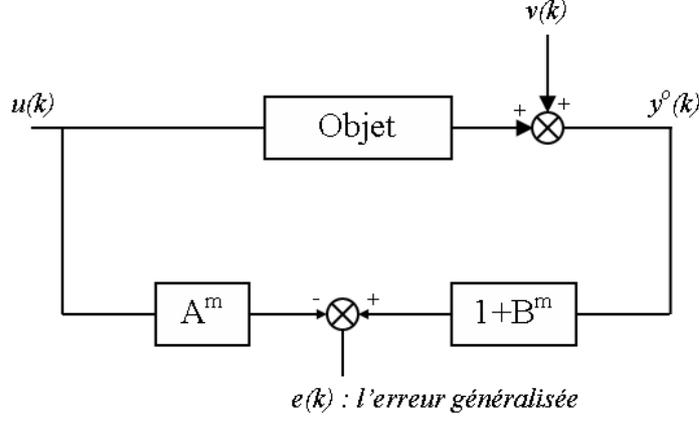


FIG. 2.24 – Méthode des moindres carrés

La mise en équation matricielle du problème fournit :

$$\begin{bmatrix} y^m(1) \\ y^m(2) \\ \vdots \\ y^m(M) \end{bmatrix} = \begin{bmatrix} u(1) & \cdots & u(-n+1) & | & y^o(0) & \cdots & y^o(1-n) \\ u(2) & \cdots & u(-n+2) & | & y^o(1) & \cdots & y^o(2-n) \\ \vdots & & \vdots & | & \vdots & & \vdots \\ u(M) & \cdots & u(M-n) & | & y^o(M-1) & \cdots & y^o(M-n) \end{bmatrix} \begin{bmatrix} \alpha^m(0) \\ \vdots \\ \alpha^m(n) \\ \hline -\beta^m(1) \\ \vdots \\ -\beta^m(n) \end{bmatrix}$$

$$\mathbf{Y}^m = \mathbf{A} \times \mathbf{P}^m \tag{2.43}$$

et en notant $\mathbf{Y}^o = [y^o(1), y^o(2), \dots, y^o(M)]^T$, il vient :

$$\hat{\mathbf{P}}^m = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}^o \tag{2.44}$$

Afin d'analyser le biais de cette méthode, supposons que l'objet soit décrit par l'équation stochastique suivante :

$$y^o(k) = - \sum_{i=1}^n \beta_i^m y^o(k-i) + \sum_{j=1}^n \alpha_j^m u(k-j) + \underbrace{\sum_{l=1}^n c_l \nu(k-l)}_{\zeta(k)} \tag{2.45}$$

Soit en notation matricielle

$$\mathbf{Y}^o = \mathbf{A} \mathbf{P}^o + \zeta \tag{2.46}$$

2.6. OBTENTION D'UN MODÈLE PARAMÉTRIQUE

On remarque immédiatement que la séquence $\zeta(k)$ est corrélée avec la matrice \mathbf{A} car dans le calcul du biais, le produit matriciel :

$$\mathbf{A}^T \boldsymbol{\zeta} = \begin{bmatrix} u(1) & u(2) & \dots & u(M) \\ \dots & \dots & \dots & \dots \\ y^o(1) & y^o(2) & \dots & y^o(M) \\ \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} \zeta(1) \\ \zeta(2) \\ \vdots \\ \zeta(M) \end{bmatrix}$$

intervient et fait apparaître des termes de la forme $y^o(i)\zeta(k+i)$ avec $k > 1$.

$y^o(i)$ dépend de $\zeta(i)$ et puisque $\zeta(i)$ n'est pas en général un bruit blanc il apparaît une corrélation entre $y^o(i)$ et $\zeta(k+i)$ pour $k \geq 1$.

En conclusion, la méthode des moindres carrés appliquée directement à la détermination des coefficients d'une équation aux différences conduit le plus souvent à un biais d'autant plus important que les perturbations et bruits en sortie de l'objet sont élevés.

Une solution pratique consiste à effectuer un *prétraitement des informations par corrélation* et ensuite appliquer la méthodes des moindres carrés (cf. figure 2.25). Il suffit de substituer la séquence ϕ_{wu} à u et ϕ_{wy^o} à y^o .

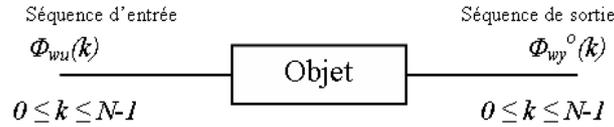


FIG. 2.25 – Moindres carrés après traitement par corrélation

Notons que le problème est simplifié dans l'application 1 de 2.5.2 car la méthode de corrélation est susceptible de fournir directement une estimation de la séquence de pondération de l'objet. Dans ce cas la séquence d'entrée fictive est $u(k) = \delta(k)$ et la sortie est l'estimation des $a^o(k)$. On remarque alors, étant donné la structure particulière de la matrice \mathbf{A} , que les coefficients β_i^m peuvent être déterminés en premier lieu indépendamment des α_i^m . Ce qui a pour intérêt essentiel de réduire la dimension des matrices à inverser.

Chapitre 3

Méthodes Itératives

Dans ce chapitre nous allons voir comment il est possible d'éliminer le biais relatif à l'identification d'un modèle paramétrique (voir la méthode des moindres carrés) en introduisant des calculs itératifs.

3.1 Méthodes stochastiques

Ces méthodes reviennent à caractériser simultanément un modèle déterministe aux commandes $u(i)$ et un modèle stochastique relatif aux perturbations $v(i)$.

3.1.1 Principe

Le biais de la méthode des moindres carrés provient du fait que la séquence $\zeta(k)$ n'était pas un bruit blanc.

En effet, nous avons supposé que l'objet était décrit par l'équation :

$$Y^o(z) = \frac{A(z)}{1+B(z)} \cdot U(z) + \underbrace{\frac{C(z)}{1+B(z)} \cdot \nu(z)}_{v(z)}$$

où $\nu(k)$ est un bruit blanc.

Supposons que $C(z)$ soit connu : il suffit alors d'introduire un filtrage inverse $1/C(z)$ comme l'indique la figure 3.1, afin d'éliminer le biais dans l'estimation des paramètres par la méthode des moindres carrés.

$$\begin{aligned} Y^*(z) &= \left[\frac{A(z)}{1+B(z)} \cdot U(z) + \frac{C(z)}{1+B(z)} \cdot \nu(z) \right] \cdot \frac{1}{C(z)} \\ Y^*(z) &= \frac{A(z)}{1+B(z)} \cdot U^*(z) + \frac{\nu(z)}{1+B(z)} \end{aligned}$$

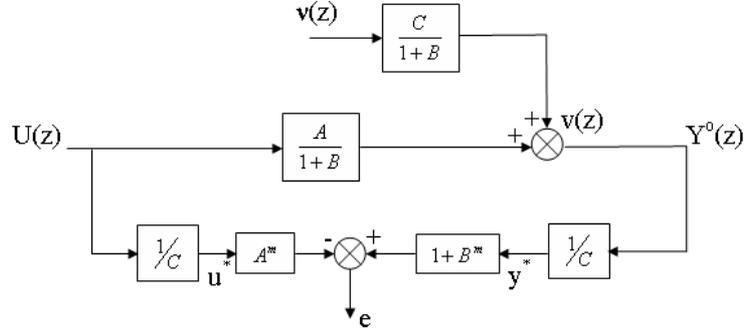


FIG. 3.1 – Elimination du biais par filtrage inverse

c'est-à-dire :

$$y^*(k) = - \sum_1^n \beta_i^o . y^*(k-i) + \sum_0^n \alpha_j^o . u^*(k-j) + \nu(k)$$

En considérant $u^*(k)$ comme mesure de l'entrée et $y^*(k)$ comme mesure de la sortie objet, le modèle de prédiction que l'on est conduit à faire est :

$$y^m(k) = - \sum_1^n \beta_i^m . y^*(k-i) + \sum_0^n \alpha_j^m . u^*(k-j)$$

d'où la fonction de coût à minimiser :

$$Q = \sum_1^M e^2(k) = \sum_1^M [y^*(k) - y^m(k)]^2 = [\mathbf{Y}^* - \mathbf{A} . \mathbf{P}^m]^T . [\mathbf{Y}^* - \mathbf{A} . \mathbf{P}^m]$$

à partir de la mise en équation matricielle :

$$\begin{aligned} \mathbf{Y}^m &= \mathbf{A} . \mathbf{P}^m && \text{relative au modèle} \\ \mathbf{Y}^* &= \mathbf{A} . \mathbf{P}^o + \boldsymbol{\nu} && \text{relative à l'objet} \end{aligned}$$

Cette fois, le bruit $\nu(k)$ étant un bruit blanc, il n'y a plus de corrélation entre la matrice \mathbf{A}^T et le vecteur $\boldsymbol{\nu}$. Par la même, le biais disparaît (ceci est obtenu en blanchissant le résidu $e(k)$).

En fait, le modèle du bruit $\{v(k)\}$ est inconnu a priori et le principe précédent ne peut être appliqué directement. Les méthodes suivantes cherchent à approximer $C(z)$ de manière itérative.

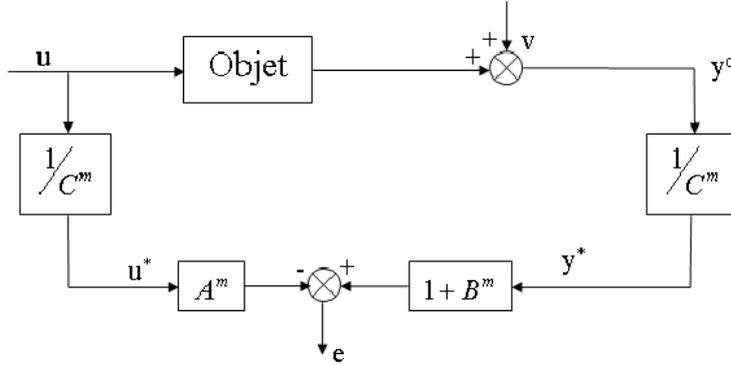


FIG. 3.2 – Méthode du maximum du vraisemblance

3.1.2 Méthode du maximum de vraisemblance (Astrom et Bohlin 1965)

Considérons le schéma de minimisation donné par la figure 3.2, où $C^m(z)$ fait partie du modèle. C^m est inconnue a priori et le bloc $1/C^m$ représente un *filtrage autoregressif*. Le critère à minimiser est :

$$Q = \sum_1^M e^2(k)$$

Le schéma de la méthode des moindres carrés est appliqué à partir des variables $y^*(k)$ et $u^*(k)$ d'après le modèle de prédiction (linéaire en paramètre) :

$$y^m(k) = - \sum_1^n \beta_i^m \cdot y^*(k-i) + \sum_0^n \alpha_j^m \cdot u^*(k-j)$$

$$e(k) = y^*(k) - y^m(k) = y^*(k) + \sum_1^n \beta_i^m \cdot y^*(k-i) - \sum_0^n \alpha_j^m \cdot u^*(k-j)$$

Seulement ici, $y^*(k)$ et $u^*(k)$ s'expriment à partir des variables observées ($u(k)$ et $y^o(k)$) par les équations récursives :

$$\begin{cases} y^*(k) = - \sum_1^n c_l^m \cdot y^*(k-l) + y^o(k) \\ u^*(k) = - \sum_1^n c_l^m \cdot u^*(k-l) + u(k) \end{cases}$$

où c_l^m sont des inconnues.

En explicitant l'écriture de l'erreur de prédiction $e(k)$ en fonction des variables mesurées $u(k)$ et $y^o(k)$, on remarque que $e(k)$ n'est pas une fonction linéaire du

3.1. MÉTHODES STOCHASTIQUES

vecteur paramètre modèle \mathbf{P}^m :

$$\mathbf{P}^m = \left[\alpha_0^m \quad \dots \quad \alpha_n^m \quad \vdots \quad \beta_1^m \quad \dots \quad \beta_n^m \quad \vdots \quad c_1^m \quad \dots \quad c_n^m \right]^T \quad (3.1)$$

La minimisation de Q ne peut être conduite directement et exige l'emploi d'algorithmes de programmation non linéaire. Voyons, par exemple, l'application de l'algorithme du gradient à la $K^{ième}$ itération :

$$\mathbf{P}^m(K+1) = \mathbf{P}^m(K) - \lambda^2 \cdot \left(\frac{\partial Q}{\partial \mathbf{P}^m} \right)_{\mathbf{P}^m = \mathbf{P}^m(K)}$$

Il s'agit d'exprimer $\frac{\partial Q}{\partial \mathbf{P}^m}$:

$$\frac{\partial Q}{\partial \mathbf{P}^m} = \frac{\partial}{\partial \mathbf{P}^m} \left[\sum_1^M e^2(k) \right] = \left[\sum_1^M 2 \cdot e(k) \cdot \frac{\partial e(k)}{\partial \mathbf{P}^m} \right]$$

distinguons $\frac{\partial e(k)}{\partial \alpha_j^m}$, $\frac{\partial e(k)}{\partial \beta_i^m}$, $\frac{\partial e(k)}{\partial c_l^m}$, avec :

$$e(k) = y^*(k) + \sum_1^n \beta_i^m \cdot y^*(k-i) - \sum_0^n \alpha_j^m \cdot u^*(k-j)$$

soit :

$$E(z) = \left(1 + \sum_1^n \beta_i^m \cdot z^{-i} \right) \cdot Y^*(z) - \left(\sum_0^n \alpha_j^m \cdot z^{-j} \right) \cdot U^*(z)$$

avec :

$$Y^*(z) = \frac{Y^o(z)}{1 + \sum_1^n c_l^m \cdot z^{-l}} \quad \text{et} \quad U^*(z) = \frac{U(z)}{1 + \sum_1^n c_l^m \cdot z^{-l}}$$

d'où :

$$\begin{cases} \mathcal{Z} \left(\frac{\partial e(k)}{\partial \alpha_j^m} \right) = -z^{-j} \cdot U^*(z) \\ \mathcal{Z} \left(\frac{\partial e(k)}{\partial \beta_i^m} \right) = +z^{-i} \cdot Y^*(z) \\ \mathcal{Z} \left(\frac{\partial e(k)}{\partial c_l^m} \right) = -\frac{z^{-l} \cdot E(z)}{1 + \sum_1^n c_l^m \cdot z^{-l}} \end{cases} \quad (3.2)$$

Ces équations 3.2 montrent comment sont calculées les composantes du vecteur gradient à partir des équations aux différences.

Les équations du modèle sont :

$$\begin{cases} y^*(k) = -\sum_1^n c_l^m \cdot y^*(k-l) + y^o(k) \\ u^*(k) = -\sum_1^n c_l^m \cdot u^*(k-l) + u(k) \\ e(k) = y^*(k) + \sum_1^n \beta_i^m \cdot y^*(k-i) - \sum_0^n \alpha_j^m \cdot u^*(k-j) \end{cases}$$

et les équations de sensibilité :

$$\begin{cases} \left(\frac{\partial e(k)}{\partial \alpha_j^m} \right) = -u^*(k-j) \\ \left(\frac{\partial e(k)}{\partial \beta_i^m} \right) = y^*(k-i) \\ \left(\frac{\partial e(k)}{\partial c_l^m} \right) = \sum_1^n c_l^m \cdot \frac{\partial e(k-l)}{\partial c_l^m} - e(k-l) \end{cases}$$

d'où :

$$\frac{\partial Q}{\partial p_i^m} = \left[\sum_1^M 2 \cdot e(k) \cdot \frac{\partial e(k)}{\partial p_i^m} \right]$$

A partir de ces éléments spécifiques à la méthode du maximum de vraisemblance, on se reportera aux *méthodes du modèle* pour une étude plus détaillée de la mise en oeuvre numérique de cette méthode. Ces méthodes du modèle seront explicitées plus loin dans ce chapitre.

Justifions toutefois l'appellation "*Maximum de Vraisemblance*".

Prenons le modèle caractérisé par l'équation stochastique :

$$y^m(k) = - \sum_1^n \beta_i^m \cdot y^m(k-i) + \sum_0^n \alpha_j^m \cdot u(k-j) + \nu(k) + \sum_1^n c_l^m \cdot \nu(k-l)$$

où $\nu(k)$ est un bruit blanc.

Ce qui peut s'écrire matriciellement sur l'horizon $[1, M]$, en supposons les conditions initiales nulles :

$$\begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \beta_1^m & 1 & 0 & \dots & \dots & 0 \\ \beta_2^m & \beta_1^m & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & 0 & \beta_n^m & \dots & 1 \end{bmatrix} \begin{bmatrix} y^m(1) \\ y^m(2) \\ \vdots \\ \vdots \\ \vdots \\ y^m(M) \end{bmatrix} = \begin{bmatrix} \alpha_0^m & 0 & \dots & \dots & \dots & 0 \\ \alpha_1^m & \alpha_0^m & 0 & \dots & \dots & 0 \\ \alpha_2^m & \alpha_1^m & \alpha_0^m & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n^m & \dots & \alpha_0^m \end{bmatrix} \begin{bmatrix} u(1) \\ u(2) \\ \vdots \\ \vdots \\ \vdots \\ u(M) \end{bmatrix} + \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ c_1^m & 1 & 0 & \dots & \dots & 0 \\ c_2^m & c_1^m & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & 0 & c_n^m & \dots & 1 \end{bmatrix} \begin{bmatrix} \nu(1) \\ \nu(2) \\ \vdots \\ \vdots \\ \vdots \\ \nu(M) \end{bmatrix}$$

$$\mathbf{B} \cdot \mathbf{Y}^m = \mathbf{A} \cdot \mathbf{U} + \mathbf{C} \cdot \boldsymbol{\nu}$$

soit :

$$\mathbf{Y}^m = \mathbf{B}^{-1} \cdot \mathbf{A} \cdot \mathbf{U} + \mathbf{B}^{-1} \cdot \mathbf{C} \cdot \boldsymbol{\nu} \quad (3.3)$$

3.1. MÉTHODES STOCHASTIQUES

On remarque par ailleurs que cette équation est non linéaire par rapport aux paramètres et que :

$$\mathbf{B}^{-1} \cdot \mathbf{C} \cdot \boldsymbol{\nu} = [v(1) \quad v(2) \quad \dots \quad v(M)]^T$$

Supposons que les expériences effectuées sur l'objet se déroulent toujours dans des conditions identiques :

- Conditions initiales nulles,
- Séquence de commande figée à $\{u(1), \dots, u(M)\}$,
- Horizon d'observation $[1, M]$.

L'expérience va alors fournir un vecteur de mesure \mathbf{Y}^o , résultat d'une expérience aléatoire.

La vraisemblance du modèle $\mathbf{Y}^m = \mathbf{B}^{-1} \cdot \mathbf{A} \cdot \mathbf{U} + \mathbf{B}^{-1} \cdot \mathbf{C} \cdot \boldsymbol{\nu}$ va se mesurer par la probabilité que l'on a d'avoir ce \mathbf{Y}^o comme résultat d'expérience (c'est-à-dire d'avoir $\mathbf{Y}^o = \mathbf{B}^{-1} \cdot \mathbf{A} \cdot \mathbf{U} + \mathbf{B}^{-1} \cdot \mathbf{C} \cdot \boldsymbol{\nu}$).

L'estimée au sens du maximum de vraisemblance est le vecteur paramètre qui maximise la densité de probabilité conditionnelle $P(\mathbf{Y}^o / \mathbf{P}^m)$ où \mathbf{Y}^o est le résultat numérique d'une expérience relative à l'objet.

Afin d'établir analytiquement cette condition, exprimons la loi de probabilité de \mathbf{Y}^m à partir de l'équation 3.3, dans le cas où $\boldsymbol{\nu}$ est un vecteur gaussien, centré et tel que la séquence $\nu(k)$ soit indépendante ($\nu(k)$ est un bruit blanc).

La variance de $\boldsymbol{\nu}$ s'écrit : $E\{\boldsymbol{\nu} \cdot \boldsymbol{\nu}^T\} = \sigma^2 \cdot \mathbf{I}$

d'où :

$$\bar{\mathbf{Y}} = E\{\mathbf{Y}^m\} = \mathbf{B}^{-1} \cdot \mathbf{A} \cdot \mathbf{U}$$

et :

$$\boldsymbol{\Sigma}_{\mathbf{Y}^m \mathbf{Y}^m} = E\{(\mathbf{Y}^m - \bar{\mathbf{Y}})(\mathbf{Y}^m - \bar{\mathbf{Y}})^T\} = \sigma^2 \cdot \mathbf{B}^{-1} \cdot \mathbf{C} \cdot [\mathbf{B}^{-1} \cdot \mathbf{C}]^T$$

Le vecteur $\boldsymbol{\nu}$ étant gaussien, le vecteur $\mathbf{Y}^m = \mathbf{B}^{-1} \cdot \mathbf{A} \cdot \mathbf{U} + \mathbf{B}^{-1} \cdot \mathbf{C} \cdot \boldsymbol{\nu}$ l'est aussi et est caractérisé par l'équation de probabilité :

$$P(\mathbf{Y}^m / \mathbf{P}^m) = \frac{1}{(2\pi)^{\frac{M}{2}} |\boldsymbol{\Sigma}_{\mathbf{Y}^m \mathbf{Y}^m}|^{\frac{M}{2}}} \cdot \exp \left\{ -\frac{1}{2} \cdot [(\mathbf{Y}^m - \bar{\mathbf{Y}})^T \cdot [\boldsymbol{\Sigma}_{\mathbf{Y}^m \mathbf{Y}^m}]^{-1} \cdot (\mathbf{Y}^m - \bar{\mathbf{Y}})] \right\}$$

or :

$$[\boldsymbol{\Sigma}_{\mathbf{Y}^m \mathbf{Y}^m}]^{-1} = \frac{1}{\sigma^2} \cdot [\mathbf{B}^{-1} \cdot \mathbf{C} \cdot (\mathbf{B}^{-1} \cdot \mathbf{C})^T]^{-1} = \frac{1}{\sigma^2} \cdot (\mathbf{C}^{-1} \cdot \mathbf{B})^T \cdot \mathbf{C}^{-1} \cdot \mathbf{B}$$

et :

$$|\boldsymbol{\Sigma}_{\mathbf{Y}^m \mathbf{Y}^m}| = \sigma^2$$

car les matrices \mathbf{B} et \mathbf{C} sont des matrices triangulaires de diagonale 1, donc de déterminant unité. En conséquence :

$$P(\mathbf{Y}^m / \mathbf{P}^m) = \frac{1}{(2\pi)^{\frac{M}{2}} \sigma^M} \exp \left\{ -\frac{1}{2\sigma^2} [\mathbf{C}^{-1} \mathbf{B} \mathbf{Y}^m - \mathbf{C}^{-1} \mathbf{A} \mathbf{U}]^T [\mathbf{C}^{-1} \mathbf{B} \mathbf{Y}^m - \mathbf{C}^{-1} \mathbf{A} \mathbf{U}] \right\}$$

La probabilité d'avoir \mathbf{Y}^o (résultat numérique relatif à une expérience) comme sortie du modèle caractérisé par le vecteur paramètre 3.1 est :

$$P(\mathbf{Y}^m / \mathbf{P}^m)_{\mathbf{Y}^m = \mathbf{Y}^o} = \frac{1}{(2\pi)^{\frac{M}{2}} \sigma^M} \cdot \exp \left\{ -\frac{1}{2\sigma^2} [\mathbf{C}^{-1}\mathbf{B}\mathbf{Y}^o - \mathbf{C}^{-1}\mathbf{A}\mathbf{U}]^T [\mathbf{C}^{-1}\mathbf{B}\mathbf{Y}^o - \mathbf{C}^{-1}\mathbf{A}\mathbf{U}] \right\}$$

Maximiser cette densité de probabilité est équivalent à maximiser son logarithme népérien :

$$\mathbf{L} = \log(P(\mathbf{Y}^o / \mathbf{P}^m)) = \lambda - M \log \sigma - \frac{1}{2\sigma^2} [\mathbf{C}^{-1}\mathbf{B}\mathbf{Y}^o - \mathbf{C}^{-1}\mathbf{A}\mathbf{U}]^T [\mathbf{C}^{-1}\mathbf{B}\mathbf{Y}^o - \mathbf{C}^{-1}\mathbf{A}\mathbf{U}]$$

Le maximum de cette fonction est obtenu pour :

$$\begin{cases} \frac{\partial \mathbf{L}}{\partial \mathbf{P}^m} = 0 \\ \frac{\partial \mathbf{L}}{\partial \sigma} = 0 \end{cases}$$

La fonction \mathbf{L} peut être interprétée comme une fonction de coût et l'on établit immédiatement la relation :

$$Q = \sum_1^M e^2(k) = [\mathbf{C}^{-1} \cdot \mathbf{B} \cdot \mathbf{Y}^o - \mathbf{C}^{-1} \cdot \mathbf{A} \cdot \mathbf{U}]^T \cdot [\mathbf{C}^{-1} \cdot \mathbf{B} \cdot \mathbf{Y}^o - \mathbf{C}^{-1} \cdot \mathbf{A} \cdot \mathbf{U}]$$

Ce qui justifie les équivalences suivantes :

$$\begin{cases} \frac{\partial \mathbf{L}}{\partial \mathbf{P}^m} = 0 \iff \frac{\partial Q}{\partial \mathbf{P}^m} = 0 \\ \frac{\partial \mathbf{L}}{\partial \sigma} = 0 \iff \hat{\sigma}^2 = \frac{1}{M} \sum_1^M e^2(k) \end{cases}$$

L'estimation au sens du maximum de vraisemblance nous permet d'établir certaines propriétés intéressantes de convergence. Nous retiendrons simplement que, pour la plupart des séquences d'entrée expérimentales $\{u(k)\}$, il est possible d'établir le résultat suivant :

$$\hat{\mathbf{P}}^m \longrightarrow \mathbf{P}^o \quad \text{lorsque} \quad M \longrightarrow \infty$$

c'est-à-dire que l'estimateur est asymptotiquement non biaisé et à variance nulle.

Exercice 9 *Refaites les calculs précédents !*

3.1.3 Méthode des moindres carrés généralisés (Clarcke 1967)

Cette éthode revient d'abord à appliquer la méthode des moindres carrés ordinaires et à considérer ensuite le résidu $e(k)$ comme première estimation du bruit $\zeta(k)$.

3.1. MÉTHODES STOCHASTIQUES

On sait, d'après les modèles précédents que $\zeta(k)$ est déduit, par filtrage à moyenne mobile, du bruit blanc $\nu(k)$:

$$\begin{aligned}\zeta(z) &= C(z).\nu(z) \\ \zeta(k) &= \sum_1^n c_l^m.\nu(k-l) + \nu(k)\end{aligned}$$

en confondant $\zeta(k)$ avec $e(k)$ il n'est cependant plus possible d'identifier les coefficients c_l^m car $\nu(k)$ est non mesurable. Cependant si l'on cherche à approcher le filtrage précédent par un filtrage récursif du type :

$$e(k) = -\sum_1^q r_i.e(k-i) + \nu(k)$$

L'estimation des r_i est possible par une méthode des moindres carrés à partir de la mise en équation matricielle suivante :

$$\begin{bmatrix} e(q+1) \\ \vdots \\ \vdots \\ e(M) \end{bmatrix} = \begin{bmatrix} e(q) & \cdots & \cdots & e(1) \\ \vdots & & & e(2) \\ \vdots & & & \vdots \\ e(M-1) & \cdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} -r_1 \\ \vdots \\ \vdots \\ -r_q \end{bmatrix} + \begin{bmatrix} \nu(q+1) \\ \vdots \\ \vdots \\ \nu(M) \end{bmatrix}$$

$$\mathbf{e} = \mathbf{E} \times \mathbf{r} + \boldsymbol{\nu}$$

d'où :

$$\hat{\mathbf{r}} = (\mathbf{E}^T \cdot \mathbf{E})^{-1} \cdot \mathbf{E}^T \cdot \mathbf{e}$$

Si la séquence $e(k)$ était une estimation non biaisée de $\zeta(k)$, $\hat{\mathbf{r}}$ ne présenteraient pas de biais car $\nu(k)$ est un bruit blanc. Malheureusement, $e(k)$ est biaisé car c'est le résidu de la méthode des moindres carrés ordinaire.

Cependant, en opérant les itérations suivantes, l'expérience montre que l'on aboutit à une estimation non biaisée des paramètres du modèle :

- a. Après la première initialisation des paramètres $\hat{\mathbf{r}}$ décrite précédemment, on opère un filtrage inverse (à moyenne mobile) :

$$\begin{cases} u^*(k) = \sum_1^q r_j.u(k-j) + u(k) \\ y^*(k) = \sum_1^q r_i.y^o(k-j) + y^o(k) \end{cases}$$

- b. A partir des nouvelles données $y^*(k)$ et $u^*(k)$ on effectue une estimation des paramètres $\{\alpha_i^m\}$ et $\{\beta_j^m\}$ par une méthode des moindres carrés ordinaire.
- c. On construit la nouvelle séquence d'erreur $e^*(k)$ (résidu relatif à l'estimation en (b)).
- d. A partir de $e^*(k)$, on effectue une estimation d'un filtre caractérisé par $\hat{\mathbf{r}}$.

On retourne ensuite à l'étape (a) en introduisant le nouveau filtre inverse *en série avec* les filtres déterminés aux itérations précédentes.

Le schéma de la figure à ajouter symbolise la $K^{\text{ième}}$ itération.

En fin de compte (à la $K^{\text{ième}}$ itération par exemple), le filtrage introduit afin d'éliminer le biais est :

$$\frac{1}{R_1(z).R_2(z)\dots R_K(z)}$$

L'ordre de ce filtre dépend directement du nombre d'itérations et on aura intérêt à choisir un ordre faible pour les filtres élémentaires (on peut d'ailleurs choisir $q = 1$ ce qui simplifie considérablement l'estimation de \mathbf{r} car dans ce cas ($\mathbf{E}^T \cdot \mathbf{E}$ est un scalaire)).

3.1.4 Conclusion

Dans la plupart des cas pratiques le modèle mathématique des perturbations n'est pas utilisé par la suite et il faut plutôt considérer les méthodes précédentes comme des méthodes permettant d'éliminer, en identification paramétrique, l'influence de ces bruits.

3.2 Méthodes du modèle (Méthode déterministe)

Les méthodes du modèle visent à aligner le comportement du modèle sur celui de l'objet à partir du *calcul effectif de l'erreur de sortie* :

$$\epsilon(k) = y^m(k) - y^o(k)$$

La sortie du modèle est calculée à partir des équations du modèle et des *seules observations de l'entrée*.

Par exemple, un modèle de prédiction du type $y^m(k) = \beta.y^o(k-1) + u(k)$ ne permet pas l'application d'une méthode du modèle.

Par contre, le modèle récursif $y^m(k) = \beta.y^m(k-1) + u(k)$ permet, en simulation, de calculer l'erreur de sortie $\epsilon(k)$ (cf. fig.3.3).

A partir de $\epsilon(k)$ on construit le plus souvent une fonction de coût ($Q = \sum_1^M \epsilon^2(k)$ par exemple) que l'on cherche à minimiser.

On montre qu'une telle minimisation ne conduit pas à un biais systématique lorsque le bruit v n'est pas corrélé avec la séquence d'entrée u .

Si de plus v est un bruit blanc gaussien, la méthode du modèle est *optimale* au sens du maximum de vraisemblance.

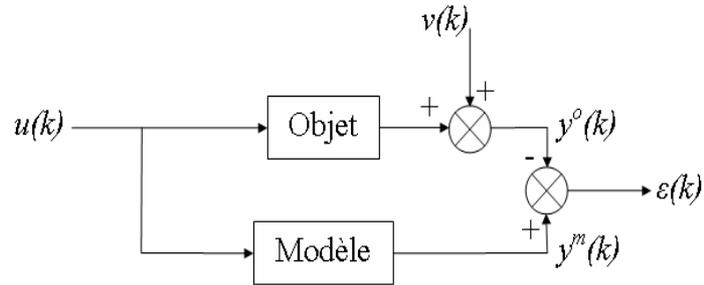


FIG. 3.3 – Méthode du Modèle

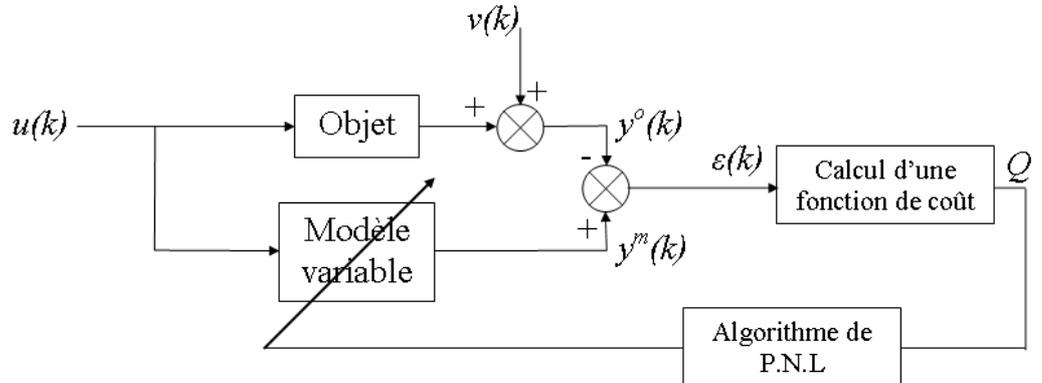


FIG. 3.4 – Principe de la méthode du modèle

3.2.1 Méthode du modèle appliquée à la détermination d'un modèle paramétrique

Le modèle paramétrique considéré ici est celui donné par l'équation aux différences :

$$y^m(k) = - \sum_{i=1}^n \beta_i^m \cdot y^m(k-i) + \sum_{j=0}^n \alpha_j^m \cdot u(k-j)$$

ou encore :

$$\frac{Y^m(z)}{U(z)} = \frac{A(z)}{1 + B(z)}$$

Le principe de la méthode réside dans la minimisation d'une fonction de coût par des méthodes de programmation non linéaire (méthodes itératives) (cf. fig. 3.4).

La mise en oeuvre de cette méthode exige donc :

1. Le choix d'une fonction de coût.
2. L'acquisition de séquences "entrée" et "sortie" de l'objet (réalisation d'une expérience).
3. Le calcul de la séquence de sortie du modèle afin d'élaborer l'erreur $\epsilon(k)$ puis la fonction de coût.
4. La modification des paramètres du modèle d'après un algorithme de P.N.L.
5. Le retour à l'étape 3 pour une mise en oeuvre en "temps différé", ou bien à l'étape 2 pour une mise en oeuvre de l'identification "en ligne".

Analysons ces différents points :

1. Questions liées au choix du critère

En général on choisit $Q = \sum_{i=1}^M \epsilon^2(k)$ (ou plus rarement $Q = \sum_{i=1}^M |\epsilon(k)|$). Le critère quadratique présente l'avantage de permettre une interprétation statistique aisée de l'estimation des paramètres (moindres carrés, maximum de vraisemblance), mais aussi de conduire à des développements analytiques simples.

L'horizon M de calcul de critère est à priori choisi de l'ordre de grandeur de la mémoire de l'objet.

L'identification "en ligne" pose en outre le problème de la stationnarité de la fonction de coût et par la même du choix des séquences d'entrée : En effet, dans ce mode de fonctionnement la diminution de la fonction de coût peut être simplement due à une diminution de l'énergie du signal d'entrée sur l'horizon M . La stationnarité de la fonction de coût est donc liée à la stationnarité du signal d'entrée.

A noter qu'en opérant en temps différé ce problème ne se pose pas car les séquences d'entrée et de sortie sont identiques pour toute la durée de la minimisation.

2. Acquisition des séquences "entrée" et "sortie"

Un bon signal d'entrée doit être riche en information c'est-à-dire sensibilisant vis-à-vis des paramètres du modèle, sur l'horizon $[1, M]$. Notons qu'une séquence binaire pseudo aléatoire **SBPA**, de période M , possède à la fois la richesse en information et la stationnarité sur l'horizon M .

Les caractéristiques du signal de sortie sont liés aux perturbations $v(k)$. Si l'on veut que l'estimation des paramètres du modèle (en temps différé, c'est-à-dire en opérant à partir d'enregistrements relatifs à une seule expérience) soit proche de la valeur optimale, il faut que le bruit $v(k)$ soit à variations rapides par rapport à l'horizon M (c'est-à-dire que ce bruit soit proche d'un bruit blanc).

Si le bruit $v(k)$ ne peut être considéré comme blanc par rapport à M , il est

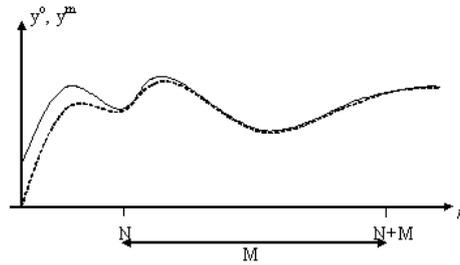


FIG. 3.5 – Sortie objet (C.I. $\neq 0$)- Sortie modèle (C.I. = 0)

nécessaire d'augmenter cet horizon.

A noter que l'identification "en ligne" ne présente pas tout à fait la même difficulté car on réalise un nombre d'expériences égal au nombre d'itération de l'algorithme *P.N.L.*

3. Simulation numérique du modèle

Cette simulation est effectuée en toute généralité à partir de l'équation aux différences :

$$y^m(k) = - \sum_{i=1}^n \beta_i^m . y^m(k-i) + \sum_{j=0}^n \alpha_j^m . u(k-j)$$

Le calcul de $y^m(k)$ exige la donnée de la séquence $u(k-j)$ mais aussi des valeurs précédentes de la sortie du modèle.

Pour $k < n$, il est nécessaire de se fixer des conditions initiales relatives à $y^m(-j)$ pour $0 \leq j \leq n$.

Le plus souvent on prendra $y^m(-j) = 0$ (modèle au repos à l'origine). Cette solution est correcte si l'objet est aussi initialement au repos, sinon les sorties du modèle ne seront prises en compte, dans le calcul du critère, qu'à partir d'un instant N correspondant approximativement à la mémoire de l'objet (cf. fig 3.5)

4. Minimisation du critère

Les méthodes de recherche de minimum sont extrêmement nombreuses. Retenons qu'il n'existe pas de méthode universelle dont la convergence soit assurée quel que soit le problème posé. Voyons cependant la mise en oeuvre de deux méthodes analytiques simples : la méthode du gradient et la méthode de Gauss-Newton.

Méthode du gradient :

$$P^m(K+1) = P^m(k) - \lambda^2 \cdot \left(\frac{\partial Q}{\partial P^m} \right)_{P^M = P^M(K)}$$

Méthode de Gauss-Newton :

$$P^m(K+1) = P^m(k) - \lambda^2 \cdot \left(B \cdot \frac{\partial Q}{\partial P^m} \right)_{P^M = P^M(K)}$$

où :

- $\frac{\partial Q}{\partial P^m}$ est le vecteur gradient.
- B est une matrice définie positive tendant vers l'inverse de la matrice Hessienne lorsque l'on se rapproche de l'optimum.
- K est le numéro de l'itération.
- Le calcul du gradient est mené en s'aidant des fonctions de sensibilité :

$$\frac{\partial Q}{\partial p_i^m} = \frac{\partial}{\partial p_i^m} \sum_{k=1}^M \epsilon^2(k) = 2 \cdot \sum_{k=1}^M \epsilon(k) \cdot \frac{\partial y^m(k)}{\partial p_i^m} = 2 \cdot \sum_{k=1}^M \epsilon(k) \cdot \sigma_i(k)$$

en posant $\sigma_i(k) = \frac{\partial y^m(k)}{\partial p_i^m}$, σ_i est appelé fonction de sensibilité.

Les dérivées secondes, éléments de la matrice Hessienne, peuvent être approchées par des expressions dépendant uniquement des $\sigma_i(k)$:

$$\frac{\partial^2 Q}{\partial p_i^m \partial p_j^m} = 2 \cdot \sum_{k=1}^M \sigma_j(k) \cdot \sigma_i(k) + 2 \cdot \sum_{k=1}^M \epsilon(k) \cdot \frac{\partial \sigma_i(k)}{\partial p_j^m}$$

soit :

$$\frac{\partial^2 Q}{\partial p_i^m \partial p_j^m} \simeq 2 \cdot \sum_{k=1}^M \sigma_j(k) \cdot \sigma_i(k)$$

La matrice B de l'algorithme de Gauss-Newton peut être choisie telle que $B^{-1} = [b_{ij}]$ avec $b_{ij} = 2 \cdot \sum_{k=1}^M \sigma_j(k) \cdot \sigma_i(k)$.

On note alors que B^{-1} est définie positive par construction et tend vers le Hessien lorsque $\epsilon(k) \rightarrow 0$.

5. Calcul des fonctions de sensibilité

L'équation de modèle est

$$\frac{Y^m(z)}{U(z)} = \frac{\sum_{j=0}^n \alpha_j^m \cdot z^{-j}}{1 + \sum_{i=1}^n \beta_i^m \cdot z^{-i}}$$

3.2. MÉTHODES DU MODÈLE (MÉTHODE DÉTERMINISTE)

Posons :

$$\sigma_i(z) = \mathcal{Z}(\sigma_i(k)) = \sum \sigma_i(k) \cdot z^{-k} = \frac{\partial Y^m(z)}{\partial p_i^m}$$

avec :

$$P^m = [\alpha_0^m \ \dots \ \alpha_n^m \mid \beta_1^m \ \dots \ \beta_n^m] = [p_1^m \ \dots \ p_{2n-1}^m]$$

d'où :

$$\sigma_{\alpha_i^m}(z) = z^{-i} \cdot \frac{U(z)}{1 + \sum_{j=1}^n \beta_j^m \cdot z^{-j}}$$

et :

$$\sigma_{\beta_j^m}(z) = -z^{-j} \cdot \frac{Y^m(z)}{1 + \sum_{j=1}^n \beta_j^m \cdot z^{-j}}$$

Les fonctions de sensibilité sont alors générées à partir d'équations aux différences (équations de sensibilité) comme le montre l'exemple suivant.

Exemple 10

$$\frac{Y^m(z)}{U(z)} = \frac{\alpha_0^m + \alpha_1^m \cdot z^{-1}}{1 + \beta_1^m \cdot z^{-1}}$$

d'où :

$$\left\{ \begin{array}{l} y^m(k) = -\beta_1^m \cdot y(k-1) + \alpha_0^m \cdot u(k) + \alpha_1^m \cdot u(k-1) \\ \sigma_{\alpha_0}(k) = -\beta_1^m \cdot \sigma_{\alpha_0}(k-1) + u(k) \\ \sigma_{\alpha_1}(k) = -\beta_1^m \cdot \sigma_{\alpha_1}(k-1) + u(k-1) \\ \sigma_{\beta_1}(k) = -\beta_1^m \cdot \sigma_{\beta_1}(k-1) + u(k-1) \end{array} \right\} \text{ modèle de sensibilité}$$

Ces équations (du modèle et de sensibilité) sont modifiées à chaque itération selon la loi d'ajustement adoptée.

6. Conclusion

La méthode du modèle présente l'avantage de se rattacher à des notions très concrètes qui peuvent être abordées dans un cadre purement déterministe.

La difficulté de mise en oeuvre venant essentiellement de la mauvaise convergence des algorithmes de P.N.L, on aura intérêt à choisir le modèle le plus efficace de manière à réduire au maximum le nombre de paramètres à identifier.

Rappelons que lorsque l'objet est continu, il est possible de transposer tout ou partie des résultats précédents en analogique, mais l'on considérera aussi avec attention les méthodes de simulation numérique d'un système continu.

3.2.2 Méthode du modèle appliquée à la détermination d'une séquence de pondération

Nous avons déjà vu que la méthode des moindres carrés ordinaire apportait une solution non biaisée au problème de l'identification des coefficients d'une séquence de pondération. On avait remarqué, par ailleurs, que la minimisation effective était celle de *l'erreur de sortie*. La méthode des moindres carrés appliquée dans ce cas est donc une méthode du modèle non itérative.

Par contre, si l'on met en oeuvre l'algorithme des moindres carrés récurrents, la méthode devient itérative et applicable en temps réel.

Nous allons donner une autre version itérative de la méthode du modèle fondée sur la minimisation de la fonction de coût (dite encore *distance de structure*) :

$$Q = \sum_{i=0}^{N-1} [a^m(i) - a^o(i)]^2$$

où $a^m(i)$ [$a^o(i)$] est le $i^{\text{ième}}$ coefficient de la séquence de pondération du modèle (de l'objet) et N est la mémoire approximative de l'objet.

Plaçons nous à la $K^{\text{ième}}$ itération : $Q(K) = \sum_{i=0}^{N-1} [a_K^m(i) - a^o(i)]^2$, puis à la $(K+1)^{\text{ième}}$ itération $Q(K+1) = \sum_{i=0}^{N-1} [a_{K+1}^m(i) - a^o(i)]^2$

Recherchons une loi de variation des paramètres du modèle de manière à rendre la différence $\Delta Q(K) = Q(K+1) - Q(K)$ la plus négative possible :

$$\begin{aligned} \Delta Q(K) &= \sum_{i=0}^{N-1} [a_K^m(i) + a_{K+1}^m(i) - 2.a^o(i)] \cdot [a_{K+1}^m(i) - a_K^m(i)] \\ \Delta Q(K) &= \left\{ \sum_{i=0}^{N-1} [a_{K+1}^m(i) - a_K^m(i)]^2 + 2 \underbrace{[a_K^m(i) - a^o(i)]}_{\text{différences inconnues}} [a_{K+1}^m(i) - a_K^m(i)] \right\} \end{aligned}$$

Afin d'éliminer les différences inconnues, imposons à la variation des paramètres de suivre la loi :

$$a_{K+1}^m(i) = a_K^m(i) + \theta u(K-i) \quad (3.4)$$

En effet dans ces conditions :

$$2 \sum_{i=0}^{N-1} [a_K^m(i) - a^o(i)] \cdot u(K-i) = 2\theta\epsilon(K) \quad (3.5)$$

d'où :

$$\Delta Q(K) = \theta^2 \sum_{i=0}^{N-1} u^2(K-i) + 2\theta\epsilon(K) \quad (3.6)$$

3.2. MÉTHODES DU MODÈLE (MÉTHODE DÉTERMINISTE)

La valeur de θ qui maximise la variation de $|\Delta Q|$ est :

$$\theta_{opt} = -\frac{\epsilon(K)}{\sum_{i=0}^{N-1} u^2(K-i)} \quad (3.7)$$

D'où la loi de variation (en introduisant un coefficient de convergence)

$$a_{K+1}^m(i) = a_K^m(i) - \lambda^2 \frac{\epsilon(K)u(K-i)}{\sum_{i=0}^{N-1} u^2(K-i)} \quad (3.8)$$

L'expérience a montré une très bonne convergence de cet algorithme pour $\lambda^2 = 0.3$.

Notons toutefois que :

- Etant donnée l'allure particulière des coefficients d'une séquence de pondération ($a(i) \rightarrow 0$ quand $i \rightarrow \infty$), la loi d'ajustement précédente conduit à des termes de correction de valeur relative trop importante pour $i \rightarrow N$. D'où la nécessité de pondérer le coefficient λ^2 par une fonction décroissante de i (par exemple on prendra $\lambda^2 = 0.3 \left(1 - \frac{i}{N}\right)$).
- Lorsque l'énergie de l'entrée est faible sur l'horizon N le terme $\frac{1}{\sum u^2(K-i)}$ risque l'être très important et l'on aura intérêt à le remplacer par $\frac{1}{A + \sum_{i=0}^{N-1} u^2(K-i)}$ où A est de l'ordre de grandeur de l'énergie du bruit sur l'horizon N ($A \simeq \sum_0^{N-1} v^2(k)$).
- Cet algorithme, applicable en temps réel, est proche de la formulation récurrente des moindres carrés dans sa phase d'initialisation.
- Le biais de l'estimation est nul lorsque le bruit $v(k)$ (ramené globalement en sortie de l'objet) n'est pas corrélé avec l'entrée $u(k)$.

Chapitre 4

Filtre de Kalman

Le filtre de Kalman est un estimateur d'état dans un environnement stochastique. Il a été introduit par **R. Kalman** (1960).

4.1 Principe

Un système dynamique bruité peut être décrit par les équations :

$$\begin{cases} x(k+1) = A.x(k) + B.u(k) + G.w(k) \\ y(k) = C.x(k) + v(k) \end{cases} \quad (4.1)$$

où :

- x : vecteur d'état de dimension n ,
- u : vecteur de commande de dimension m ,
- y : vecteur de mesure de dimension p ,
- w : vecteur de bruit du système de dimension r ,
- v : vecteur des bruits de mesure de dimension p .

w est un bruit blanc de moyenne nulle et de matrice de covariance $Q = E\{w.w^T\}$, qu'on écrit $w \sim (0, Q)$. Il représente *les erreurs de modélisation et les perturbations*.

v est un bruit blanc $v \sim (0, R)$ et il représente *l'imprécision des capteurs*. On suppose que v et w ne sont pas corrélés, c'est-à-dire $E\{v.w^T\} = 0$.

L'état initial du système x_0 est lui aussi inconnu, mais on en dispose de quelques connaissances sous forme de valeur moyenne (prédite) \bar{x}_0 et de matrice de covariance P_0 ; $x_0 \sim (\bar{x}_0, P_0)$. On suppose que x_0 est indépendant de v et de w (c'est-à-dire : $E\{x_0.v^T\} = E\{x_0.w^T\} = 0$).

Le but est de concevoir un estimateur qui fournit des estimations de l'état $x(k)$, en tenant compte des dynamiques données par l'équation 4.1 et des mesures de sortie $y(k)$.

Propagation des moyennes et des covariances

La moyenne de l'état x se propage comme suit :

$$\begin{aligned}\bar{x}(k+1) &= A\bar{x}(k) + B\bar{u}(k) + G\bar{w}(k) \\ \bar{x}(k+1) &= A\bar{x}(k) + B.u(k) \\ \bar{x}(0) &= \bar{x}_0\end{aligned}\tag{4.2}$$

Ceci implique que la moyenne de x se propage suivant des dynamiques déterministes.

Pour trouver comment se propage la covariance de x , écrivons (où E est l'opérateur de l'espérance mathématique) :

$$\begin{aligned}P_{x(k+1)} &= E \left\{ [x(k+1) - \bar{x}(k+1)] [x(k+1) - \bar{x}(k+1)]^T \right\} \\ P_{x(k+1)} &= E \left\{ [A(x(k) - \bar{x}(k)) + Gw(k)] [A(x(k) - \bar{x}(k)) + Gw(k)]^T \right\} \\ P_{x(k+1)} &= AE \left\{ (x(k) - \bar{x}(k)) (x(k) - \bar{x}(k))^T \right\} A^T + GE \left\{ w(k)w^T(k) \right\} G^T \\ &\quad + GE \left\{ w(k) (x(k) - \bar{x}(k))^T \right\} A^T + AE \left\{ (x(k) - \bar{x}(k)) .w^T(k) \right\} G^T \\ P_{x(k+1)} &= AP_{x(k)}A^T + GQG^T\end{aligned}\tag{4.3}$$

La valeur moyenne de la sortie est :

$$\bar{y}(k) = C.\bar{x}(k)\tag{4.4}$$

La covariance entre l'état x et la sortie y est :

$$\begin{aligned}P_{x(k),y(k)} &= E \left\{ (x(k) - \bar{x}(k)) (y(k) - \bar{y}(k))^T \right\} \\ P_{x(k),y(k)} &= E \left\{ (x(k) - \bar{x}(k)) [C(x(k) - \bar{x}(k)) + v(k)]^T \right\} \\ P_{x(k),y(k)} &= P_{x(k)}C^T\end{aligned}$$

La covariance de la sortie y est :

$$\begin{aligned}P_{y(k)} &= E \left\{ (y(k) - \bar{y}(k)) (y(k) - \bar{y}(k))^T \right\} \\ P_{y(k)} &= E \left\{ [C(x(k) - \bar{x}(k)) + v(k)] [C(x(k) - \bar{x}(k)) + v(k)]^T \right\} \\ P_{y(k)} &= CP_{x(k)}C^T + R\end{aligned}$$

Estimation linéaire optimale de $x(k)$ à partir de $y(k)$

Supposons que, pour une matrice F et un vecteur g donnés :

$$\hat{x}(k) = Fy(k) + g$$

On définit $\hat{x}(k|k-1)$ comme étant l'estimation de $x(k)$ à partir des informations jusqu'à l'instant $k-1$.

Pour trouver le meilleur choix de F et de g , nous allons minimiser l'erreur quadratique moyenne :

$$\begin{aligned} J &= E \{ \hat{x}^T(k) \hat{x}(k) \} \\ J &= E \left\{ [x(k) - \hat{x}(k)]^T [x(k) - \hat{x}(k)] \right\} \\ J &= tr \left\{ E [x(k) - \hat{x}(k)] [x(k) - \hat{x}(k)]^T \right\} \\ J &= tr \left\{ E [x(k) - Fy(k) - g] [x(k) - Fy(k) - g]^T \right\} \\ J &= tr \left\{ E [x(k) - \hat{x}(k|k-1) - (Fy(k) + g - \hat{x}(k|k-1))] [x(k) - \hat{x}(k|k-1) - (Fy(k) + g - \hat{x}(k|k-1))]^T \right\} \end{aligned}$$

où $tr\{\cdot\}$ représente la *trace* de la matrice $\{\cdot\}$. Après quelques simplifications algébriques, on obtient :

$$J = tr \left\{ \begin{array}{l} P(k|k-1) + F(P_{y(k)} + \bar{y}(k)\bar{y}^T(k))F^T + (g - \hat{x}(k|k-1))(g - \hat{x}(k|k-1))^T \\ + 2F\bar{y}(k)(g - \hat{x}(k|k-1))^T - 2FP_{y(k)x(k)} \end{array} \right\}$$

Notons que nous avons utilisé les définitions suivantes :

$$\begin{aligned} \bar{y}(k) &= C\hat{x}(k|k-1) \\ P_{y(k)x(k)} &= E \left\{ (y(k) - \bar{y}(k))(x(k) - \hat{x}(k|k-1))^T \right\} \\ P(k|k-1) &= E \left\{ (x(k) - \hat{x}(k|k-1))(x(k) - \hat{x}(k|k-1))^T \right\} \end{aligned}$$

En utilisant les identités matricielles (valables pour toutes matrices F, H et D) :

$$\begin{aligned} \frac{d}{dF} \{tr [FHF^T]\} &= 2HF \\ \frac{d}{dF} \{tr [DFH]\} &= D^T H^T \end{aligned}$$

Nous avons pour minimum :

$$\begin{aligned} \frac{\partial J}{\partial g} &= 2(g - \hat{x}(k|k-1)) + 2F\bar{y}(k) = 0 \\ \frac{\partial J}{\partial F} &= 2F \cdot (P_{y(k)} + \bar{y}(k)\bar{y}(k)^T) - 2P_{x(k)y(k)} + 2(g - \hat{x}(k|k-1))\bar{y}(k)^T = 0 \end{aligned}$$

Notons que : $P_{x(k)y(k)} = P_{y(k)x(k)}^T$

La solution est donc :

$$\begin{aligned} g &= \hat{x}(k|k-1) - F\bar{y}(k) \\ F &= P_{x(k)y(k)} \cdot P_{y(k)}^{-1} \end{aligned}$$

Ce qui donne, après les substitutions appropriées :

$$\hat{x}(k|k) = \hat{x}(k|k-1) + P(k|k-1).C^T . [C.P(k|k-1).C^T + R]^{-1} . [y(k) - C.\hat{x}(k|k-1)]$$

L'erreur de covariance postérieure

Pour voir jusqu'à quel point l'estimation $\hat{x}(k|k)$ est précise, l'erreur de covariance postérieure $P(k|k)$ peut être calculée :

$$\begin{aligned} P(k|k) &= E \left\{ [x(k) - \hat{x}(k|k)] [x(k) - \hat{x}(k|k)]^T \right\} \\ P(k|k) &= E \left\{ \left[\begin{array}{l} x(k) - \hat{x}(k|k-1) - P_{x(k)y(k)}.P_{y(k)}^{-1} . (y(k) - \bar{y}(k)) \\ x(k) - \hat{x}(k|k-1) - P_{x(k)y(k)}.P_{y(k)}^{-1} . (y(k) - \bar{y}(k)) \end{array} \right]^T \right\} \\ P(k|k) &= P(k|k-1) - P(k|k-1).C^T . (C.P(k|k-1).C^T + R)^{-1} . C.P(k|k-1) \end{aligned}$$

4.2 L'algorithme discret du filtre de Kalman

- *Initialisation* : On fixe $x(0) = 0$, $P(0) = P_{x0}$, $k = 1$
- *Etape 1* (time update) : Etant donné $x(k-1|k-1)$ et $P(k-1|k-1)$, on applique les équations 4.2 et 4.3 (Effets des dynamiques du système) :

$$\begin{aligned} \hat{x}(k|k-1) &= A.\hat{x}(k-1|k-1) + B.u(k-1) \\ P(k|k-1) &= AP(k-1|k-1)A^T + GQG^T \end{aligned}$$

pour obtenir $\hat{x}(k|k-1)$, $P(k|k-1)$.

- *Etape 2* : Ensuite, après l'obtention des nouvelles mesures $y(k)$, on applique l'actualisation de mesure suivante (effets des mesures) :

$$\begin{aligned} K(k) &= P(k|k-1).C^T . (C.P(k|k-1).C^T + R)^{-1} \\ P(k|k) &= (I - K(k).C) . P(k|k-1) \\ \hat{x}(k|k) &= \hat{x}(k|k-1) + K(k) . [y(k) - C.\hat{x}(k|k-1)] \end{aligned}$$

Pour obtenir les estimations optimales $P(k|k)$, $\hat{x}(k|k)$.

$K(k)$ est appelé le *gain de Kalman*.

Faire $k = k + 1$ et retourner à l'étape 1.

4.3 Schéma structurel du Filtre de Kalman

La figure 4.1 donne un schéma synoptique du filtre de Kalman.

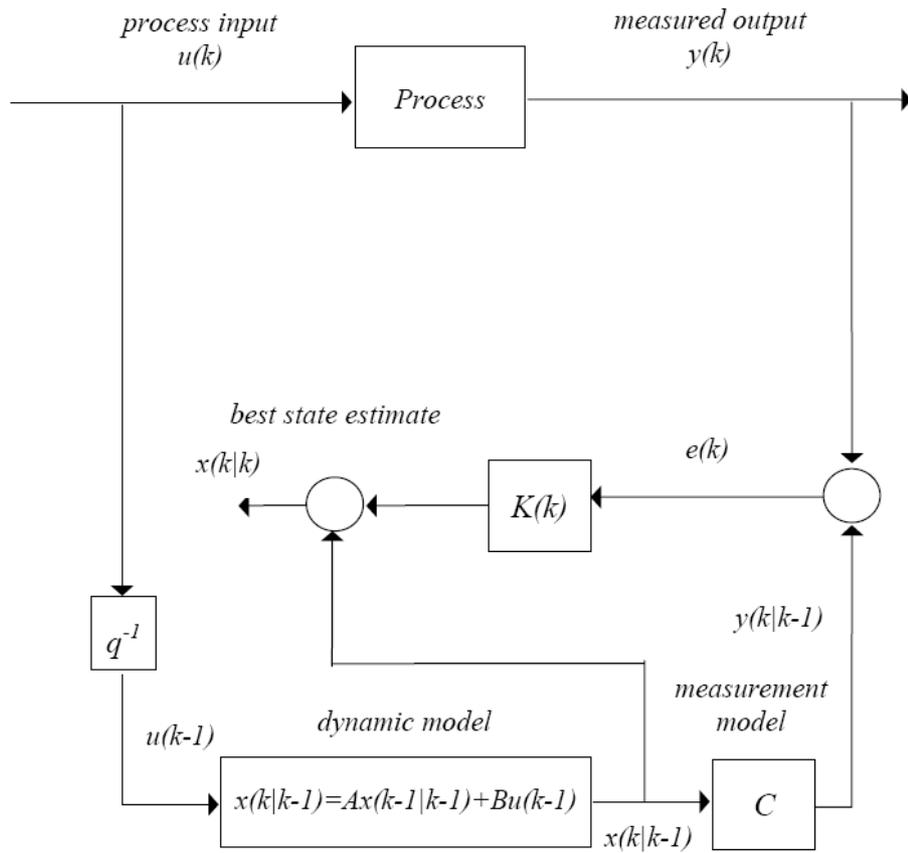


FIG. 4.1 – Schéma structurel du Filtre de Kalman

4.4 Filtre de Kalman sous-optimal du régime permanent

Même si le modèle du système (A, B, C) est invariant dans le temps, le filtre de Kalman optimal est variant dans le temps, parce que le gain de Kalman $K(k)$ est fonction du temps.

Il est souvent satisfaisant d'utiliser un filtre simplifié invariant, avec un gain constant K .

Au régime permanent statistique, l'erreur de covariance à priori $P(k|k-1)$ atteint une valeur finale constante, qu'on appelle P . L'équation de covariance peut être écrite comme suit :

$$P = A [P - PC^T(CPC^T + R)^{-1}CP] A^T + GQG \quad (4.5)$$

L'équation 4.5 est appelée **l'équation algébrique de Ricatti**, dont la solution est P .

Le gain de Kalman au régime permanent est une matrice constante $(n \times p)$.

$$K = PC^T (CPC^T + R)^{-1}$$

Le filtre de Kalman au régime permanent est le système invariant :

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K [y(k) - C\hat{x}(k|k-1)]$$

Sous **Matlab**, en utilisant la *Control Systems Toolbox*, l'expression :

$$\gg K = dlqe(A, G, C, Q, R)$$

permet d'obtenir le gain optimal du filtre de Kalman.

Exemple 11 *Etant donné le système :*

$$\begin{aligned} x(k+1) &= Ax(k) + B(u(k) + w(k)) \\ x(0) &= [1 \ 2]^T \end{aligned}$$

avec :

$$\begin{aligned} A &= \begin{bmatrix} 0.8 & 0.1 \\ 0 & 0.2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [1 \ 0], \quad h = 0.1 \\ v &\sim (0, 0.0001), \quad w \sim (0, 0.0001) \end{aligned}$$

Le gain de Kalman constant (au régime permanent) est :

$$K = \begin{bmatrix} 0.0358 \\ 0.0237 \end{bmatrix}$$

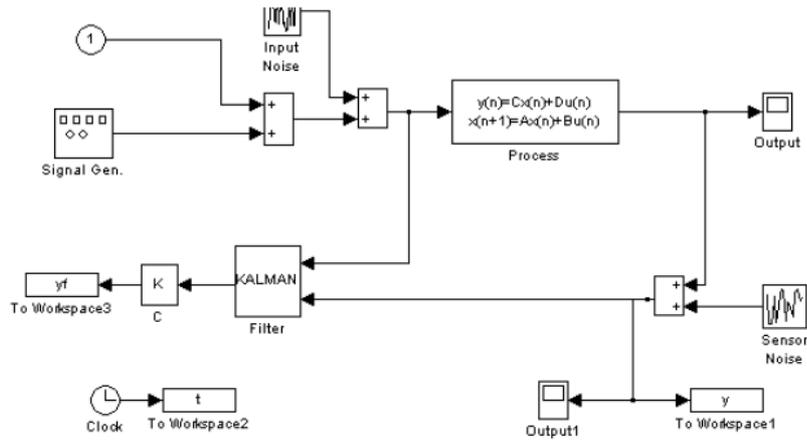


FIG. 4.2 – Exemple Simulink du filtre de Kalman

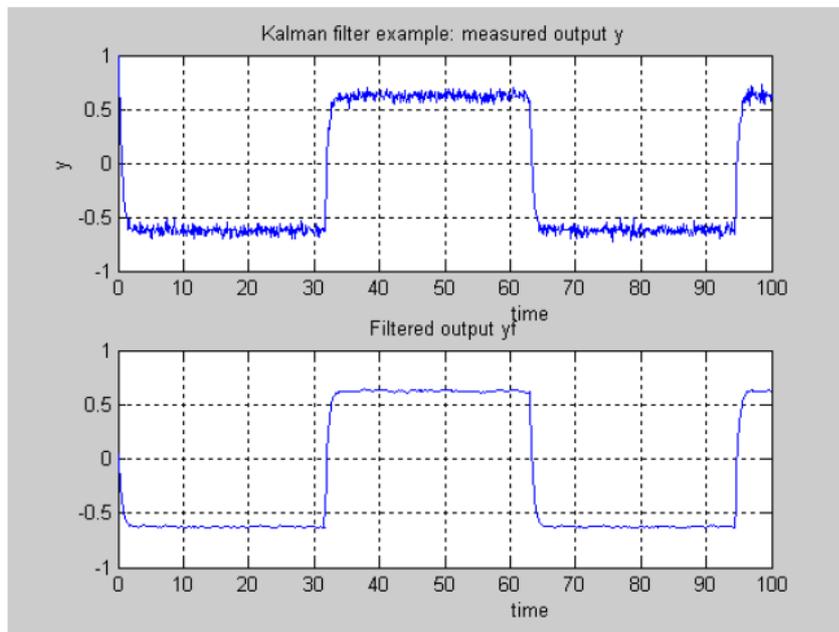


FIG. 4.3 –