Analyse et correction des Systèmes linéaires continus ou échantillonnés à l'aide des variables d'état

Gonzalo Cabodevila
gonzalo.cabodevila@femto-st.fr

2ème année
Semestre vert
Automatique avancée
filière EAOI

École Nationale Supérieure de
Mécanique et des Microtechniques
26, chemin de l'Épitaphe
25030 Besançon cedex – FRANCE
http://intranet-tice.ens2m.fr

Technologies de l'Information et de la Communication pour l'Enseignement

Table des matières

Chapitre 1

Exemple introductif : fil rouge

1.1 Différentes représentations d'un système physique

Soit un moteur à courant continu commandé par l'inducteur.

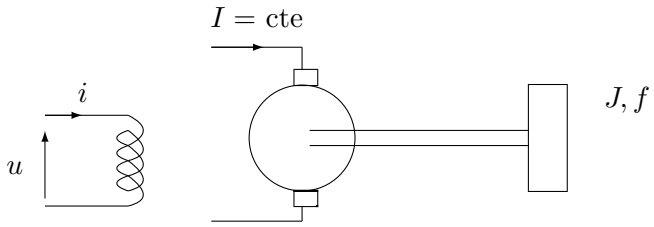


FIGURE 1.1 – Moteur à courant continu commandé par l'inducteur

— commande : u
— sortie : ω

Le système est monovariable, linéaire invariant dans le temps, il peut donc être représenté par une équation différentielle à coefficients constants.

1.1.1 Equations différentielles

Le système représenté en figure 1.1 est décrit par les équations suivantes :

$$u = Ri + L\frac{di}{dt} \tag{1.1}$$

$$J\frac{d\omega}{dt} + f\omega = \gamma \tag{1.2}$$

$$\gamma = ki \tag{1.3}$$

$$J\frac{d^2\omega}{dt^2} + f\frac{d\omega}{dt} = k\frac{di}{dt} = \frac{k}{L}(u - Ri) = \frac{k}{L}\left(u - \frac{R}{k}\left(\frac{d\omega}{dt} + f\omega\right)\right) \tag{1.4}$$

$$J\frac{d^2\omega}{dt^2} + \left(f + \frac{RJ}{L}\right)\frac{d\omega}{dt} + \frac{Rf}{L}\omega = \frac{k}{L}u \tag{1.5}$$

$$\frac{d^2\omega}{dt^2} + \left(\frac{f}{J} + \frac{R}{L}\right)\frac{d\omega}{dt} + \frac{Rf}{LJ}\omega = \frac{k}{LJ}u \tag{1.6}$$

$$\frac{d^2\omega}{dt^2} + a_1\frac{d\omega}{dt} + a_0\omega = b_0 u \tag{1.7}$$

Dès lors $\omega(t)$ est connu si u(t) et les deux conditions initiales ($\omega(0)$ et $\frac{d\omega(0)}{dt}$) sont connues.

1.1.2 Fonction de transfert

En utilisant la transformation de Laplace :

$$\mathcal{L}\left[e(t)\right] = \int_0^\infty e^{-pt} e(t) dt, \tag{1.8}$$

$$\mathcal{L}\left[\frac{de(t)}{dt}\right] = pE(p) - e(0), \tag{1.9}$$

nous obtenons la transformée de (1.7) :

$$p^2\Omega(p) - p\Omega(0) - \dot{\Omega}(0) + a_1(p\Omega(p) - \Omega(0) + a_0\Omega(p) = b_0 U(p) \tag{1.10}$$

d'où :

$$\Omega(p) = \frac{b_0}{p^2 + a_1 p + a_0} U(p) + \frac{\Omega(0)(a_1 + p) + \dot{\Omega}(0)}{p^2 + a_1 p + a_0} \tag{1.11}$$

Si les conditions initiales sont nulles :

$$\Omega(p) = \frac{b_0}{p^2 + a_1 p + a_0} U(p) \tag{1.12}$$

$$\frac{\Omega(p)}{U(p)} = \frac{b_0}{p^2 + a_1 p + a_0} = H(p) \tag{1.13}$$

$H(p)$ est la fonction de transfert du système.

1.1.3 Réponse impulsionnelle

Si les conditions initiales sont nulles :

$$\Omega(p) = H(p) \times U(p) \tag{1.14}$$

En repassant en temporel, la multiplication est transformée en une convolution

$$\omega(t) = h(t) \star u(t) = \int_0^\infty h(\tau) u(t - \tau) d\tau \tag{1.15}$$

donc

$$h(t) = \frac{k}{RJ - Lf} \left(e^{-\frac{f}{J}t} - e^{-\frac{R}{L}t} \right) \tag{1.16}$$

1.1.4 Représentation d'état

Si l'on désire réaliser une simulation analogique du système à partir d'intégrateurs, l'équation (1.7) peut se mettre sous la forme :

$$\frac{d^2\omega}{dt^2} = b_0 u - a_1 \frac{d\omega}{dt} - a_0 \omega \tag{1.17}$$

d'où le schéma suivant,
Les variables d'état sont les sorties des intégrateurs.

$$x_1 = \omega, \qquad x_2 = \frac{d\omega}{dt} = \dot{\omega} \tag{1.18}$$

Le système peut être représenté par les deux équations suivantes,

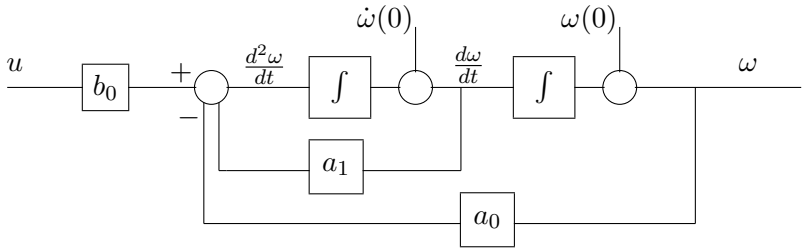


FIGURE 1.2 – Schéma d'un simulateur analogique

$$\dot{x_1} = x_2 \tag{1.19}$$
$$\dot{x_2} = b_0 u - a_0 x_1 - a_1 x_2 \tag{1.20}$$

La représentation utilisée classiquement est la représentation matricielle, on définit alors un vecteur d'état,

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1.21}$$

Equation d'état :

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b_0 \end{bmatrix} u \tag{1.22}$$

Equation de sortie :

$$\omega = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1.23}$$

équation d'état + équation de sortie = représentation d'état

Dans le cas général, l'écriture de la représentation d'état est la suivante,

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{1.24}$$
$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \tag{1.25}$$

Dans les cas qui nous intéressent c'est-à-dire les systèmes à une seule entrée et une seule sortie, l'écriture générale de la représentation d'état est la suivante,

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}u \tag{1.26}$$
$$y = \mathbf{C}\underline{x} + \mathbf{D}u \tag{1.27}$$

Dimensions : si le vecteur d'état est de dimension n, alors
- **A** est de dimension n lignes $\times n$ colonnes,
- **B** est de dimension n lignes $\times 1$ colonne,
- **C** est de dimension 1 ligne $\times n$ colonnes,
- **D** est une constante (très souvent nulle).

1.2 Propriétés de la représentation d'état

1.2.1 Non unicité de la représentation d'état

La représentation d'état n'est pas unique, dans l'exemple traité jusqu'à présent nous avons choisi le vecteur d'état suivant

$$x = \begin{bmatrix} \omega \\ \dot{\omega} \end{bmatrix} \tag{1.28}$$

nous aurions pu choisir un autre vecteur d'état, par exemple

$$x = \begin{bmatrix} i \\ \omega \end{bmatrix} \tag{1.29}$$

En effet, en reprenant les équations fondamentales du système

$$u = Ri + L\frac{di}{dt} \tag{1.30}$$

$$J\frac{d\omega}{dt} + f\omega = ki \tag{1.31}$$

avec ce nouveau vecteur d'état elle peuvent être écrites sous la forme

$$U = Rx_1 + L\dot{x_1} \tag{1.32}$$

$$J\dot{x_2} + fx_2 = kx_1 \tag{1.33}$$

d'où la représentation d'état
Equation d'état :

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} -\frac{R}{L} & 0 \\ \frac{k}{J} & -\frac{f}{J} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} u \tag{1.34}$$

Equation de sortie :

$$\omega = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1.35}$$

Notez que nous n'aurions pas pu prendre i et $\frac{di}{dt}$ comme vecteur d'état car $\frac{di}{dt}$ n'est pas une sortie d'intégrateur.
En fait, on peut prendre comme vecteur d'état n'importe quelle combinaison linéaire d'un vecteur d'état valable.
Soit $x' = \mathbf{M}x$

$$\dot{\underline{x'}} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}\underline{x'} + \mathbf{M}^{-1}\mathbf{B}u \tag{1.36}$$

$$y = \mathbf{C}\mathbf{M}\underline{x'} + \mathbf{D}u \tag{1.37}$$

1.2.2 Matrice de transfert

En prenant la transformée de Laplace de la représentation d'état, on obtient,

$$p\underline{X}(p) - \underline{x}(0) = \mathbf{A}\underline{X}(p) + \mathbf{B}U(p) \tag{1.38}$$

$$Y(p) = \mathbf{C}\underline{X}(p) + \mathbf{D}U(p) \tag{1.39}$$

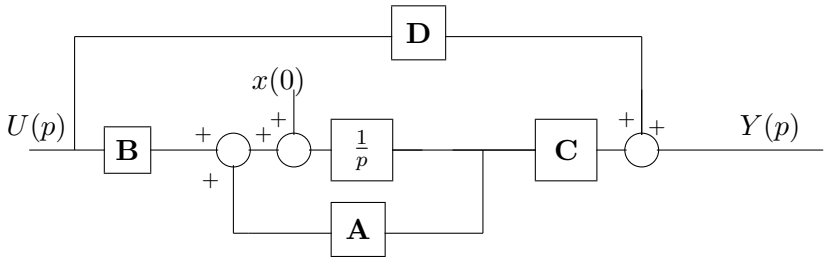


FIGURE 1.3 – Schéma général après transformation de Laplace

les équations (1.38) et (1.39) se réécrivent sous la forme

$$[p\mathbf{I}-\mathbf{A}]\,\underline{X}(p) = \underline{x}(0)+\mathbf{B}U(p) \quad (1.40)$$
$$Y(p) = \mathbf{C}\underline{X}(p)+\mathbf{D}U(p) \quad (1.41)$$

$$\underline{X}(p) = =[p\mathbf{I}-\mathbf{A}]^{-1}\underline{x}(0)+[p\mathbf{I}-\mathbf{A}]^{-1}\mathbf{B}U(p) \quad (1.42)$$
$$Y(p) = \mathbf{C}[p\mathbf{I}-\mathbf{A}]^{-1}\underbrace{\underline{x}(0)}_{\text{C I}}+\underbrace{\Big[\mathbf{C}[p\mathbf{I}-\mathbf{A}]^{-1}\mathbf{B}+\mathbf{D}\Big]}_{\text{matrice de transfert}}U(p) \quad (1.43)$$

en posant

$$\mathbf{T}=\Big[\mathbf{C}[p\mathbf{I}-\mathbf{A}]^{-1}\mathbf{B}+\mathbf{D}\Big] \quad (1.44)$$

on obtient

$$Y(p)=\mathbf{T}U(p)+\mathbf{C}[p\mathbf{I}-\mathbf{A}]^{-1}\underline{x}(0) \quad (1.45)$$

Note : dans le cas multivariable (plusieurs entrée, plusieurs sorties)

$$\mathbf{T}_{(i,j)}(p)=\frac{Y_i(p)}{U_j(p)} \quad (1.46)$$

Dans le cas SISO la matrice $\mathbf{T}$ représente la fonction de transfert du système.
La matrice $\mathbf{T}$ est unique, elle ne dépend pas de la représentation d'état choisie, elle ne dépend que des entrées et des sorties.
Démonstration :
soit

$$p\underline{X}(p) = \mathbf{A}\underline{X}(p)+\mathbf{B}U(p) \quad (1.47)$$
$$Y(p) = \mathbf{C}\underline{X}(p)+\mathbf{D}U(p) \quad (1.48)$$

posons

$$\underline{X}=\mathbf{K}\underline{X}' \quad (1.49)$$

donc

$$p\underline{X}'(p) = \mathbf{K}^{-1}\mathbf{A}\mathbf{K}\underline{X}'(p)+\mathbf{K}^{-1}\mathbf{B}U(p) \quad (1.50)$$
$$Y(p) = \mathbf{C}\mathbf{K}\underline{X}'(p)+\mathbf{D}U(p) \quad (1.51)$$

$$\begin{array}{rcll} \mathbf{T} & = & \mathbf{CK}\left[p\mathbf{I}-\mathbf{K}^{-1}\mathbf{AK}\right]^{-1}\mathbf{K}^{-1}\mathbf{B}+\mathbf{D} & \text{en utilisant } [\mathbf{AB}]^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1} \\ & = & \mathbf{CK}[\mathbf{K}p\mathbf{I}-\mathbf{AK}]^{-1}\mathbf{B}+\mathbf{D} & \text{en utilisant } [\mathbf{AB}]^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1} \\ & = & \mathbf{C}\left[\mathbf{K}p\mathbf{I}\mathbf{K}^{-1}-\mathbf{A}\right]^{-1}\mathbf{B}+\mathbf{D} & p\mathbf{I} \text{ est une matrice diagonale} \\ & = & \mathbf{C}[p\mathbf{I}-\mathbf{A}]^{-1}\mathbf{B}+\mathbf{D} & \text{CQFD} \end{array} \tag{1.52}$$

T est bien indépendante de la représentation d'état choisie.

1.3 Intérêt de cette représentation

1. Systèmes linéaires invariants
   — → simulation du système, analogique ou numérique
   — → extension aux systèmes multivariables
   — →notions nouvelles de commandabilité et d'observabilité
2. Systèmes linéaires variants (**A**, **B**, **C**, **D** dépendent de t)
   — → simulation
3. Systèmes non linéaires
   — → simulation, étude de la dynamique
4. Qu'est-ce que l'état ?
   — l'état d'un système à un instant donné représente la mémoire minimale du passé nécessaire à la détermination du futur.
   — les variables d'état doivent apporter une description interne du système et on choisit celles pour lesquelles on peut définir l'état initial, c'est-à-dire celles qui décrivent des "réservoirs" d'énergie

1.4 Résolution des équations d'état

soit le système :

$$\begin{array}{rcl} \dot{\underline{x}} & = & \mathbf{A}\underline{x}+\mathbf{B}u \\ y & = & \mathbf{C}\underline{x}+\mathbf{D}u \end{array} \tag{1.53}$$

1.4.1 Cas simple

Dans le cas simple ou x ne contient qu'un seule composante, ce système devient

$$\frac{dx}{dt}=ax+bu \tag{1.54}$$

En régime libre, la solution est

$$x(t)=ke^{at} \tag{1.55}$$

En régime forcé, la méthode de la variation de la constante donne

$$x(t) = k(t)e^{at} \tag{1.56}$$

$$ak(t)e^{at}+k'(t)e^{at} = ak(t)e^{at}+bu(t) \tag{1.57}$$

$$k'(t) = bu(t)e^{-at} \tag{1.58}$$

$$k(t) = \int_0^t bu(\tau)e^{-a\tau}d\tau+cte \tag{1.59}$$

d'où

$$x(t) = e^{at}\left[\int_0^t bu(\tau)e^{-a\tau}d\tau + cte\right] \tag{1.60}$$

en identifiant $x(0) = x_0$

$$x(t) = e^{at}\left[\int_0^t bu(\tau)e^{-a\tau}d\tau\right] + e^{at}x_0 \tag{1.61}$$

en généralisant

$$x(t) = \int_{t_0}^t bu(\tau)e^{-a(t-\tau)}d\tau + e^{a(t-t_0)}x(t_0). \tag{1.62}$$

1.4.2 Cas général

$$\underline{x}(t) = e^{\mathbf{A}t}\left[\int_{t_0}^{t} e^{-\mathbf{A}\tau}\mathbf{B}u(\tau)d\tau\right] + e^{\mathbf{A}(t-t_0)}\underline{x}(t_0) \tag{1.63}$$

Définition :

$$e^{\mathbf{A}t} = 1 + \mathbf{A}t + \frac{1}{2}(\mathbf{A}t)^2 + \frac{1}{3!}(\mathbf{A}t)^3 + \cdots \tag{1.64}$$

propriétés de $e^{\mathbf{A}t}$

$$e^{\mathbf{A}t}e^{\mathbf{B}t} = e^{(\mathbf{A}+\mathbf{B})t} \qquad \text{si} \quad \mathbf{AB} = \mathbf{BA} \tag{1.65}$$

$$\frac{d}{dt}e^{\mathbf{A}t} = \mathbf{A}e^{\mathbf{A}t} \tag{1.66}$$

$$\int_{t_1}^{t_2} e^{\mathbf{A}\tau}d\tau = \mathbf{A}^{-1}\left[e^{\mathbf{A}t_2} - e^{\mathbf{A}t_1}\right] = \left[e^{\mathbf{A}t_2} - e^{\mathbf{A}t_1}\right]\mathbf{A}^{-1} \tag{1.67}$$

1.4.3 Généralisation aux systèmes variants dans le temps

En posant $\Phi(t, t_0)$ solution de l'équation

$$\dot{\Phi}(t, t_0) = \mathbf{A}(t)\Phi(t, t_0) \tag{1.68}$$

alors

$$\underline{x}(t) = \int_{t_0}^{t} \Phi(t, \tau)\mathbf{B}(\tau)u(\tau)d\tau + \Phi(t, t_0)x(t_0) \tag{1.69}$$

Sauf dans quelques cas particuliers, cette équation est irrésolvable.
dans le cas particulier où la matrice $\mathbf{A}$ est diagonale la solution est de la forme

$$\Phi(t, t_0) = \exp\left(\int_{t_0}^{t} \mathbf{A}(\tau)d\tau\right) \tag{1.70}$$

1.4.4 Simulation sur calculateur

Equation générale, en posant $t = kT_e$

$$\underline{x}((k+1)T_e) = e^{\mathbf{A}(k+1)T_e}\left[\int_{kT_e}^{(k+1)T_e} e^{-\mathbf{A}\tau}\mathbf{B}u(\tau)d\tau\right] + e^{\mathbf{A}T_e}\underline{x}(kT_e) \tag{1.71}$$

Hypothèse $u(t) = cte$ entre kT_e et $(k+1)T_e$,

$$\underline{x}((k+1)T_e) = e^{\mathbf{A}(k+1)T_e}\left[e^{-\mathbf{A}(k+1)T_e} - e^{-\mathbf{A}kT_e}\right]\mathbf{A}^{-1}\mathbf{B}u(kT_e) + e^{\mathbf{A}T_e}\underline{x}(kT_e) \tag{1.72}$$

$$\underline{x}((k+1)T_e) = -\left[1 - e^{-\mathbf{A}T_e}\right]\mathbf{A}^{-1}\mathbf{B}u(kT_e) + e^{\mathbf{A}T_e}\underline{x}(kT_e) \tag{1.73}$$

$$\underline{x}((k+1)T_e) = \left(T_e + \frac{{T_e}^2}{2}\mathbf{A} + \frac{{T_e}^3}{3!}(\mathbf{A})^2 + \cdots\right)\mathbf{B}u(kT_e) + e^{\mathbf{A}T_e}\underline{x}(kT_e) \tag{1.74}$$

simplification : Méthode d'Euler

$$\underline{x}((k+1)T_e) = -\left[1 - e^{-\mathbf{A}T_e}\right]\mathbf{A}^{-1}\mathbf{B}u(kT_e) + e^{\mathbf{A}T_e}\underline{x}(kT_e) \tag{1.75}$$

avec un développement au premier degré,

$$\underline{x}((k+1)T_e) = T_e\mathbf{B}u(kT_e) + (1 + \mathbf{A}T_e)\underline{x}(kT_e) \tag{1.76}$$

Vérification dans le cas simple

$$\frac{dx}{dt} = \frac{x((k+1)T_e) - x(kT_e)}{T_e} \tag{1.77}$$

et

$$\frac{dx}{dt} = ax + bu \tag{1.78}$$

$$x((k+1)T_e) - x(kT_e) = T_e a x(kT_e) + T_e b u(kT_e) \tag{1.79}$$

$$x((k+1)T_e) = (T_e a + 1)x(kT_e) + T_e b u(kT_e) \tag{1.80}$$

Vous êtes maintenant capables simuler sur un ordinateur le comportement d'un système donné sous la forme d'une représentation d'état. Notez que, si le système est invariant dans le temps, le choix de T_e n'est pas crucial. Les termes $-\left[1 - e^{-\mathbf{A}T_e}\right]\mathbf{A}^{-1}\mathbf{B}$ et$e^{\mathbf{A}T_e}$ de l'équation (1.75) peuvent être précalculés une fois pour toutes. Dans le cas contraire il vaut mieux choisir T_e suffisamment petit pour que l'équation (1.76) soit valable.

Chapitre 2

Obtention des équations d'état

2.1 Méthode directe

La méthode qui donne le maximum d'informations est celle qui consiste à déterminer la représentation d'état directement à partir des équations de la physique appliquées au système considéré.

2.2 A partir de la fonction de transfert

Souvent, les systèmes physiques sont donnés sous la forme d'une fonction de transfert, les paragraphes suivants traitent du passage de la fonction de transfert vers la représentation d'état.
Tout système monovariable peut être représenté sous la forme

$$\frac{Y(p)}{U(p)} = \frac{b_0 + b_1 p + b_2 p^2 + \cdots + b_m p^m}{a_0 + a_1 p + a_2 p^2 + \cdots + a_n p^n} \tag{2.1}$$

En général $m < n$, car les systèmes physiques sont filtrants, quelquefois , $m = n$.

2.2.1 Forme 1 : forme canonique de commandabilité

L'équation (2.1) peut se décomposer sous la forme suivante

$$\frac{Y(p)}{W(p)} \times \frac{W(p)}{U(p)} \quad \text{avec} \quad \frac{W(p)}{U(p)} = \frac{1}{a_0 + a_1 p + a_2 p^2 + \cdots + a_n p^n} \tag{2.2}$$

sous cette forme, nous pouvons en déduire que

$$U(p) = a_0 W(p) + a_1 p W(p) + a_2 p^2 W(p) + \cdots + a_n p^n W(p) \tag{2.3}$$

donc,

$$U(p) = a_0 W(p) + a_1 \dot{W}(p) + a_2 \ddot{W}(p) + \cdots + a_n \overset{\cdot n}{W}(p) \tag{2.4}$$

d'où le choix du vecteur d'état

$$\underline{x} = \begin{bmatrix} W \\ \dot{W} \\ \vdots \\ \overset{\cdot (n-1)}{W} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2.5}$$

on en déduit

$$\begin{gathered} \dot{x_1} = x_2 \\ \dot{x_2} = x_3 \\ \vdots \\ \dot{x_n} = \frac{1}{a_n} U(p) - \frac{a_0}{a_n} x_1 - \frac{a_1}{a_n} x_2 + \cdots - \frac{a_{n-1}}{a_n} x_n \end{gathered} \tag{2.6}$$

$$y = b_0x_1 + b_1x_2 + b_2x_3 + \cdots + b_{n-1}x_n + b_n\left(\frac{1}{a_n}U(p) - \frac{a_0}{a_n}x_1 - \frac{a_1}{a_n}x_2 + \cdots - \frac{a_{n-1}}{a_n}x_n\right) \tag{2.7}$$

la représentation d'état est alors, dans le cas $n = m$

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \vdots \\ \dot{x_{n-1}} \\ \dot{x_n} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & -\frac{a_2}{a_n} & \dots & -\frac{a_{n-1}}{a_n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{a_n} \end{bmatrix} u \tag{2.8}$$

$$y = \left[b_0 - \frac{a_0}{a_n}b_n, \cdots, b_{n-1} - \frac{a_{n-1}}{a_n}b_n\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \frac{b_n}{a_n}u \tag{2.9}$$

Dans le cas nettement plus fréquent où $m < n$ et après simplification par a_n, le système est plus simple

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \vdots \\ \dot{x_{n-1}} \\ \dot{x_n} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u \tag{2.10}$$

$$y = [b_0, b_1, \cdots, b_m, 0, \cdots, 0] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} \tag{2.11}$$

2.2.2 Forme 2 : forme canonique d'observabilité

Dans le cas où $m < n$ et après simplification par a_n,

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \vdots \\ \dot{x_{n-1}} \\ \dot{x_n} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & & \ddots & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} u \tag{2.12}$$

$$y = [0, \cdots, 0, 1] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} \tag{2.13}$$

Noter la symétrie des deux formes canoniques, transposition par rapport à la diagonale principale.

2.2.3 Représentation modale

La représentation modale consiste à choisir le vecteur d'état tel que la matrice **A** soit diagonale et fasse apparaître les valeurs propres du système.
Valeurs propres d'une matrice,

$$\det\left[\lambda\mathbf{I}-\mathbf{A}\right]=0 \tag{2.14}$$

S'il y a n valeurs propres distinctes, il y a aussi n vecteurs propres $\underline{v_i}$ tels que

$$\left[\lambda_i\mathbf{I}-\mathbf{A}\right]\underline{v_i}=0 \tag{2.15}$$

d'où l'on en déduit une matrice de passage

$$\mathbf{P}_{(n,n)}=\left[\underline{v_1},\underline{v_2},\cdots,\underline{v_n}\right] \tag{2.16}$$

le nouveau vecteur d'état est x' tel que

$$x=\mathbf{P}x' \tag{2.17}$$

La nouvelle matrice d'état est donc

$$\mathbf{A}'=\mathbf{P}^{-1}\mathbf{A}\mathbf{P}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n\end{bmatrix} \tag{2.18}$$

2.2.4 Forme canonique de Jordan (forme diagonale)

Méthode : décomposition en éléments simples de la fraction rationnelle $G(p)$

Cas où tous les pôles sont simples

$$Y(p)=\left(\sum_{i=1}^{n}\frac{r_i}{p-\lambda_i}+r_0\right)U(p)$$

où $r_0\neq 0$ si $m=n$
Choix des variables d'état :

$$X_i(p)=\frac{1}{p-\lambda_i}U(p),\qquad i=1,2,\cdots,n$$

donc

$$Y(p)=\sum_{i=1}^{n}r_iX_i(p)+r_0U(p)$$

En prenant la transformée de Laplace inverse

$$\begin{array}{rcl}\dot{x_i} & = & \lambda_i x_i+u,\qquad i=1,2,\cdots,n\\ y & = & \sum_{i=1}^{n}r_ix_i+r_0u\end{array}$$

Ecriture matricielle

$$\begin{pmatrix}\dot{x_1}\\ \dot{x_2}\\ \vdots\\ \dot{x_{n-1}}\\ \dot{x_n}\end{pmatrix}=\begin{pmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & 0 & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_{n-1}\\ x_n\end{pmatrix}+\begin{pmatrix}1\\ 1\\ \vdots\\ 1\\ 1\end{pmatrix}u \tag{2.19}$$

$$y = (r_1, r_2, \cdots, r_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} + r_0 u \tag{2.20}$$

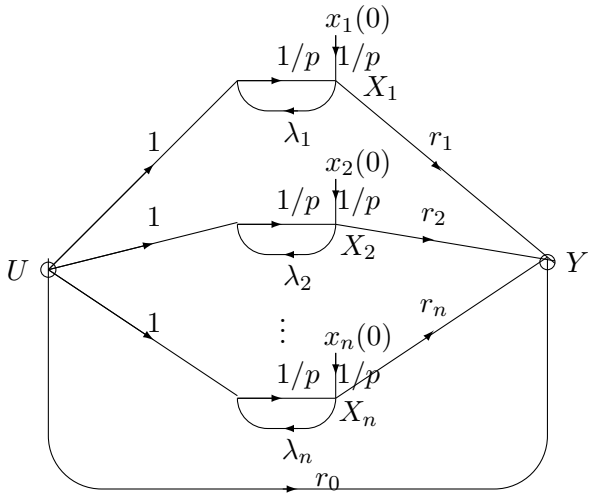


FIGURE 2.1 – Graphe de fluence de la représentation d'état sous la forme canonique de Jordan

Cas d'un pôle multiple

soit λ_1 d'ordre q donc

$$Y(p) = \left[\frac{r_1}{p-\lambda_1} + \cdots + \frac{r_q}{(p-\lambda_1)^q} + \sum_{i=q+1}^{n} \frac{r_i}{p-\lambda_i} + r_0 \right] U(p)$$

Choix des variables d'état

$$\begin{aligned} X_1 &= \frac{1}{p-\lambda_1}U, \quad X_2 = \frac{1}{(p-\lambda_1)^2}U, \cdots, X_q = \frac{1}{(p-\lambda_1)^q}U \\ X_i &= \frac{1}{p-\lambda_i}U, \quad i = q+1, \cdots, n \end{aligned}$$

d'où

$$X_2 = \frac{1}{p-\lambda_1}X_1, \quad X_3 = \frac{1}{p-\lambda_1}X_2, \cdots, X_q = \frac{1}{p-\lambda_1}X_{q-1}$$

En prenant la transformée de Laplace inverse

$$\begin{aligned} \dot{x}_1 &= \lambda_1 x_1 + u \\ \dot{x}_2 &= \lambda_1 x_2 + x_1 \\ \vdots &= \quad \vdots \\ \dot{x}_q &= \lambda_1 x_q + x_{q-1} \\ \dot{x}_i &= \lambda_i x_i + u, \qquad i = q+1, \cdots, n \\ y &= \textstyle\sum_{i=1}^{n} r_i x_i + r_0 u \end{aligned}$$

Ecriture matricielle

$$\begin{pmatrix} \dot{x_1} \\ \dot{x_2} \\ \vdots \\ \dot{x_q} \\ \dot{x_{q+1}} \\ \vdots \\ \dot{x_n} \end{pmatrix} = \begin{pmatrix} \lambda_1 & & & & & & \\ 1 & \lambda_1 & & & & & \\ & \ddots & \ddots & & & & \\ & & 1 & \lambda_1 & & & \\ & & & & \lambda_{q+1} & & \\ & & & & & \ddots & \\ & & & & & & \lambda_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \\ x_{q+1} \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} u \qquad (2.21)$$

$$y = (r_1, r_2, \cdots, r_q, r_{q+1}, \cdots, r_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \\ x_{q+1} \\ \vdots \\ x_n \end{pmatrix} + r_0 u \qquad (2.22)$$

Matrice du système obtenue : forme réduite de Jordan
Graphe correspondant :

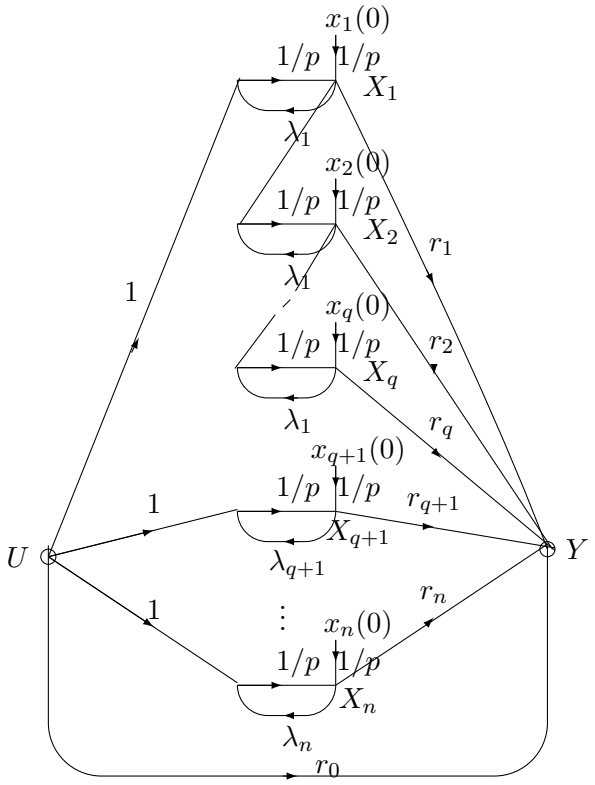


FIGURE 2.2 – Graphe de fluence de la représentation d'état sous la forme canonique de Jordan

Remarque 1 : Dans le cas de plusieurs pôles multiples, à chaque pôle correspond un sous bloc de Jordan carré $q \times q$, avec la valeur de ce pôle sur la diagonale et 1 sur le sous diagonale.
Remarque 2 : Les variables d'état ne sont plus entièrement découplées.

Cas d'une paire de pôles complexes conjugués

2 solutions :

1. Etablir une forme réduite de Jordan comme précédemment avec des variables d'état complexes.
2. Rechercher une représentation autre que diagonale, où tous les coefficients sont réels. (voir §4).

Exemple :
Equation d'état :

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} a+jb & 0 \\ 0 & a-jb \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u \tag{2.23}$$

Equation de sortie :

$$\omega = \begin{bmatrix} c & \overline{c} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2.24}$$

Dans cet exemple, les variables d'état sont des fonctions complexes conjuguées.
Le changement de variable $\underline{z} = \mathbf{M}\underline{x}$ avec $\mathbf{M} = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}$ permet de transformer le système précédent en :
Equation d'état :

$$\begin{bmatrix} \dot{z_1} \\ \dot{z_2} \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} u \tag{2.25}$$

Equation de sortie :

$$\omega = \begin{bmatrix} Re(c) & Im(c) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{2.26}$$

En résumé

$$\mathbf{A} = \begin{bmatrix} a & b & 0 & 0 & 0 \\ -b & a & 0 & 0 & 0 \\ 0 & 0 & c & 0 & 0 \\ 0 & 0 & 0 & d & 1 \\ 0 & 0 & 0 & 0 & d \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{pôles complexes conjugués} \quad a \pm jb \\ \left.\begin{matrix} \\ \end{matrix}\right\} \text{pôle réel} \quad c \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{pôles réels d'ordre 2} \quad d \end{matrix} \tag{2.27}$$

;

$$\mathbf{B} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{pôles complexes conjugués} \quad a \pm jb \\ \left.\begin{matrix} \\ \end{matrix}\right\} \text{pôle réel} \quad c \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} \text{pôles réels d'ordre 2} \quad d \end{matrix} \tag{2.28}$$

Chapitre 3

Commandabilité et observabilité des systèmes

Problème :

- Peut-on à partir des commandes $u(t)$ contrôler l'état du système c'est-à-dire $\underline{x}(t)$?
  - $\rightarrow$ problème de commandabilité.
- Peut-on à partir de l'observation de $y(t)$ remonter à l'état du système $\underline{x}(t)$?
  - $\rightarrow$ problème d'observabilité.

3.1 Définitions

- Un système est totalement commandable s'il existe $u(t)$ qui conduise en un temps fini d'un état $x(t_0)$ à un état $x(t_1)$ quels que soient $x(t_0)$ et $x(t_1)$.
- Un système est observable, si l'observation de $y(t)$ entre t_0 et t_1 permet de retrouver $x(t_0)$

Exemple

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & & \vdots \\ \vdots & & a_{3,3} & 0 \\ 0 & \cdots & 0 & a_{4,4} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & c_{1,2} & 0 & 0 \\ 0 & 0 & 0 & c_{2,4} \end{bmatrix} \tag{3.1}$$

Le schéma représentatif de ce système et donné en figure 3.1.
Condition de commandabilité : $\mathbf{Q}_S$ est de rang n avec

$$\mathbf{Q}_S = \left[\mathbf{B}, \mathbf{AB}, \mathbf{A}^2\mathbf{B}, \cdots, \mathbf{A}^{n-1}\mathbf{B}\right] \tag{3.2}$$

Application à l'exemple du moteur commandé par l'inducteur,
Equation d'état :

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b_0 \end{bmatrix} u \tag{3.3}$$

Calcul de $\mathbf{Q}_S$

$$\mathbf{B} = \begin{bmatrix} 0 \\ b_0 \end{bmatrix}, \quad \mathbf{AB} = \begin{bmatrix} b_0 \\ -b_0 a_1 \end{bmatrix} \tag{3.4}$$

d'où

$$\mathbf{Q}_S = \begin{bmatrix} 0 & b_0 \\ b_0 & -b_0 a_1 \end{bmatrix} \tag{3.5}$$

et

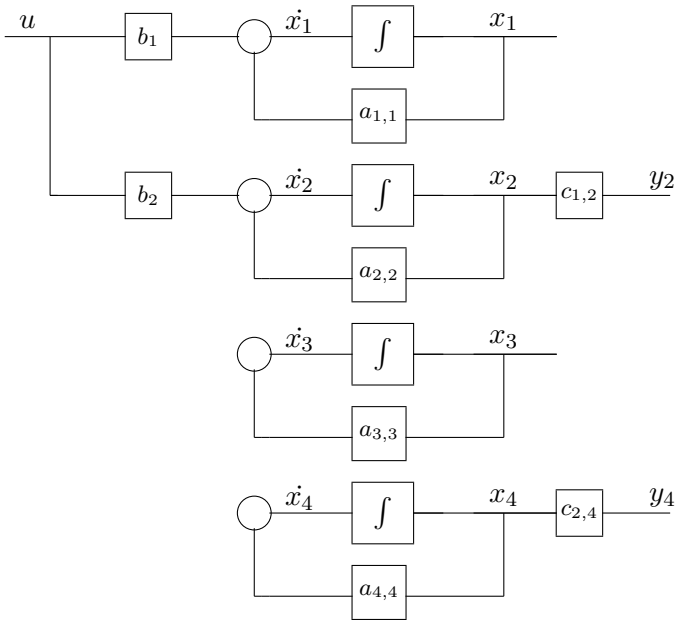


FIGURE 3.1 – Schéma d'un simulateur analogique

$$\det \mathbf{Q}_S = -{b_0}^2 \tag{3.6}$$

donc le système est commandable (si $b_0 \neq 0$).
Condition d'observabilité : $\mathbf{Q}_B$ est de rang n avec

$$\mathbf{Q}_B = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix} \tag{3.7}$$

Application à l'exemple du moteur commandé par l'inducteur,
Equation de sortie :

$$\omega = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{3.8}$$

donc

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{CA} = \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{3.9}$$

donc

$$\mathbf{Q}_B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.10}$$

et comme $\det \mathbf{Q}_B = 1$, $\mathbf{Q}_B$ est de rang n et donc le système est observable.

3.2 Que faire si un système n'est pas observable et/ou commandable

Il arrive souvent qu'après une première analyse du système celui-ci s'avère être non commandable ou non observable. Deux solutions s'offrent à vous.

3.2.1 Retour sur conception

La première solution consiste à ajouter ou changer des organes de commande (non commandable) ou des capteurs (non observable).

3.2.2 Réduction de modèles

La deuxième solution n'est valable que si la maîtrise complète de l'état n'est pas importante. Dans ce cas, la réduction du modèle est envisageable. Une façon simple de procéder et de déterminer la représentation d'état à partir de l'équation différentielle (qui "élimine" les inobservabilités) ou de la fonction de transfert du système (qui "élimine" les inobservabilités et les non commandabilités).

Chapitre 4

Transformation en l'une des formes canoniques

4.1 Diagonalisation de la matrice A

Hypothèses :

— le système est décrit par les équations suivantes :

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{4.1}$$
$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \tag{4.2}$$

— Les valeurs propres de $\mathbf{A}$ $(\lambda_1, \cdots, \lambda_n)$ sont toutes différentes.

Il existe alors une transformation $\mathbf{T}$ définie par

$$\underline{x} = \mathbf{T}\underline{x}^*$$

avec : $\underline{x}$ = ancien vecteur d'état,
$\underline{x}^*$ = nouveau vecteur d'état qui diagonalise $\mathbf{A}$

Dans ce cas $\mathbf{T}$ est la matrice modale.

On en déduit une nouvelle représentation d'état.

$$\dot{\underline{x}}^* = \mathbf{\Lambda}\underline{x}^* + \widehat{\mathbf{B}}\underline{u} \tag{4.3}$$
$$\underline{y} = \widehat{\mathbf{C}}\underline{x}^* + \widehat{\mathbf{D}}\underline{u} \tag{4.4}$$

avec :
$$\mathbf{\Lambda} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$
$$\widehat{\mathbf{B}} = \mathbf{T}^{-1}\mathbf{B}$$
$$\widehat{\mathbf{C}} = \mathbf{C}\mathbf{T}$$
$$\widehat{\mathbf{D}} = \mathbf{D}$$

• Cas où $\mathbf{A}$ est une matrice compagne (une forme canonique) avec des valeurs propres toutes différentes, alors $\mathbf{T}$ est une matrice de Vandermonde :

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ {\lambda_1}^2 & {\lambda_2}^2 & \cdots & {\lambda_n}^2 \\ \vdots & & & \vdots \\ {\lambda_1}^{n-1} & {\lambda_2}^{n-1} & \cdots & {\lambda_n}^{n-1} \end{pmatrix}$$

• Cas général.

Vecteur propre $\underline{v}_i$ associé à la valeur propre λ_i, obtenu par

$$\mathbf{A}\underline{v}_i = \lambda_i \underline{v}_i$$

c'est-à-dire

$$(\lambda_i \mathbf{I} - \mathbf{A})\underline{v}_i = 0.$$

Rappel : les λ_i sont obtenus par la résolution de l'équation caractéristique de la matrice $\mathbf{A}$

$$\det(\lambda_i \mathbf{I} - \mathbf{A}) = 0$$

la matrice de transformation $\mathbf{T}$ est alors donnée par

$$\mathbf{T} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & & & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{pmatrix} = (\underline{v}_1, \underline{v}_2, \cdots, \underline{v}_n)$$

4.2 Conséquences pour la commandabilité et l'observabilité

Propriété 1 : Si le système $\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u}$ possède des valeurs propres toutes différentes, il est commandable si et seulement si aucun élément du vecteur colonne $\widehat{\mathbf{B}} = \mathbf{T}^{-1}\mathbf{B}$ n'est nul.
démonstration : voir graphe du système découplé (fig. (2.2) page 22).
note : commandabilité $\longleftrightarrow$ $u(t)$ doit pouvoir influencer tout x_i.

Propriété 2 : Si le système

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{4.5}$$
$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \tag{4.6}$$

possède des valeurs propres toutes différentes, il est observable si et seulement si aucun élément du vecteur ligne $\widehat{\mathbf{C}} = \mathbf{CT}$ n'est nul.
démonstration : voir graphe du système découplé (fig. (2.2) page 22).
note : observabilité $\longleftrightarrow$ tout x_i doit pouvoir influencer $y(t)$.

Relation avec la fonction de transfert

$$\dot{x}_i^* = \lambda_i x_i^* + \widehat{b}_i u \tag{4.7}$$
$$X_i^* = \frac{1}{p - \lambda_i}\widehat{b}_i u \tag{4.8}$$
$$Y = \sum_{i=1}^{n} \widehat{c}_i X_i^* + dU = \underbrace{\left(\sum_{i=1}^{n} \frac{\widehat{b}_i \widehat{c}_i}{p - \lambda_i} + d\right)}_{G(p)} U \tag{4.9}$$

Si le système n'est pas commandable ou pas observable, au moins l'un des $\widehat{b}_i$ ou des $\widehat{c}_i$ sera nul, donc le terme correspondant du développement de $G(p)$ en éléments simples disparaîtra, donc la valeur propre λ_i correspondante ne sera pas un pôle de $G(p)$.

Propriété 3 : un système à 1 entrée et 1 sortie et à valeurs propres toutes différentes est à la fois commandable et observable

- si et seulement si toutes les valeurs propres sont des pôles de la fonction de transfert.
- si et seulement si la fonction de transfert $G(p)$, dans laquelle ne doit plus apparaître au numérateur et au dénominateur un terme identique, a un dénominateur de degré égal au nombre n des équations différentielles d'état.

4.3 Cas des valeurs propres complexes

S'il existe des valeurs propres complexes

$$\lambda_i = \sigma + j\omega \quad \text{et} \tag{4.10}$$
$$\lambda_{i+1} = \sigma - j\omega = \overline{\lambda_i} \tag{4.11}$$

parmi les valeurs propres de $\mathbf{A}$, il existe deux solutions possibles.

4.3.1 Diagonalisation classique

Matrice diagonale, à éléments diagonaux λ_i et λ_{i+1} complexes.
exemple : système du deuxième ordre.

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma + j\omega & 0 \\ 0 & \sigma - j\omega \end{pmatrix}$$

On montre aisément que les vecteur propres associés sont eux aussi complexes conjugués.

$$\underline{v}_i = \underline{\alpha} + j\underline{\beta} \tag{4.12}$$
$$\underline{v}_{i+1} = \underline{\alpha} - j\underline{\beta} \tag{4.13}$$
$$\tag{4.14}$$

4.3.2 Transformation modifiée : $\mathbf{T}_m$

$$\underline{x} = \mathbf{T}_m \underline{w}$$

avec $\underline{w}$ vecteur d'état qui "diagonalise" $\mathbf{A}$ et $\mathbf{T}_m = (\underline{\alpha} \;\; \underline{\beta})$
Comme

$$\begin{aligned} \mathbf{A}\underline{v}_1 &= \lambda_1 \underline{v}_1 \\ &= (\sigma + j\omega)\underline{v}_1 \\ &= (\sigma + j\omega)(\underline{\alpha} + j\underline{\beta}) \\ &= (\sigma\underline{\alpha} - \omega\underline{\beta}) + j(\omega\underline{\alpha} + \sigma\underline{\beta}) \end{aligned}$$

et que : $\mathbf{A}\underline{v}_i = \mathbf{A}\underline{\alpha} + j\mathbf{A}\underline{\beta}$
on a :

$$\begin{aligned} \mathbf{A}\underline{\alpha} &= \sigma\underline{\alpha} - \omega\underline{\beta} \\ \mathbf{A}\underline{\beta} &= \omega\underline{\alpha} + \sigma\underline{\beta} \end{aligned}$$

donc

$$\mathbf{A} \;\; (\underline{\alpha} \;\; \underline{\beta}) = (\underline{\alpha} \;\; \underline{\beta}) \begin{pmatrix} \sigma & \omega \\ -\omega & \sigma \end{pmatrix} \tag{4.15}$$
$$\mathbf{A}\mathbf{T}_m = \mathbf{T}_m \mathbf{\Lambda}_m \tag{4.16}$$

d'où

$$\mathbf{\Lambda}_m = \mathbf{T}_m^{-1}\mathbf{A}\mathbf{T}_m = \begin{pmatrix} \sigma & \omega \\ -\omega & \sigma \end{pmatrix}$$

Généralisation : système d'ordre n, ayant (à titre d'exemple) deux paires de valeurs propres complexes conjuguées $\lambda_1, \lambda_2, \overline{\lambda_1}$ et $\overline{\lambda_2}$
la transformation modifiée est :

$$\underline{x} = \mathbf{T}_m \underline{w}$$

avec

$$\mathbf{T}_m = [\mathrm{Re}(\underline{v}_1), \mathrm{Im}(\underline{v}_1), \mathrm{Re}(\underline{v}_2), \mathrm{Im}(\underline{v}_2), \underline{v}_5, \cdots, \underline{v}_n]$$

la nouvelle représentation est :

$$\underline{\dot{w}} = \mathbf{\Lambda}_m \underline{x} + \mathbf{B}_m \underline{u} \tag{4.17}$$
$$\underline{y} = \mathbf{C}_m \underline{x} + \mathbf{D}_m \underline{u} \tag{4.18}$$

avec :

$$\mathbf{\Lambda}_m = \mathbf{T}_m^{-1}\mathbf{A}\mathbf{T}_m = \begin{pmatrix} \sigma_1 & \omega_1 & & & & & \\ -\omega_1 & \sigma_1 & & & & & \\ & & \sigma_2 & \omega_2 & & & \\ & & -\omega_2 & \sigma_2 & & & \\ & & & & \lambda_5 & & \\ & & & & & \ddots & \\ & & & & & & \lambda_n \end{pmatrix}$$

et

$$\mathbf{B}_m = \mathbf{T}^{-1}\mathbf{B}, \quad \mathbf{C}_m = \mathbf{C}\mathbf{T}, \quad \mathbf{D}_m = \mathbf{D}$$

4.4 Transformation en la forme canonique d'asservissement

Hypothèses : $a_n = 1$ et $b_n = 0$.
On montre aisément que l'équation caractéristique

$$\det(p\mathbf{I} - \mathbf{A}) = 0$$

se réduit à :

$$a_0 + a_1 p + a_2 p^2 + \cdots + a_{n-1} p^{n-1} + p^n = 0$$

Dans la forme canonique d'asservissement, la dernière ligne de la matrice du système est composée des coefficients de l'équation caractéristique (excepté celui du terme de plus haut degré) changés de signe.
Théorème : Si le système est commandable on peut le mettre sous la forme canonique d'asservissement au moyen de la transformation

$$\underline{z} = \mathbf{T}^{-1}\underline{x}$$

avec

$$\mathbf{T}^{-1} = \begin{pmatrix} \underline{q}_S^T & \\ \underline{q}_S^T & \mathbf{A} \\ \underline{q}_S^T & \mathbf{A}^2 \\ \vdots & \\ \underline{q}_S^T & \mathbf{A}^{n-1} \end{pmatrix}$$

où $\underline{q}_s^T$ est la dernière ligne de l'inverse de la matrice de commandabilité $\mathbf{Q_S}^{-1}$
L'utilisation de ce théorème peut sembler lourde, mais dans la pratique il se révèle puissant, puisqu' une grande partie des termes, notamment de $\mathbf{Q_S}^{-1}$, est inutile.

4.5 Transformation en la forme canonique d'observabilité

Dans la forme canonique d'observation, la dernière colonne de la matrice du système est composée des coefficients de l'équation caractéristique (excepté celui du terme de plus haut degré) changés de signe.
Théorème : Si le système est observable on peut le mettre sous la forme canonique d'asservissement au moyen de la transformation

$$\underline{z} = \mathbf{T}\underline{x}$$

avec

$$\mathbf{T} = \begin{pmatrix} \underline{q}_B & \mathbf{A}\underline{q}_B & \mathbf{A}^2\underline{q}_B & \cdots & \mathbf{A}^{n-1}\underline{q}_B \end{pmatrix}$$

où $\underline{q}_b$ est la dernière colonne de l'inverse de la matrice d'observabilité $\mathbf{Q}_B^{-1}$

Chapitre 5

Stabilité des systèmes dynamiques linéaires

La stabilité des systèmes linéaires, dans le cas SISO, consiste le plus souvent à revenir à des méthodes classiques des systèmes représentés sous la forme d'une fonction de transfert. Attention, la fonction de transfert d'un système présentant moins d'information que sa représentation d'état, il est possible mais rare que la fonction de transfert soit stable alors que la représentation d'état présente un mode instable !

5.1 Définition

Un système est dit asymptotiquement stable, si et seulement si, écarté de sa position d'origine et soumis à une entrée nulle, celui-ci revient à sa position d'origine.

5.2 Etude de la stabilité

En reprenant l'équation 1.63 et en supposant que $u(t) = 0$ et $t_0 = 0$, nous obtenons :

$$\underline{x}(t) = e^{\mathbf{A}t}\underline{x}(0)$$

D'après la définition de la stabilité, il faut que $\underline{x}(t) \to 0$ donc que $e^{\mathbf{A}t} \to 0$.
Si $\mathbf{A}$ est diagonalisable ($\mathbf{\Lambda}$) alors il existe une matrice $\mathbf{T}$ telle que :

$$\mathbf{\Lambda} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \tag{5.1}$$

donc

$$\underline{x}(t) = e^{\mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}t}\underline{x}(0) \to 0$$

donc

$$\exp \begin{bmatrix} \lambda_1 t & 0 & \cdots & 0 \\ 0 & \lambda_2 t & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n t \end{bmatrix} \to 0 \tag{5.2}$$

donc il faut et il suffit que les λ_i soient à partie réelle négative, donc que les valeurs propres de la matrice $\mathbf{A}$ soient à partie réelle négative.

En d'autres termes, pour qu'un système soit asymptotiquement stable en boucle ouverte, il faut et il suffit que tous ses modes soient stables. L'extension aux systèmes non diagonalisables (blocs de Jordan) est immédiate puisque les termes de l'exponentielle d'une matrice triangulaire sont une combinaison linéaire des $e^{\lambda_i t}$.

Forme modale

La lecture de la stabilité est immédiate, puisque les valeurs propres de la matrice sont sur la diagonale.

Forme canonique de commandabilité

La matrice $\mathbf{A}$ se présente alors sous la forme :

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -\frac{a_0}{a_n} & -\frac{a_1}{a_n} & -\frac{a_2}{a_n} & \dots & -\frac{a_{n-1}}{a_n} \end{bmatrix}$$

le calcul des valeurs propres de la matrice revient à calculer les zéros du polynôme

$$P(\lambda) = -\frac{a_0}{a_n} - \frac{a_1}{a_n}\lambda - \frac{a_2}{a_n}\lambda^2 - \cdots - \frac{a_{n-1}}{a_n}\lambda^{n-1}$$

En fait, seul le signe des parties réelles des valeurs propres nous intéresse pour l'étude de la stabilité, donc il suffit d'appliquer le critère de Routh sur le polynôme P.

Forme canonique d'observabilité

Par transposition de la matrice $\mathbf{A}$, on revient au cas précédent.

5.3 Stabilité au sens de Lyapounov

L'origine est un état stable si pour tout $\varepsilon > 0$, il existe un nombre $\delta(\varepsilon, t_0) > 0$ tel que $\|\underline{x}(t_0)\| < \delta$ entraîne $\|\underline{x}(t)\| < \varepsilon, \quad \forall t > t_0$. Cet état est asymptotiquement stable si, de plus, il existe un nombre $\delta_a(t_0)$ tel que $\|\underline{x}(t_0)\| < \delta_a(t_0)$ entraîne $\lim_{t\to\infty} \|\underline{x}(t)\| = 0$.

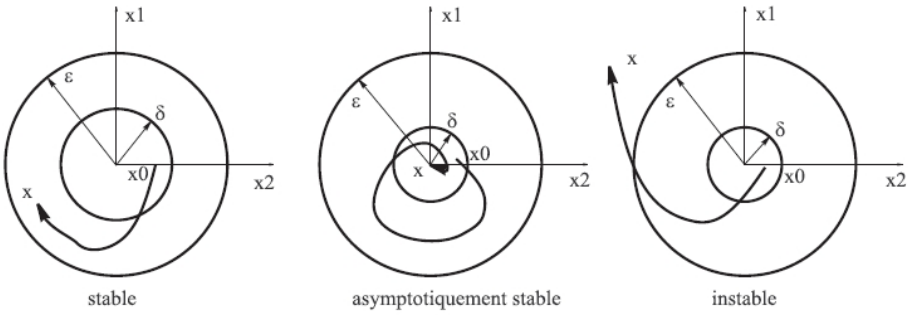


FIGURE 5.1 – Illustration de la stabilité dans le plan de phase

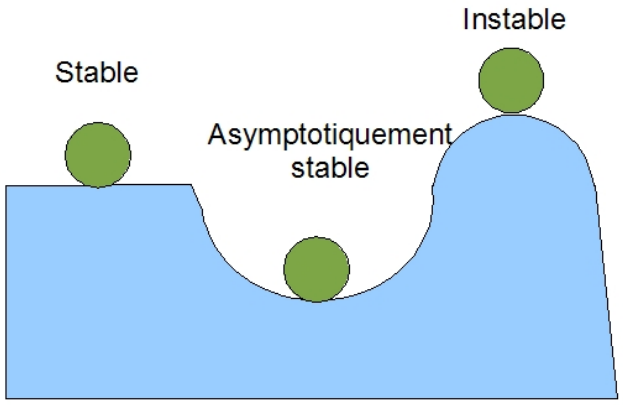


FIGURE 5.2 – Illustration de la stabilité d'une bille sur un profil

5.3.1 Théorème

S'il existe une fonction $V(\underline{x})$ telle que :

1. $V(\underline{x}) > 0, \quad \forall \underline{x} \neq 0$,
2. $V(0) = 0$,
3. $V(\underline{x}) \to \infty$ pour $\|\underline{x}\| \to \infty$,

alors l'équilibre en $\underline{x} = 0$ est localement asymptotiquement stable.
Dans le cas des systèmes linéaires invariants dans le temps, la stabilité est globale.

5.3.2 Interprétation physique

Comme nous l'avons dit précédemment, les variables d'état représentent les réservoirs d'énergie du système ou du moins une combinaison linéaire de ceux-ci. Si, à l'origine des temps le système contient de l'énergie alors $\underline{x}(0) \neq 0$. La stabilité est alors synonyme de décroissance de la quantité d'énergie présente dans le système, celle-ci peut converger vers une valeur non nulle. La stabilité asymptotique par contre implique la convergence vers 0.

Supposons que la quantité d'énergie présente dans le système soit $V(t)$, il suffit alors de démontrer que $\dot{V}(t) < 0$ pour que le système soit asymptotiquement stable ($\dot{V}(t) \leq 0$ pour que le système soit stable).

5.3.3 Applications aux systèmes linéaires

Posons

$$V(\underline{x}) = \underline{x}^T \mathbf{P}(t) \underline{x}$$

où $\mathbf{P}$ est une matrice symétrique définie positive, sa dérivée par rapport au temps s'écrit :

$$\dot{V}(\underline{x}) = \dot{\underline{x}}^T \mathbf{P} \underline{x} + \underline{x}^T \mathbf{P} \dot{\underline{x}} + \underline{x}^T \dot{\mathbf{P}} \underline{x} = \underline{x}^T (\mathbf{A}^T \mathbf{P} + \mathbf{P}\mathbf{A} + \dot{\mathbf{P}}) \underline{x}$$

Sa dérivée est toujours négative s'il existe une matrice définie positive $\mathbf{Q}$ telle que :

$$-\mathbf{Q} = \mathbf{A}^T \mathbf{P} + \mathbf{P}\mathbf{A} + \dot{\mathbf{P}}$$

Dans le cas des systèmes linéaires invariants l'équation précédente se simplifie en

$$-\mathbf{Q} = \mathbf{A}^T \mathbf{P} + \mathbf{P}\mathbf{A} \tag{5.3}$$

Dans la pratique la méthode est la suivante :

1. Choisir une matrice $\mathbf{Q}$ arbitraire, symétrique, définie positive (la matrice $\mathbf{I}$ conviendra le plus souvent),
2. déterminer ensuite $\mathbf{P}$ à partir de (5.3), cela revient à résoudre n(n+1)/2 équations,
3. le système est asymptotiquement stable si $\mathbf{P}$ est définie positive.

Cette méthode peut sembler lourde au regard de celle présentées précédemment, mais c'est la seule qui reste valable dans le cas des systèmes non linéaires et/ou variants dans le temps. La recherche de la fonction de Lyapounov $V(\underline{x}, t)$ devient alors la clef du problème.

5.3.4 Fil rouge

Le système est défini par sa matrice

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix}$$

posons

$$\mathbf{P} = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix} \quad \text{et} \quad \mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

l'application de la relation (5.3) donne

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} -2\, a_0\, p_2 & -a_0\, p_3 + p_1 - a_1\, p_2 \\ -a_0\, p_3 + p_1 - a_1\, p_2 & 2\, p_2 - 2\, a_1\, p_3 \end{bmatrix}$$

la résolution de 3 équations à 3 inconnues donne la matrice suivante

$$\mathbf{P} = \begin{bmatrix} \frac{{a_0}^2 + a_0 + {a_1}^2}{2 a_0\, a_1} & \frac{1}{2a_0} \\ \frac{1}{2a_0} & \frac{a_0 + 1}{2 a_0\, a_1} \end{bmatrix}$$

sachant que a_0 et a_1 sont positifs, la matrice $\mathbf{P}$ est bien définie positive, le système est stable.

Chapitre 6

Commande des systèmes

6.1 Placement de pôles

Hypothèses : le système est décrit par les équations suivantes

$$\begin{aligned} \underline{\dot{x}} &= \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \\ \underline{y} &= \mathbf{C}\underline{x} \end{aligned}$$

L'objectif de la synthèse est une régulation (donc pas un asservissement), donc il s'agit de maintenir y proche d'une valeur de consigne y_c. Le schéma de la régulation est donné figure 6.1.

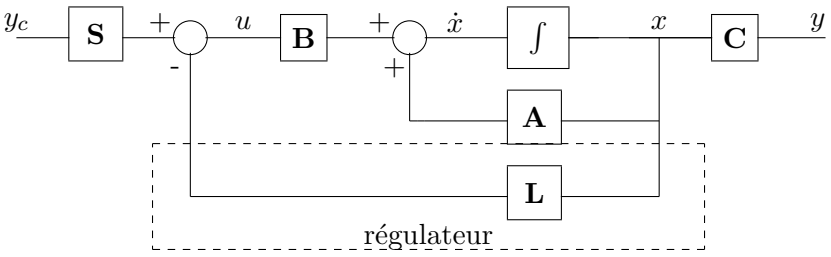


FIGURE 6.1 – Régulation continue par retour d'état

6.1.1 Calcul du régulateur L

Supposons que la fonction de transfert du système en boucle ouverte soit :

$$F_{BO}(p) = \frac{Y(p)}{U(p)} = \frac{b_0 + b_1 p + b_2 p^2 + \cdots + b_m p^m}{a_0 + a_1 p + a_2 p^2 + \cdots + a_n p^n},$$

avec $m < n$
alors la représentation d'état sous la forme canonique de commandabilité est (après simplification par a_n) :

$$\underline{\dot{x}} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u \tag{6.1}$$

$$y = [b_0, b_1, \cdots, b_m, 0, \cdots, 0]\, \underline{x} \tag{6.2}$$

On remarque que la dernière ligne de la matrice **A** contient les coefficients du dénominateur de la fonction de transfert.
Le schéma 6.1 donne comme commande :

$$u = -\mathbf{L}\underline{x} + \mathbf{S}\underline{y}_c = -(l_0 \;\; l_1 \;\; l_2 \;\; \cdots \;\; l_{n-1})\underline{x} + \mathbf{S}\underline{y}_c.$$

Avec cette commande, le système en boucle fermée admet pour représentation d'état :

$$\dot{\underline{x}} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -l_0 - a_0 & -l_1 - a_1 & -l_2 - a_2 & \cdots & -l_{n-1} - a_{n-1} \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \mathbf{S}\underline{y}_c \tag{6.3}$$

$$y = [b_0, b_1, \cdots, b_m, 0, \cdots, 0]\,\underline{x}. \tag{6.4}$$

Ainsi la fonction de transfert du système en boucle fermée s'écrit :

$$F_{BF}(p) = \frac{Y(p)}{\mathbf{S}y_c(p)} = \frac{b_0 + b_1 p + b_2 p^2 + \cdots + b_m p^m}{(l_0 + a_0) + (l_1 + a_1)p + (l_2 + a_2)p^2 + \cdots + p^n},$$

donc, le régulateur permet d'imposer arbitrairement le polynôme caractéristique (donc la dynamique) de la fonction de transfert du système en boucle fermée.
Supposons que le polynôme choisi soit :

$$P(p) = \alpha_0 + \alpha_1 p + \alpha_2 p^2 + \cdots + p^n,$$

Le calcul du régulateur est immédiat puisque :

$$(l_i + a_i) = \alpha_i,$$

donc

$$l_i = \alpha_i - a_i.$$

Si le système n'est pas directement sous la forme de commandabilité, il suffit de s'y ramener (voir § 4.4). Le correcteur obtenu précédemment devra subir la transformation inverse à celle utilisée pour obtenir la forme de commandabilité soit :

$$\mathbf{L}^* = \mathbf{L}\mathbf{T}^{-1}$$

6.1.2 Calcul de la matrice de préfiltre S

Le modèle du système en boucle fermée est :

$$\begin{aligned} \dot{\underline{x}} &= (\mathbf{A} - \mathbf{B}\mathbf{L})\underline{x} + \mathbf{B}\mathbf{S}\underline{y}_c \\ \underline{y} &= \mathbf{C}\underline{x} \end{aligned}$$

Si le système est stable, alors $\lim_{t\to\infty} \dot{\underline{x}} = 0$ donc,

$$\lim_{t\to\infty} \underline{y}(t) = -\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{L})^{-1}\mathbf{B}\mathbf{S}\underline{y}_c$$

or, on désire que $\lim_{t\to\infty} \underline{y}(t) = \underline{y}_c$, donc,

$$\mathbf{S}^{-1} = -\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{L})^{-1}\mathbf{B}$$

$$\mathbf{S} = -\Big(\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{L})^{-1}\mathbf{B}\Big)^{-1}$$

6.2 Cas d'une représentation quelconque du système à asservir

6.2.1 Transformation en la forme canonique de commandabilité

$$\underline{z} = \mathbf{T}^{-1}\underline{x}$$

avec

$$\mathbf{T}^{-1} = \begin{pmatrix} \underline{q}_S^T & \\ \underline{q}_S^T & \mathbf{A} \\ \underline{q}_S^T & \mathbf{A}^2 \\ \vdots & \\ \underline{q}_S^T & \mathbf{A}^{n-1} \end{pmatrix}$$

où $\underline{q}_s^T$ est la dernière ligne de l'inverse de la matrice de commandabilité $\mathbf{Q_S}^{-1}$.

Dans cette représentation

$$u = -\mathbf{L}_{FC}\underline{z} = -\mathbf{L}_{FC}\mathbf{T}^{-1}\underline{x} = -\mathbf{L}\underline{x}$$

avec

$$\mathbf{L} = \mathbf{L}_{FC}\mathbf{T}^{-1} = (\alpha_0 - a_0)\underline{q}_S^T + (\alpha_1 - a_1)\underline{q}_S^T\mathbf{A} + \cdots + (\alpha_{n-1} - a_{n-1})\underline{q}_S^T\mathbf{A}^{n-1}$$

6.2.2 Théorème de Cayley-Hamilton

Toute matrice carrée $\mathbf{A}$ satisfait sa propre équation caractéristique.

Si

$$P(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}) = \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1\lambda + \alpha_0 = 0$$

alors

$$P(\mathbf{A}) = \det(\mathbf{A}\mathbf{I} - \mathbf{A}) = \mathbf{A}^n + \alpha_{n-1}\mathbf{A}^{n-1} + \cdots + \alpha_1\mathbf{A} + \alpha_0\mathbf{I} = 0$$

En appliquant ce théorème aux résultats précédents, $\mathbf{A}$ et $\mathbf{A}_{FC}$sont semblables et ont la même équation caractéristique, donc $\mathbf{A}$ doit satisfaire l'équation caractéristique de $\mathbf{A}_{FC}$.
Donc

$$\mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I} = 0$$

$$\underline{q}_S^T\mathbf{A}^n = -a_{n-1}\underline{q}_S^T\mathbf{A}^{n-1} - \cdots - a_1\underline{q}_S^T\mathbf{A} - a_0\underline{q}_S^T$$

d'où, par substitution : Théorème d'Ackermann

$$\mathbf{L} = \underline{q}_S^T\mathbf{A}^n + \alpha_{n-1}\underline{q}_S^T\mathbf{A}^{n-1} + \cdots + \alpha_1\underline{q}_S^T\mathbf{A} + \alpha_0\underline{q}_S^T$$

$$\mathbf{L} = \underline{q}_S^T P(\mathbf{A})$$

Théorème : Pour un système à une entrée et une sortie les zéros du système restent inchangés par le placement des pôles

6.3 Commande Modale

6.3.1 Définition

Soit la transformation $\underline{x} = \mathbf{T}\underline{x}^\star$ $\mathbf{T}$ est la transformation modale qui diagonalise $\mathbf{A}$ si toutes les valeurs propres sont différentes ou du moins la met sous forme de blocs de Jordan si toutes les valeurs propres ne sont pas différentes.
Dans la nouvelle base, les équations sont

$$\dot{\underline{x}}_i^\star = \lambda_i \underline{x}_i^\star + u_i^\star$$

si le système est diagonalisable.
Si les u_{d_i} sont indépendantes les une des autres, on peut asservir chaque variable d'état $\underline{x}_{d_i}$, et donc chaque mode du système.

6.3.2 Méthode de synthèse

Si l'on dispose de p grandeurs de commande indépendantes, on peut asservir p coordonnées modales.

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \Rightarrow \dot{\underline{x}}^* = \mathbf{\Lambda}\underline{x}^* + \widehat{\mathbf{B}}\underline{u}$$

avec :

$$\begin{aligned} \mathbf{\Lambda} &= \mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \\ \widehat{\mathbf{B}} &= \mathbf{T}^{-1}\mathbf{B} \end{aligned}$$

Choix des p coordonnées ramenées en contre réaction :
- les plus proches de l'axe imaginaire (système plus rapide),
- celles qui sont le plus influencées par une perturbation (système moins sensible au perturbations)
- ...

le système peut être mis sous la forme :

$$\begin{pmatrix} \dot{\underline{x}}_p^\star \\ \dot{\underline{x}}_{n-p}^\star \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_p & 0 \\ 0 & \mathbf{\Lambda}_{n-p} \end{pmatrix} \begin{pmatrix} \underline{x}_p^\star \\ \underline{x}_{n-p}^\star \end{pmatrix} + \begin{pmatrix} \widehat{\mathbf{B}}_p \\ \widehat{\mathbf{B}}_{n-p} \end{pmatrix} \underline{u}$$

avec :
$\mathbf{\Lambda}_p$: sous matrice $(p \times p)$ de $\mathbf{\Lambda}$

$\widehat{\mathbf{B}}_p$: sous matrice $(p \times p)$ de $\widehat{\mathbf{B}}$
Principe : déterminer la contre réaction de ces p coordonnées de telle manière que :

1. les valeurs propres du système λ_1 à λ_p soient remplacées par les valeurs propres souhaitées $\lambda_1^\star$ à $\lambda_p^\star$
2. le système bouclé reste diagonal (en ces p variables, donc chaque mode est régulé indépendamment)

donc le système bouclé doit avoir, en mode de régulation pour les p coordonnées concernées une équation de la forme :

$$\dot{\underline{x}}_p^\star = \mathbf{\Lambda}_p^\star \underline{x}_p^\star$$

avec $\mathbf{\Lambda}_p^\star = \begin{pmatrix} \lambda_1^\star & & 0 \\ & \ddots & \\ 0 & & \lambda_p^\star \end{pmatrix}$

donc

$$\mathbf{\Lambda}_p \underline{x}_p^\star + \underline{u}^\star = \mathbf{\Lambda}_p^\star \underline{x}_p^\star$$

donc

$$\underline{u}^\star = -(\mathbf{\Lambda}_p - \mathbf{\Lambda}_p^\star)\underline{x}_p^\star = -\begin{pmatrix} \lambda_1 - \lambda_1^\star & & 0 \\ & \ddots & \\ 0 & & \lambda_p - \lambda_p^\star \end{pmatrix} \begin{pmatrix} \underline{x}_1^\star \\ \underline{x}_p^\star \end{pmatrix}$$

Problème : revenir à $\underline{u}$ notre véritable vecteur de commande
$\mathbf{B}$ $(n \times p)$ est de rang p ; $\mathbf{T}^{-1}$ est régulière donc : $\widehat{\mathbf{B}}(n \times p)$ est de rang p donc $\widehat{\mathbf{B}}_p(p \times p)$ est régulière

$$\underline{u} = \widehat{\mathbf{B}}_p^{-1}\underline{u}^\star = -\widehat{\mathbf{B}}_p^{-1}(\mathbf{\Lambda}_p - \mathbf{\Lambda}_p^\star)\underline{u}^\star$$

d'où

$$\dot{\underline{x}}_p^\star = \mathbf{\Lambda}_p^\star \underline{x}_p^\star \tag{6.5}$$

$$\dot{\underline{x}}_{n-p}^\star = -\widehat{\mathbf{B}}_{n-p}\widehat{\mathbf{B}}_p^{-1}(\mathbf{\Lambda}_p - \mathbf{\Lambda}_p^\star)\underline{x}_p^\star + \mathbf{\Lambda}_{n-p}\underline{x}_{n-p}^\star \tag{6.6}$$

— l'équation (6.5) montre que les p premières coordonnées modales sont bien asservies individuellement ; les valeurs propres sont bien celles choisies $(\lambda_i^\star)$,
— l'équation (6.6) montre que les $n-p$ autres coordonnées modales sont influencées par les p premières.

Vérification de cette deuxième constatation. Calculons les valeurs propres du système bouclé :

$$\det \begin{pmatrix} p\mathbf{I}_p - \mathbf{\Lambda}_p^\star & 0 \\ \widehat{\mathbf{B}}_{n-p}\widehat{\mathbf{B}}_p^{-1}(\mathbf{\Lambda}_p - \mathbf{\Lambda}_p^\star) & p\mathbf{I}_p - \mathbf{\Lambda}_{n-p} \end{pmatrix} = 0$$

$$\det\left(p\mathbf{I}_p - \mathbf{\Lambda}_p^\star\right) \times \det\left(p\mathbf{I}_p - \mathbf{\Lambda}_{n-p}\right) = 0$$

Les valeurs propres du système sont donc :
$\lambda_1^\star, \cdots, \lambda_p^\star$: les p premières qui ont été déplacées par la commande modale,
$\lambda_{p+1}, \cdots, \lambda_n$: les $n-p$ autres qui restent inchangées.
La commande modale ne présente pas d'effet indésirable du fait du couplage entre les $n-p$ coordonnées modales restantes et les p premières.

Synthèse du régulateur modal

Retour à la représentation de départ :

$$\underline{x}^\star = \mathbf{T}^{-1}\underline{x}$$

$$\underline{u} = -\widehat{\mathbf{B}}_p^{-1}(\mathbf{\Lambda}_p - \mathbf{\Lambda}_p^\star)\underline{x}^\star = -\widehat{\mathbf{B}}_p^{-1}(\mathbf{\Lambda}_p - \mathbf{\Lambda}_p^\star)\mathbf{T}^{-1}\underline{x} = -\mathbf{L}\underline{x}$$

$$\mathbf{L} = \widehat{\mathbf{B}}_p^{-1}\begin{pmatrix} \lambda_1 - \lambda_1^\star & & 0 \\ & \ddots & \\ 0 & & \lambda_p - \lambda_p^\star \end{pmatrix}\mathbf{T}^{-1}\underline{x}$$

6.4 Choix des pôles

Le choix de pôles en boucle fermée tient plus de l'art que de la science. En effet, les phénomènes à prendre en compte sont nombreux, très dépendants du système considéré et des performances désirées.

Voici quelques règles fondamentales à respecter :

1. Les pôles choisis doivent être stables,
2. pas trop proches de l'axe des imaginaires, sinon la moindre variation de modèle peut provoquer une instabilité,
3. Les pôles complexes conjugués seront choisis pour présenter un dépassement convenable (typiquement : inférieur à 20 %) sinon le régime transitoire sera long
4. pas trop rapides (typiquement : 4 à 10 fois plus rapides que les pôles en B0), il est peu probable que le modèle que vous avez soit encore valable au-delà de ce domaine et/ou que la commande ne sature pas.

La figure 6.2 résume ces quelques règles.

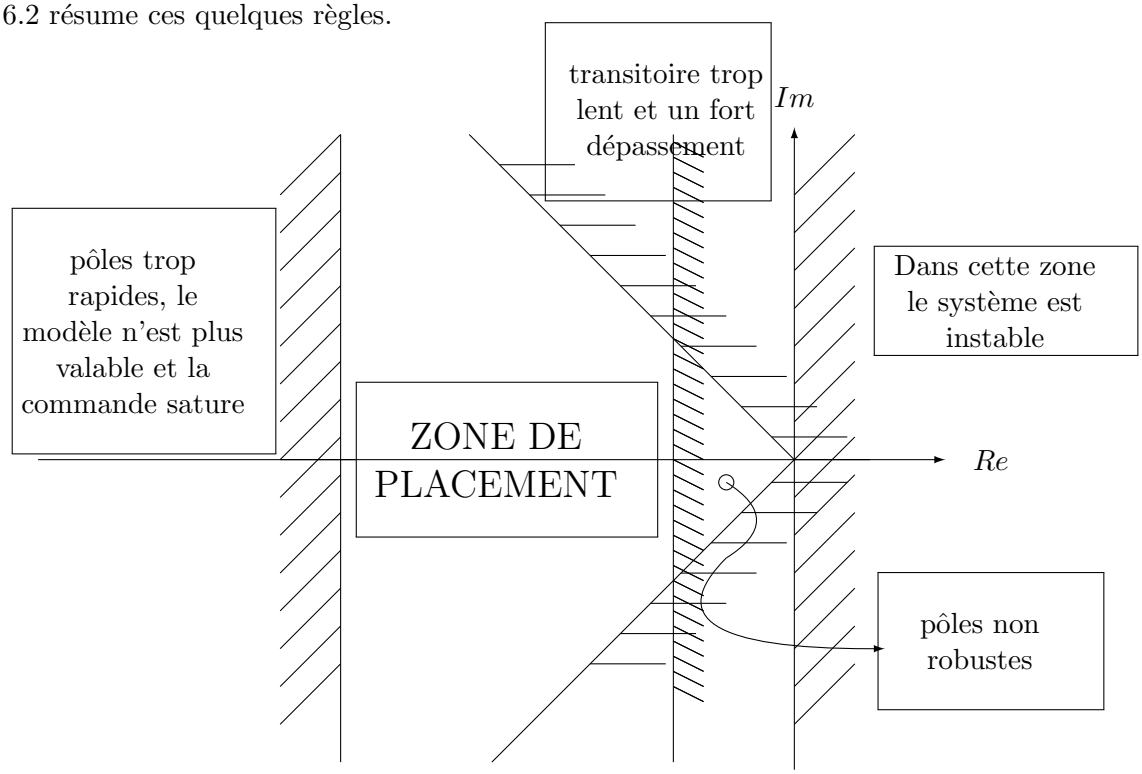
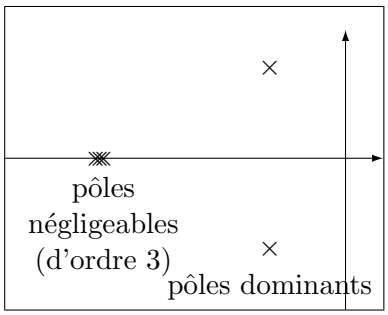


FIGURE 6.2 – Règles de placement de pôles

Le choix des pôles au sein de la zone de placement est tout autant un art et les méthodes sont nombreuses. Afin de réduire vos maux de tête lors du choix je vous propose 3 méthodes ayant fait leurs preuves, mais sachez qu'il en existe d'autres.

Les exemples sont donnés pour un système d'ordre 5.

6.4.1 Pôles complexes conjugués dominants



Le grand avantage de cette méthode est une simplification des calculs et une bonne maîtrise du comportement en boucle fermée du système (comportement comme les pôles dominants). Les résultats en boucle fermée sont quelques fois surprenants si les pôles négligés ne l'étaient pas vraiment. Bien sûr si le comportement souhaité est de type premier ordre, il suffit de choisir un pôle simple comme pôle dominant.

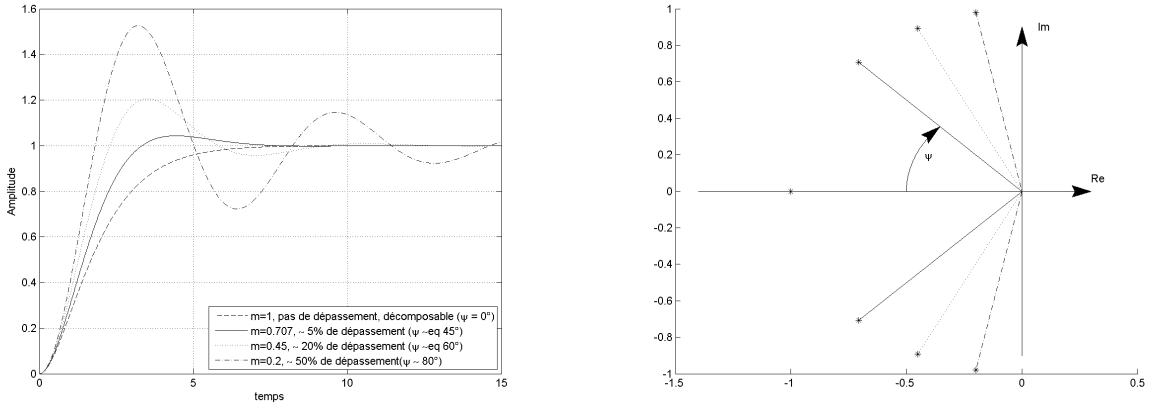


FIGURE 6.3 – Relation entre dépassement et position des pôles dans le plan complexe

Quelques relations pour la détermination des pôles :
Pour $H(p) = \frac{1}{1+\frac{2m}{\omega_0}p+\frac{p^2}{\omega_0^2}}$
Temps de réponse à 5% :

$$T_{r5\%} = \frac{3}{m\omega_0}$$

Temps de montée (10 à 90% de la valeur finale) :

$$T_m = \frac{\pi}{2\omega_0\sqrt{1-m^2}}$$

Premier dépassement

$$D1 = 100e^{-\frac{m\pi}{\sqrt{1-m^2}}}$$

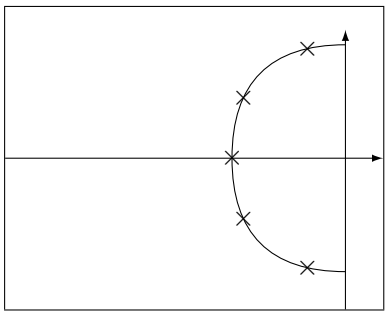
la partie réelle des pôles négligeables sera choisie égale entre 3 et 10 plus grande que la partie réelle des pôles dominants soit :

$$P_{neg} = 3 \text{ à } 10 \times m\omega_0$$

TABLE 6.1 – Polynômes normalisés de Butterworth à l'ordre n (remplacer p par p/ω_0)

n	Polynôme	$D\%$	$\omega_0 T(D\%)$	$\omega_0 T_{r5\%}$
1	$(p+1)$	0		3
2	$(p^2+1.4142p+1)$	4.3	4.5	2.8
3	$(p+1)(p^2+p+1)$	8.1	4.9	5.9
4	$(p^2+0.7654p+1)(p^2+1.8478p+1)$	10.8	5.5	6.7
5	$(p+1)(p^2+0.6180p+1)(p^2+1.6180p+1)$	12.7	6.3	7.5
6	$(p^2+0.5176p+1)(p^2+1.4142p+1)(p^2+1.9319p+1)$	14.1	6.9	10.7

6.4.2 Maximalement plat



Premier dépassement

Cette méthode qui revient à calculer un filtre de Butterworth (qui ne présente pas d'ondulation dans la bande passante d'où son nom : maximalement plat) fonctionne bien dans le cas des systèmes électromécaniques. Le dépassement n'est pas réglable et est fonction de l'ordre du système (voir tableau 6.1), seuls les temps (montée, établissement, ...) sont réglables par le choix de ω_0. Les pôles sont placés sur un demi cercle. Bien sûr, la même méthode existe avec les polynômes de Tchebyshev, dans ce cas, le dépassement est réglable indépendamment du temps d'établissement.

$$D1 = -0.20n^2 + 4n - 2.35$$

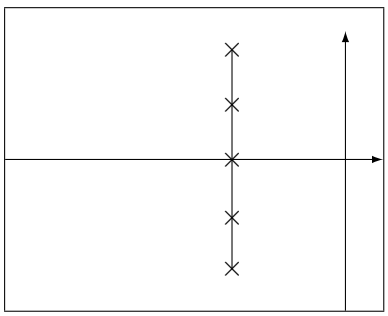
Temps du premier maximum :

$$T_{D1} = \frac{0.016n^2 + 0.52n + 3.3}{\omega_0}$$

les pôles sont donnés par :

$$\text{pour } \; 0 \leq k \geq n \;\; , \;\; p_k = \omega_0 e^{\frac{j(2k+n-1)\pi}{2n}}$$

6.4.3 Pôles à partie réelle identique



Cette méthode, proche de la méthode des pôles dominants, consiste à équirépartir l'ensemble des pôles sur la même verticale. En l'absence de zéros influents dans le système, la réponse à l'échelon est toujours apériodique. Les pôles "négligeables" atténuent le dépassement des pôles "dominants". Utile en cas de bruit, car les pôles "négligeables" (qui ne le sont plus donc !) filtrent fortement ces bruits.

6.4.4 Polynômes de Naslin

C'est encore dans les vieux pots que l'on fait les meilleures soupes. Cette méthode perdure depuis les années 60 et donne de bons résultats quel que soit le domaine d'application.

Rappels : On choisit α et ω_0 en fonction des caractéristiques voulues pour la boucle fermée.

Temps de montée (10 à 90% de la valeur finale) :

$$T_m = \frac{2.2}{\omega_0}$$

Premier dépassement

$$D1\% = 10^{4.8-2\alpha}$$

les coefficients du polynôme sont alors :

$$\text{pour} \quad 0 \leq k \leq n \quad a_k = \alpha^{\frac{-k(k-1)}{2}} \omega_0^{-k}$$

En cas de zéro dans la fonctions de transfert, cette méthode permet de "précompenser" l'influence de ce zéro. Si le numérateur de la fonction de transfert est de la forme $b_0 + b_1 p$, alors $\omega_0^{'} = \frac{b_0}{b_1}$. Il suffit alors de reprendre les formules précédentes en remplaçant α par α_e et ω_0 par ω_{0_c}.

$$\alpha_e = 1.5 + \frac{\omega_0^{'}}{4\omega_0}(\alpha - 1.5)$$

$$\omega_{0_c} = \left(\frac{1}{\omega_0} - \frac{1}{2\omega_0^{'}}\right)^{-1}$$

Si ω_{0_c} est négatif, seule la méthode "try and error" pour le choix de α et ω_0 reste valable !

6.5 Commande optimale

Cette partie n'est donnée qu'à titre de culture générale. La commande optimale est une base fondamentale de l'automatique, pour continuer dans cette voie, une autoformation est nécessaire !

6.5.1 Définition

L'objectif de cette commande est de minimiser un critère quadratique de la forme :

$$J = \int_0^\infty (\underline{x}^T \mathbf{Q} \underline{x} + \underline{u}^T \mathbf{R} \underline{u}) dt \tag{6.7}$$

avec : $\mathbf{Q}$ = matrice symétrique définie positive
$\mathbf{R}$ = matrice symétrique non négative
la commande u est alors définie par l'équation suivante :

$$u = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{K}\underline{x}$$

où $\mathbf{K}$ est une matrice symétrique, solution définie négative de l'équation de Riccati :

$$\mathbf{KA} + \mathbf{A}^T\mathbf{K} - \mathbf{KBR}^{-1}\mathbf{B}^T\mathbf{K} + \mathbf{Q} = 0$$

6.5.2 Stabilité de la commande optimale

Les commandes optimales à critère quadratique conduisent toujours à un système stable en boucle fermée. La marge de phase est toujours supérieure à 60 degrés.

6.5.3 Choix des matrice R et Q

La forme du critère (eq. (6.7)) représente typiquement un critère énergétique. Aussi la commande optimale vise-t'elle à minimiser l'énergie requise pour faire passer le système d'un état à un autre. Dans ce cas les matrices $\mathbf{R}$ et $\mathbf{Q}$ seront le plus souvent choisies diagonales, les coefficients de la diagonale ayant une signification physique (inductance, masse ...).
Plus généralement, cette méthode de synthèse s'applique lorsque l'on désire minimiser une ou plusieurs grandeurs de commande et/ou un ou plusieurs états pour des problèmes de saturation, de sécurité ou de durée de vie du système (limitation de la vitesse d'un moteur, limitation de la pression dans un réservoir ...).

6.5.4 Exemple : fil rouge.

Calcul d'une commande qui minimise les pertes énergétiques dans le système. Les pertes sont :
$\frac{1}{2}Ri^2$: pertes Joule dans la résistance d'inducteur
$\frac{1}{2}f_v\omega^2$: pertes mécaniques par frottements visqueux
Le critère est donc

$$J = \int_0^\infty (Ri^2 + f_v\omega^2)dt$$

sous forme matricielle

$$J = \int_0^\infty (\underline{x}^T \begin{pmatrix} 0 & 0 \\ 0 & f_v \end{pmatrix} \underline{x} + \underline{u}^T(R)\underline{u})dt \tag{6.8}$$

En posant (**K** symétrique)

$$\mathbf{K} = \begin{pmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{pmatrix},$$

le calcul de la commande

$$u = -\begin{pmatrix} \frac{k_{12}\,b_0}{r} & \frac{k_{22}\,b_0}{r} \end{pmatrix} \underline{x}$$

montre que seuls les coefficients k_{12} et k_{22} sont à déterminer. En développant l'équation de Riccati, nous obtenons :

$$\begin{bmatrix} -2\,k_{12}\,a_0 - \frac{{k_{12}}^2{b_0}^2}{r} & k_{11} - k_{12}\,a_1 - k_{22}\,a_0 - \frac{k_{12}\,{b_0}^2 k_{22}}{r} \\ k_{11} - k_{12}\,a_1 - k_{22}\,a_0 - \frac{k_{12}\,{b_0}^2 k_{22}}{r} & 2\,k_{12} - 2\,k_{22}\,a_1 - \frac{{k_{22}}^2{b_0}^2}{r} + f_v \end{bmatrix} = 0$$

Nous en déduisons que :

$$\begin{aligned} k_{12} &= 0 \\ k_{22} &= 1/2\,\frac{-2\,a_1\,r + 2\sqrt{{a_1}^2r^2 + {b_0}^2 f_v\,r}}{{b_0}^2} \\ &\text{ou} \\ k_{22} &= 1/2\,\frac{-2\,a_1\,r - 2\sqrt{{a_1}^2r^2 + {b_0}^2 f_v\,r}}{{b_0}^2} \end{aligned}$$

comme **K** est définie négative, la solution est :

$$\begin{aligned} k_{12} &= 0 \\ k_{22} &= 1/2\,\frac{-2\,a_1\,r - 2\sqrt{{a_1}^2r^2 + {b_0}^2 f_v\,r}}{{b_0}^2} \end{aligned}$$

donc la commande est :

$$u = -\begin{pmatrix} 0 & \frac{b_0}{2r}\,\frac{-2\,a_1\,r - 2\sqrt{{a_1}^2r^2 + {b_0}^2 f_v\,r}}{{b_0}^2} \end{pmatrix} \underline{x}$$

Chapitre 7

Synthèse d'observateurs d'état

7.1 Introduction au problème de la reconstruction d'état

Dans tout ce qui précède, nous sommes partis du principe que nous avions accès à toutes les composantes du vecteur d'état. Nous avons donc supposé que le système est complètement instrumenté. En réalité, les systèmes physiques sont très peu instrumentés, les raisons sont :

- le coût,
- la difficulté d'accéder à certaines variables
- la fiabilité,
- l'encombrement, ...

Nous allons donc voir comment "reconstruire" l'état à partir des commandes appliquées au système réel et les quelques mesures du vecteur d'état effectuées.
Soit le système,

$$\left|\begin{array}{rcl} \dot{\underline{x}} &=& \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \\ \underline{y} &=& \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \end{array}\right. \quad \text{ou} \quad \left|\begin{array}{rcl} \underline{x}_{k+1} &=& \mathbf{\Phi}\underline{x}_k + \mathbf{\Gamma}\underline{u}_k \\ \underline{y}_k &=& \mathbf{C}\underline{x}_k + \mathbf{D}\underline{u}_k \end{array}\right. \tag{7.1}$$

Comment obtenir $\underline{x}$?

7.1.1 Par calcul direct

Avec cette méthode x est obtenu par la formule suivante :

$$\underline{x} \stackrel{?}{=} \mathbf{C}^{-1}(\underline{y} - \mathbf{D}\underline{u})$$

Ce calcul est impossible car en général $\mathbf{C}$ n'est pas inversible, dans les systèmes SISO, $\mathbf{C}$ est un vecteur, donc jamais inversible.

7.1.2 Par simulation du processus

L'idée consiste à dire que puisque nous possédons un modèle du système, il suffit de simuler le processus dans le calculateur. Ainsi, les différentes variables comme le vecteur d'état sont parfaitement accessibles. Cette méthode est illustrée par le schéma 7.1.
Les indices $_r$ (réelle) et $_m$ (mesurée) sont là pour rappeler qu'il existe toujours une différence entre la réalité et le modèle utilisé. Par la suite, comme dans tout ce qui précède, nous ne ferons aucune différence entre les deux mais ayez toujours à l'esprit qu'elle existe.
En fait cette méthode ne fonctionne pas du tout car le modèle, pour précis qu'il soit, est toujours faux (imprécisions de mesure des coefficients des matrices, non linéarités du système, ...).Aussi après quelques temps de simulation le simulateur donne n'importe quoi. Néanmoins dans quelques cas, ce sera la seule solution, pour obtenir une variable non mesurée, on parle alors de reconstruction par simulation. La robustesse du processus corrigé est alors faible et la synthèse de la commande doit prendre en compte cet aspect (commande robuste).

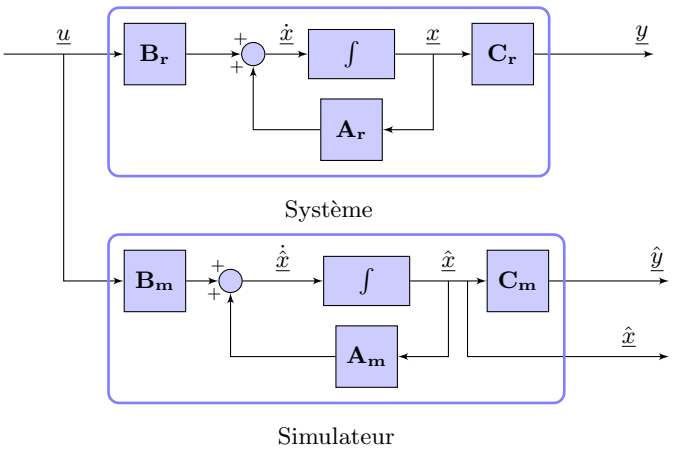


FIGURE 7.1 – Schéma général d'un simulateur

7.1.3 Par simulation du processus et asservissement sur les parties connues du vecteur d'état

L'idée consiste comme précédemment à simuler le système, mais à l'asservir de façon à ce que les sorties mesurées concordent avec les sorties simulées, en injectant à l'entrée l'erreur d'estimation pondérée par un gain.
C'est exactement ce que l'on appelle un observateur de Luenberger.

7.2 Observateurs de Luenberger

Définition

On appelle observateur (de Luenberger) du système,

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{7.2}$$
$$\underline{y} = \mathbf{C}\underline{x} \tag{7.3}$$

un système qui est décrit par le schéma fonctionnel 7.2,

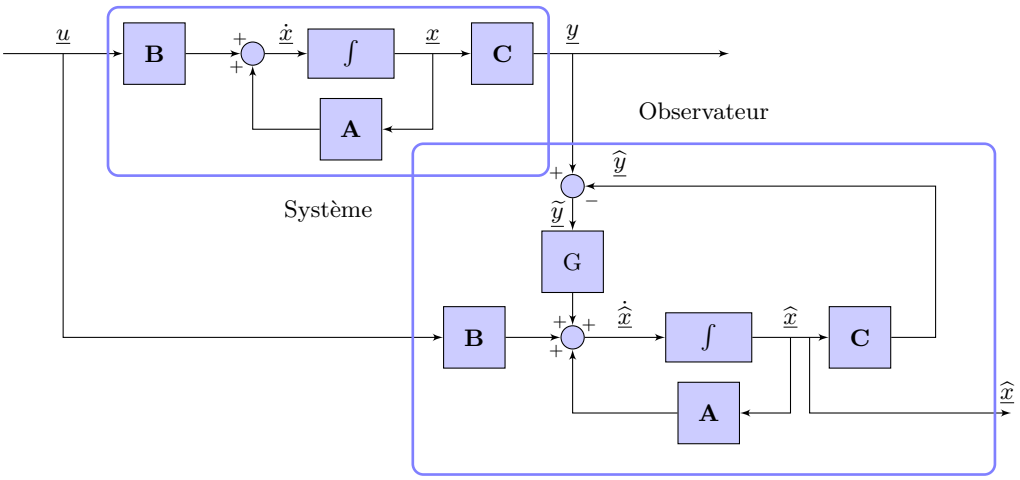


FIGURE 7.2 – Schéma général d'un observateur de Luenberger complet.

l'équation différentielle d'état de l'observateur est :

$$\begin{aligned}\dot{\underline{\hat{x}}} &= \mathbf{A}\underline{\hat{x}} + \mathbf{B}\underline{u} + \mathbf{G}\underline{\tilde{y}} \\ &= \mathbf{A}\underline{\hat{x}} + \mathbf{B}\underline{u} + \mathbf{G}\underline{y} - \mathbf{G}\underline{\hat{y}} \\ &= \mathbf{A}\underline{\hat{x}} + \mathbf{B}\underline{u} + \mathbf{G}\underline{y} - \mathbf{GC}\underline{\hat{x}} \\ &= (\mathbf{A} - \mathbf{GC})\underline{\hat{x}} + \mathbf{B}\underline{u} + \mathbf{G}\underline{y}\end{aligned}$$

En définissant

$$\mathbf{A} - \mathbf{GC} = \mathbf{F}$$

$$\dot{\underline{\hat{x}}} = \mathbf{F}\underline{\hat{x}} + \mathbf{G}\underline{y} + \mathbf{B}\underline{u} \tag{7.4}$$

où les valeurs propres de la matrice $\mathbf{F}$ sont stables.
$\underline{\hat{x}}(t)$ est une estimation de $\underline{x}(t)$,
on définit $\underline{\tilde{x}} = \underline{x} - \underline{\hat{x}}$ l'erreur d'estimation et $\underline{\tilde{x}} \to 0$ pour $t \to \infty$.
Il ne reste plus qu'à définir la matrice de "retour" $\mathbf{G}$.

Dualité

$$\begin{array}{rcl}
\text{synthèse d'un régulateur d'état} & \longrightarrow & \text{synthèse d'un observateur} \\
\mathbf{A} & \longrightarrow & \mathbf{A}^{\mathrm{T}} \\
\mathbf{B} & \longrightarrow & \mathbf{C}^{\mathrm{T}} \\
\mathbf{L} & \longrightarrow & \mathbf{G}^{\mathrm{T}} \\
\det\left[p\mathbf{I}-(\mathbf{A}-\mathbf{BR})\right] & \longrightarrow & \det\left[p\mathbf{I}-(\mathbf{A}^{\mathrm{T}}-\mathbf{C}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}})\right] \\
\left[\mathbf{B},\mathbf{AB},\mathbf{A}^2\mathbf{B},\cdots,\mathbf{A}^{n-1}\mathbf{B}\right] & \longrightarrow & \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^{n-1} \end{bmatrix}^{\mathrm{T}}
\end{array}$$

Toutes les méthodes de calcul d'un régulateur sont valables pour le calcul de la matrice $\mathbf{G}^{\mathrm{T}}$.

7.3 Observateurs d'ordre réduit

Il est parfois intéressant de n'observer que la partie de l'état qui n'est pas accessible afin de minimiser le temps de calcul. On définit alors un observateur d'ordre réduit.

7.3.1 Hypothèses

soit le système :

$$\begin{array}{rcl} \dot{\underline{x}} & = & \mathbf{A}\underline{x}+\mathbf{B}\underline{u} \\ \underline{y} & = & \mathbf{C}\underline{x} \end{array} \tag{7.5}$$

avec :

$$\mathbf{C} = \underbrace{\left.\begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}\right\}}_{n-q\qquad\qquad q} \quad q = [0 \;\; \mathbf{I}_q] \tag{7.6}$$

(7.7)

Bien entendu, il est probable qu'il faille procéder à une transformation de coordonnées pour mettre $\mathbf{C}$ sous la forme ci-dessus.

Partitionnement

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{n-q} \\ \cdots\cdots \\ x_{n-q+1} \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \underline{v} \\ \cdots \\ \underline{y} \end{bmatrix} \begin{matrix} \} \, n-q \\ \\ \} \, q \end{matrix}$$

$$\dot{\underline{x}} = \begin{bmatrix} \dot{\underline{v}} \\ \cdots \\ \dot{\underline{y}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \cdots & \cdots \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \underline{v} \\ \cdots \\ \underline{y} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \cdots \\ \mathbf{B}_2 \end{bmatrix} u \begin{matrix} \} \, n-q \\ \\ \} \, q \end{matrix}$$

soit :

$$\dot{\underline{v}} = \mathbf{A}_{11}\underline{v} + \mathbf{A}_{12}\underline{y} + \mathbf{B}_1\underline{u} \tag{7.8}$$
$$\dot{\underline{y}} = \mathbf{A}_{21}\underline{v} + \mathbf{A}_{22}\underline{y} + \mathbf{B}_2\underline{u} \tag{7.9}$$

Il reste à estimer $\underline{v}$, ($n-q$ composantes) uniquement.
Idée de base : Considérer ces deux équations comme l'équation différentielle d'état et l'équation de sortie d'un système qui aurait

$\underline{v}$	:	comme vecteur d'état
$\mathbf{A}_{12}\underline{y} + \mathbf{B}_1\underline{u}$	:	comme vecteur d'entrée
$\dot{\underline{y}} - \mathbf{A}_{22}\underline{y} - \mathbf{B}_2\underline{u}$	:	comme vecteur de sortie

$$\underbrace{\dot{\underline{v}}}_{\dot{X}} = \underbrace{\mathbf{A}_{11}}_{A}\underbrace{\underline{v}}_{X} + \underbrace{\mathbf{A}_{12}\underline{y} + \mathbf{B}_1\underline{u}}_{BU} \tag{7.10}$$
$$\underbrace{\dot{\underline{y}} - \mathbf{A}_{22}\underline{y} - \mathbf{B}_2\underline{u}}_{Y} = \underbrace{\mathbf{A}_{21}}_{C}\underbrace{\underline{v}}_{X} \tag{7.11}$$

Développons un observateur complet (d'ordre $n-q$) pour ce système

$$\dot{\widehat{\underline{v}}} = (\mathbf{A}_{11} - \mathbf{G}\mathbf{A}_{21})\widehat{\underline{v}} + \mathbf{A}_{12}\underline{y} + \mathbf{B}_1\underline{u} + \mathbf{G}(\dot{\underline{y}} - \mathbf{A}_{22}\underline{y} - \mathbf{B}_2\underline{u})$$

Posons : $\underline{z} = \widehat{\underline{v}} - \mathbf{G}\underline{y}$
D'où les équations d'état de l'observateur :

$$\dot{\underline{z}} = (\mathbf{A}_{11} - \mathbf{G}\mathbf{A}_{21})\underline{z} + [(\mathbf{A}_{11} - \mathbf{G}\mathbf{A}_{21})\mathbf{G}\mathbf{A}_{12} - \mathbf{G}\mathbf{A}_{22})]\,\underline{y}(\mathbf{B}_1 - \mathbf{G}\mathbf{B}_2)\underline{u} \tag{7.12}$$
$$\widehat{\underline{v}} = z + \mathbf{G}\underline{y} \tag{7.13}$$

dont le schéma fonctionnel est donné en figure 7.3

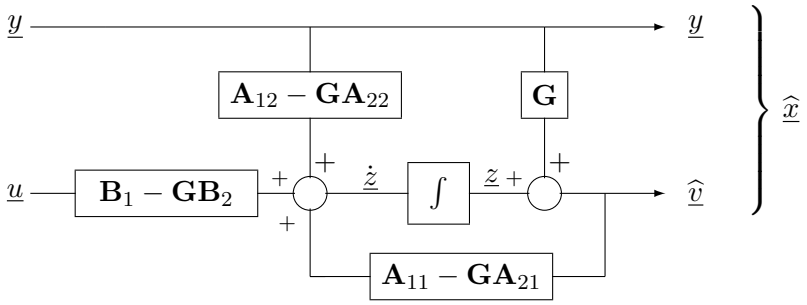


FIGURE 7.3 – Schéma fonctionnel d'un observateur de Luenberger d'ordre réduit

Remarque : La construction précédente fournit un observateur de ce type, avec la matrice de système

$$\mathbf{F} = \mathbf{A}_{11} - \mathbf{G}\mathbf{A}_{21}$$

7.4 Observateur généralisé

Pour observer un système linéaire invariant, on introduit un second système linéaire invariant, d'ordre r et ayant l'équation d'état suivante :

$$\dot{\underline{z}}(t) = \mathbf{F}\underline{z}(t) + \mathbf{G}\underline{y}(t) + \mathbf{E}\underline{u}(t)$$

On dit qu'un tel système est un observateur du premier système si, étant donné une matrice de transformation arbitraire $\mathbf{K}$ de dimension $(r \times n)$, son vecteur d'état représente une estimation de $\mathbf{K}\underline{x}$:

$$\underline{z} = \mathbf{K}\widehat{\underline{x}}$$

c'est-à-dire que $\underline{z} \to \mathbf{K}\widehat{\underline{x}}$ pour $t \to \infty$
$\underline{z}$ doit donc être solution d'une équation homogène, de la forme

$$(\underline{z} - \dot{\mathbf{K}}\underline{x}) = \mathbf{F}(\underline{z} - \mathbf{K}\underline{x})$$

$$\dot{\underline{z}} = \mathbf{F}\underline{z} + \mathbf{K}\dot{\underline{x}} - \mathbf{F}\mathbf{K}\underline{x}$$

et

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u}$$

donc

$$\dot{\underline{z}} = \mathbf{F}\underline{z} + (\mathbf{K}\mathbf{A} - \mathbf{F}\mathbf{K})\widehat{\underline{x}} - \mathbf{K}\mathbf{B}\underline{u}$$

d'où l'on en déduit que

$$\mathbf{E} = \mathbf{K}\mathbf{B}$$

Cherchons à introduire $\underline{y} = \mathbf{C}\underline{x}$, on cherche à trouver une matrice $\mathbf{G}$ telle que

$$(\mathbf{K}\mathbf{A} - \mathbf{F}\mathbf{K})\underline{x} = \mathbf{G}\underline{y} = \mathbf{G}\mathbf{C}\underline{x}$$

donc

$$(\mathbf{K}\mathbf{A} - \mathbf{F}\mathbf{K}) = \mathbf{G}\mathbf{C}$$

d'où l'équation d'état de l'observateur généralisé

$$\dot{\underline{z}} = \mathbf{F}\underline{z} + \mathbf{G}\underline{y} + \mathbf{K}\mathbf{B}\underline{u}$$

7.5 Equation d'état d'un système asservi avec observateur

Loi de commande

$$\underline{u} = -\mathbf{L}\underline{\widehat{x}}$$

à $\underline{w} = 0$, étude de la stabilité
Vu du point de vue de l'observateur généralisé, $\underline{u}$ doit dépendre de la grandeur de sortie $\underline{y}$ du procédé et du vecteur d'état $\underline{z}$ de l'observateur :

$$\underline{u} = \mathbf{D}\underline{\widehat{y}} + \mathbf{E}\underline{\widehat{z}}$$

avec $\underline{z} = -\mathbf{K}\underline{\widehat{x}}$ et $\underline{y} = \mathbf{C}\underline{x}$ il vient

$$-\mathbf{L}\underline{\widehat{x}} = \mathbf{DC}\underline{x} + \mathbf{EK}\underline{\widehat{x}}$$

cette expression doit être valable quel que soit t, donc aussi pour $t \to \infty(\underline{\widehat{x}} = \underline{x})$ donc,

$$-\mathbf{L} = \mathbf{DC} + \mathbf{EK}$$

donc, l'équation d'état de l'ensemble observateur+ régulateur est

$$\begin{array}{rcl} \dot{\underline{z}} & = & \mathbf{F}\underline{z} + \mathbf{G}\underline{y} + \mathbf{KB}\underline{u} \\ \underline{u} & = & \mathbf{D}\underline{y} + \mathbf{E}\underline{z} \end{array} \tag{7.14}$$

schéma fonctionnel du système asservi

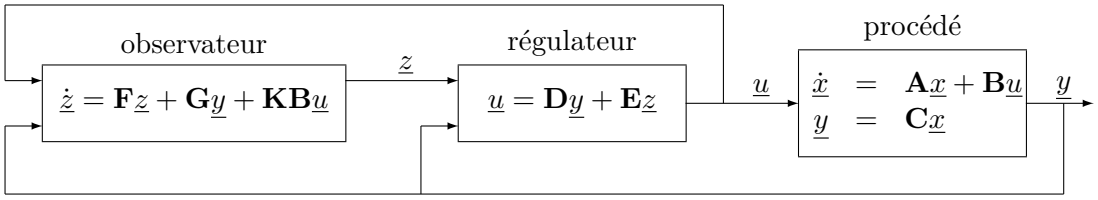


FIGURE 7.4 – Schéma fonctionnel d'un système asservi via un observateur

vecteur d'état de ce système composite : $\begin{bmatrix} \underline{x} \\ \underline{z} \end{bmatrix}$

7.5.1 Théorème de séparation

posons : $\xi = \underline{z} - \mathbf{K}\underline{x}$
pour l'observateur :

$$\begin{array}{rcll} \dot{\xi} & = & \dot{\underline{z}} - \mathbf{K}\dot{\underline{x}} & \\ & = & \mathbf{F}\underline{z} + \mathbf{G}\underline{y} + \mathbf{KB}\underline{u} - \mathbf{KA}\underline{x} - \mathbf{KB}\underline{u} & \\ & = & \mathbf{F}(\underline{z} - \mathbf{K}\underline{x}) & \text{car } (\mathbf{KA} - \mathbf{FK}) = \mathbf{GC} \\ & = & \mathbf{K}\underline{\xi} & \end{array} \tag{7.15}$$

pour le procédé :

$$\begin{array}{rcll} \dot{\underline{x}} & = & \mathbf{A}\underline{x} + \mathbf{B}\underline{u} & \\ & = & \mathbf{A}\underline{x} + \mathbf{B}(\mathbf{D}\underline{y} + \mathbf{E}\underline{z}) & \\ & = & \mathbf{A}\underline{x} + \mathbf{BDC}\underline{x} + \mathbf{BE}\underline{z} & \\ & = & \mathbf{A}\underline{x} - \mathbf{BL}\underline{x} - \mathbf{BEK}\underline{x} + \mathbf{BE}\underline{z} & \text{car } (\mathbf{DC} - \mathbf{EK}) = -\mathbf{L} \\ & = & (\mathbf{A} - \mathbf{BL})\underline{x} + \mathbf{BE}\underline{\xi} & \end{array} \tag{7.16}$$

Des deux développements précédents nous en déduisons l'équation d'état du système composite observateur + procédé.

$$\begin{pmatrix} \dot{\underline{x}} \\ \dot{\underline{\xi}} \end{pmatrix} = \begin{pmatrix} \mathbf{A}-\mathbf{BL} & \mathbf{BE} \\ 0 & \mathbf{F} \end{pmatrix} \begin{pmatrix} \underline{x} \\ \underline{\xi} \end{pmatrix}$$

dont l'équation caractéristique est :

$$\begin{vmatrix} \lambda\mathbf{I}_n - (\mathbf{A}-\mathbf{BL}) & -\mathbf{BE} \\ 0 & \lambda\mathbf{I}_r - \mathbf{F} \end{vmatrix} = 0$$

c'est-à-dire :

$$|\lambda\mathbf{I}_n - (\mathbf{A}-\mathbf{BL})| \cdot |\lambda\mathbf{I}_r - \mathbf{F}| = 0$$

d'où :
théorème de séparation :
Les valeurs propres d'un système commandé par retour d'état et comportant un observateur dans sa boucle se composent de la réunion des valeurs propres du système bouclé sans observateur et de celles de l'observateur.

7.6 Filtrage de Kalman

Soit le système,

$$\left| \begin{array}{rcl} \dot{\underline{x}} &=& \mathbf{A}\underline{x} + \mathbf{B}\underline{u} + \mathbf{K}\underline{v} \\ \underline{y} &=& \mathbf{C}\underline{x} + \mathbf{D}\underline{u} + \underline{w} \end{array} \right. \quad \text{ou} \quad \left| \begin{array}{rcl} \dot{\underline{x}}_{k+1} &=& \mathbf{\Phi}\underline{x}_k + \mathbf{\Gamma}\underline{u}_k + \mathbf{K}\underline{v}_k \\ \underline{y}_k &=& \mathbf{C}\underline{x}_k + \mathbf{D}\underline{u}_k + \underline{w}_k \end{array} \right. \tag{7.17}$$

avec :
$\underline{v}(t)$ ou $\underline{v}_k$ vecteur aléatoire superposé à l'état (bruit d'état)
$\underline{w}(t)$ ou $\underline{w}_k$ vecteur aléatoire superposé à la sortie (bruit de mesure)

Problème du filtrage linéaire :
A partir des données du problème (représentation d'état du problème) et des données statistiques concernant les bruits $\underline{v}$ et $\underline{w}$ (distribution statistique, moyenne, variance), trouver un système causal, à l'entrée duquel sont appliquées les signaux accessibles à la mesure, u et y et qui fournit à sa sortie $\widehat{\underline{x}}$ aussi proche que possible de l'état $\underline{x}$ inconnu.
La solution optimale de ce problème, au sens de la variance de $\widehat{\underline{x}} - \underline{x}$ est le filtrage optimal de Kalman.

Chapitre 8

Représentation d'état des systèmes linéaires échantillonnés

8.1 Système discret

Le schéma de la figure 8.1 représente la forme générale d'un système discret.

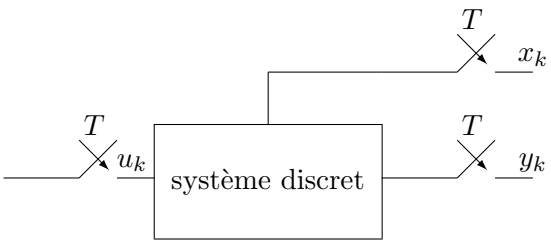


FIGURE 8.1 – Représentation générale d'un système discret

On définit :

$$x_k = \begin{pmatrix} x_1(kT_e) \\ x_2(kT_e) \\ \vdots \\ x_n(kT_e) \end{pmatrix}, \quad y_k = \begin{pmatrix} y_1(kT_e) \\ y_2(kT_e) \\ \vdots \\ y_q(kT_e) \end{pmatrix}, \quad u_k = \begin{pmatrix} u_1(kT_e) \\ u_2(kT_e) \\ \vdots \\ u_p(kT_e) \end{pmatrix}$$

La résolution de l'équation d'état donne (voir (1.63, page 14) :

$$\underline{x}(t) = e^{\mathbf{A}t}\left[\int_0^t e^{-\mathbf{A}\tau}\mathbf{B}u(\tau)d\tau\right] + e^{\mathbf{A}(t-t_0)}\underline{x}(t_0) \tag{8.1}$$

Appliquons (8.1) à l'intervalle $kT_e \leq t < (k+1)T_e$ en y faisant $t_0 = kt$, nous obtenons :

$$\underline{x}_{k+1} = e^{\mathbf{A}T_e}\underline{x}_k + \int_{kT_e}^{(k+1)T_e} e^{\mathbf{A}((k+1)T_e-\tau)}\mathbf{B}\underline{u}(\tau)d\tau$$

Supposons que $\underline{u}(t) = \underline{u}_k = cte$ pour $kT_e \leq t < (k+1)T_e$ alors,

$$\underline{x}_{k+1} = e^{\mathbf{A}T_e}\underline{x}_k + \int_{kT_e}^{(k+1)T_e} e^{\mathbf{A}((k+1)T_e-\tau)}d\tau.\mathbf{B}\underline{u}_k \tag{8.2}$$

En procédant au changement de variable : $\tau' = (k+1)T_e - \tau$ alors :

$$\int_{kT_e}^{(k+1)T_e} e^{\mathbf{A}((k+1)T_e-\tau)}d\tau = -\int_{T_e}^{0} e^{\mathbf{A}\tau'}d\tau' = \int_0^{T_e} e^{\mathbf{A}\tau'}d\tau'$$

l'équation 8.2 devient :

$$\underline{x}_{k+1} = e^{\mathbf{A}T_e}\underline{x}_k + \int_0^{T_e} e^{\mathbf{A}\tau'}d\tau'.\mathbf{B}\underline{u}_k. \tag{8.3}$$

On voit que x_{k+1} peut se mettre sous la forme :

$$\underline{x}_{k+1} = \mathbf{\Phi}\underline{x}_k + \mathbf{\Gamma}\underline{u}_k. \tag{8.4}$$

avec :

$$\mathbf{\Phi} = e^{\mathbf{A}T_e} \tag{8.5}$$

$$\mathbf{\Gamma} = \int_0^{T_e} e^{\mathbf{A}\tau'}d\tau'.\mathbf{B} \tag{8.6}$$

De même on a :

$$\underline{y}_k = \mathbf{C}\underline{x}_k + \mathbf{D}\underline{u}_k. \tag{8.7}$$

Cas particulier : si $\mathbf{A}$ est inversible,

$$\mathbf{\Gamma} = \mathbf{A}^{-1}\left[e^{\mathbf{A}\tau}\right]_0^{T_e}.\mathbf{B} = \mathbf{A}^{-1}\left(e^{\mathbf{A}T_e} - \mathbf{I}\right).\mathbf{B} \tag{8.8}$$

soit :

$$\mathbf{\Gamma} = \mathbf{A}^{-1}(\mathbf{\Phi} - \mathbf{I})\mathbf{B} \tag{8.9}$$

8.2 Résolution des équations dans le domaine du temps

En supposant que le vecteur d'état initial soit x_0, l'application répétée de (8.4) donne :

$$\underline{x}_k = \mathbf{\Phi}^k\underline{x}_0 + \sum_{i=0}^{k-1}\mathbf{\Phi}^{k-1-i}\mathbf{\Gamma}\underline{u}_i, \tag{8.10}$$

la résolution se fait alors sur ordinateur.

8.3 Application de la transformée en z

Théorème de l'avance :

$$\mathcal{Z}[\underline{x}_{k+1}] = z\underline{X}(z) - z\underline{x}_0,$$

donc la transformée en z de 8.4 est :

$$z\underline{X}(z) - z\underline{x}_0 = \mathbf{\Phi}\underline{X}(z) + \mathbf{\Gamma}\underline{U}(z),$$

soit :

$$\underline{X}(z) = (z\mathbf{I} - \mathbf{\Phi})^{-1}z\underline{x}_0 + (z\mathbf{I} - \mathbf{\Phi})^{-1}\mathbf{\Gamma}\underline{U}(z), \tag{8.11}$$

et,

$$\underline{Y}(z) = \mathbf{C}\underline{X}(z) + \mathbf{D}\underline{U}(z) \tag{8.12}$$

8.4 Matrice de transfert

Dans le cas SISO : Fonction de transfert.
En supposant $x_0 = 0$, des équations (8.11) et (8.12) on déduit :

$$\underline{Y}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{\Phi})^{-1}\mathbf{\Gamma}\underline{U}(z) + \mathbf{D}\underline{U}(z).$$

soit

$$\underline{Y}(z) = G(z)\underline{U}(z),$$

avec :

$$G(z) = \mathbf{C}(z\mathbf{I} - \mathbf{\Phi})^{-1}\mathbf{\Gamma} + \mathbf{D}.$$

$G(z)$ est la fonction (matrice dans le cas multivariable) de transfert du système échantillonné.
A comparer avec le résultat obtenu dans le cas continu :

$$\mathbf{H}(p) = \mathbf{C}(p\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$

8.5 Obtention d'un modèle d'état à partir de la fonction de transfert en z

Le modèle d'état s'obtient de la même façon qu'en continu mais avec la substitution

$$\text{intégrateur pur} \rightarrow \text{retard pur}$$

$$\frac{1}{p} \rightarrow z^{-1}$$

8.6 Résolution de l'équation d'état dans le domaine de z

La réponse libre du système ($u_k = 0$) s'obtient en utilisant (8.10) et (8.11).

(8.10) devient :	$\underline{x}_k$	$= \mathbf{\Phi}^k \underline{x}_0$
sa transformée en z est :	$\underline{X}(z)$	$= \mathcal{Z}\left[\mathbf{\Phi}^k\right]\underline{x}_0$
(8.11) devient :	$\underline{X}(z)$	$= (z\mathbf{I} - \mathbf{\Phi})^{-1} z\underline{x}_0$

en identifiant terme à terme :

$$\mathcal{Z}\left[\mathbf{\Phi}^k\right] = (z\mathbf{I} - \mathbf{\Phi})^{-1} z,$$

d'où,

$$\mathbf{\Phi}(kT_e) = \mathbf{\Phi}^k = \mathcal{Z}^{-1}\left[(z\mathbf{I} - \mathbf{\Phi})^{-1} z\right],$$

à comparer avec : $\mathcal{L}^{-1}(p\mathbf{I} - \mathbf{A})^{-1}$.

8.7 Commandabilité et observabilité

8.7.1 Commandabilité d'un système échantillonné

Un système échantillonné représenté par une équation d'état aux différences pour une entrée scalaire u_k,

$$\underline{x}_{k+1} = \mathbf{\Phi}\underline{x}_k + \mathbf{\Gamma}u_k. \tag{8.13}$$

$$\underline{y}_k = \mathbf{C}\underline{x}_k \tag{8.14}$$

$x(0)$ étant donné, le vecteur d'état solution de l'équation (8.13) s'exprime par : (en écrivant les termes successifs) :

$$\underline{x}_k = \mathbf{\Phi}^k \underline{x}(0) + \mathbf{\Gamma}\underline{u}_{k-1} + \mathbf{\Phi}\mathbf{\Gamma}\underline{u}_{k-2} + \mathbf{\Phi}^2\mathbf{\Gamma}\underline{u}_{k-3} + \cdots + \mathbf{\Phi}^{k-1}\mathbf{\Gamma}\underline{u}_0 \tag{8.15}$$

$$\underline{x}_k = \mathbf{\Phi}^k \underline{x}(0) + \sum_{i=1}^{k} \mathbf{\Phi}^{i-1}\mathbf{\Gamma}\underline{u}_{k-i} \tag{8.16}$$

que l'on peu réécrire sous la forme :

$$\underline{x}_k = \mathbf{\Phi}^k \underline{x}(0) + \mathbf{Q}_S \underline{U}^T \tag{8.17}$$

avec : $\begin{array}{rcl} \mathbf{Q}_S & = & \left[\mathbf{\Gamma}, \mathbf{\Phi}\mathbf{\Gamma}, \mathbf{\Phi}^2\mathbf{\Gamma}, \cdots, \mathbf{\Phi}^{k-1}\mathbf{\Gamma}\right] \\ \underline{U} & = & \left[\underline{u}_{k-1}, \underline{u}_{k-2}, \cdots, \underline{u}_1, \underline{u}_0\right] \end{array}$

Pour que l'état du système passe de l'état $\underline{x}_0$ à l'état $\underline{x}_k$, il faut que la séquence de commande $\underline{U}^T$ respecte

$$\mathbf{Q}_S \underline{U}^T = \underline{x}_k - \mathbf{\Phi}^k \underline{x}(0). \tag{8.18}$$

Cette séquence existe si $\mathbf{Q}_S$ est inversible donc :

$$\underline{U}^T = \mathbf{Q}_S^{-1}\left[\underline{x}_k - \mathbf{\Phi}^k \underline{x}(0)\right]. \tag{8.19}$$

Les conditions de commandabilité sont donc que :

1. $\mathbf{Q}_S$ soit carrée,
2. $\det(\mathbf{Q}_S) \neq 0$.

Ainsi la condition de commandabilité devient (n est le nombre d'états) :

$$\det \mathbf{Q}_S = \det\left[\mathbf{\Gamma}, \mathbf{\Phi}\mathbf{\Gamma}, \mathbf{\Phi}^2\mathbf{\Gamma}, \cdots, \mathbf{\Phi}^{n-1}\mathbf{\Gamma}\right] \neq 0. \tag{8.20}$$

8.7.2 Observabilité d'un système échantillonné

La sortie d'un système échantillonné peut être écrite sous la forme :

$$\underline{y}_k = \mathbf{C}\mathbf{\Phi}^k \underline{x}_0 + \sum_{i=1}^{k} \mathbf{C}\mathbf{\Phi}^{i-1}\mathbf{\Gamma}\underline{u}_{k-i} \tag{8.21}$$

$$\begin{array}{rcllll} \underline{y}_0 & = & \mathbf{C}\underline{x}_0 & & & \\ \underline{y}_1 & = & \mathbf{C}\underline{x}_1 & = \mathbf{C}\mathbf{\Phi}\underline{x}_0 + \mathbf{C}\mathbf{\Gamma}\underline{u}_0 & & \\ \underline{y}_2 & = & \mathbf{C}\underline{x}_2 & = \mathbf{C}\mathbf{\Phi}\underline{x}_1 + \mathbf{C}\mathbf{\Gamma}\underline{u}_1 & = \mathbf{C}\mathbf{\Phi}^2\underline{x}_0 + \mathbf{C}\mathbf{\Phi}\mathbf{\Gamma}\underline{u}_0 + \mathbf{C}\mathbf{\Gamma}\underline{u}_1 & \\ \vdots & & & & & \\ \underline{y}_{k-1} & = & \mathbf{C}\underline{x}_{k-1} & = \mathbf{C}\mathbf{\Phi}\underline{x}_{k-2} + \mathbf{C}\mathbf{\Gamma}\underline{u}_{k-2} & = \mathbf{C}\mathbf{\Phi}^{k-1}\underline{x}_0 + \mathbf{C}\mathbf{\Phi}^{k-2}\mathbf{\Gamma}\underline{u}_0 + \cdots + \mathbf{C}\mathbf{\Gamma}\underline{u}_{k-2} \end{array}$$

que l'on peut réécrire sous la forme :

$$\underline{Y}^T = \mathbf{Q}_B \underline{x}_0 + \mathbf{H}\underline{U_1} \tag{8.22}$$

avec : $\begin{array}{rcl} \mathbf{Q}_B^T & = & \left[\mathbf{C}, \mathbf{C}\mathbf{\Phi}, \mathbf{C}\mathbf{\Phi}^2, \cdots, \mathbf{C}\mathbf{\Phi}^{k-1}\right] \\ \underline{U_1} & = & \left[0, \underline{u}_0, \underline{u}_1, \cdots, \underline{u}_{k-2}\right] \end{array}$

donc

$$\mathbf{Q}_B \underline{x}_0 = \underline{Y}^T - \mathbf{H}\underline{U_1} \tag{8.23}$$

$$\underline{x}_0 = \mathbf{Q}_B{}^{-1}(\underline{Y}^T - \mathbf{H}\underline{U_1}) \tag{8.24}$$

Remonter à l'état initial $\underline{x}_0$ en mesurant $\underline{Y}^T$ il faut que :

1. $\mathbf{Q}_B$ soit carrée,
2. $\det(\mathbf{Q}_B) \neq 0$.

Un système est observable si et seulement si

$$\det \mathbf{Q}_B = \det \begin{bmatrix} \mathbf{C} \\ \mathbf{C\Phi} \\ \mathbf{C\Phi}^2 \\ \vdots \\ \mathbf{C\Phi}^{n-1} \end{bmatrix} \neq 0. \tag{8.25}$$

8.8 Stabilité des systèmes échantillonnés

La stabilité des systèmes échantillonnés est identique au cas continu sauf qu'un pôle stable en z est un pôle dont le module est inférieur à 1.

$$\text{stable} \Rightarrow \|z\| < 1$$

Bien entendu le critère de Routh appliqué sur la dernière ligne de la matrice sous la forme canonique de commandabilité devient le critère de Jury. Une autre solution consiste à appliquer le critère de Routh sur cette dernière ligne après lui avoir appliqué la transformée en w.
Rappel : la transformée en w consiste à remplacer z par $\frac{w-1}{w+1}$

8.9 Commandes des systèmes échantillonnés

La commande des systèmes échantillonnés se fait de la même façon que pour les systèmes continus. Les éventuelles transformations sont identiques.
Le choix des pôles en z dans le cadre d'une méthode de placement de pôles est par contre différent (voir fig. 8.2) .

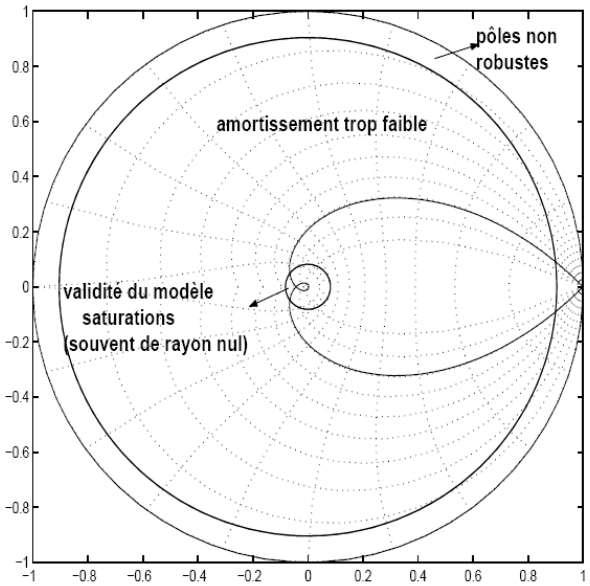


FIGURE 8.2 – Choix des pôles dans le plan en z

8.9.1 Calcul de la matrice de préfiltre

Le modèle du système en boucle fermée est :

$$\begin{aligned} \underline{x}_{k+1} &= (\mathbf{\Phi} - \mathbf{\Gamma L})\underline{x}_k + \mathbf{\Gamma S}\underline{y}_{c_k} \\ \underline{y}_k &= \mathbf{C}\underline{x}_k \end{aligned}$$

Si le système est stable, alors $\lim_{k\to\infty} \underline{x}_{k+1} = \underline{x}_k$ donc,

$$\lim_{k\to\infty} \underline{y}_k = -\mathbf{C}(\mathbf{\Phi} - \mathbf{\Gamma L} - \mathbf{I})^{-1}\mathbf{\Gamma S}\underline{y}_{c_k}$$

or, on désire que $\lim_{k\to\infty} \underline{y}_k = \underline{y}_{c_k}$, donc,

$$\mathbf{S}^{-1} = -\mathbf{C}(\mathbf{\Phi} - \mathbf{\Gamma L} - \mathbf{I})^{-1}\mathbf{\Gamma}$$

$$\mathbf{S} = -\Big(\mathbf{C}(\mathbf{\Phi} - \mathbf{\Gamma L} - \mathbf{I})^{-1}\mathbf{\Gamma}\Big)^{-1}$$

à comparer avec le résultat obtenu dans le cas continu

$$\mathbf{S} = -\Big(\mathbf{C}(\mathbf{A} - \mathbf{BL})^{-1}\mathbf{B}\Big)^{-1}$$

8.9.2 Commande optimale dans le cas discret

$$J = \sum_{n=1}^{\infty} \underline{x}_n^T \mathbf{Q}\underline{x}_n + \underline{u}_n^T \mathbf{R}\underline{u}_n \tag{8.26}$$

avec : $\mathbf{Q}$ = matrice symétrique définie positive
$\mathbf{R}$ = matrice symétrique non négative

la commande u est alors définie par l'équation suivante :

$$u_{k-1} = -(\mathbf{R} + \mathbf{\Gamma}^T\mathbf{K}\mathbf{\Gamma})^{-1}\mathbf{\Gamma K \Phi}\underline{x}_{k-1}$$

où $\mathbf{K}$ est une matrice symétrique, solution définie négative de l'équation de Riccati :

$$\mathbf{K} = \mathbf{\Phi}^T(\mathbf{K} - \mathbf{K\Gamma}(\mathbf{R} + \mathbf{\Gamma}^T\mathbf{K\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{K})\mathbf{\Phi} + \mathbf{Q}$$

Chapitre 9

Annales d'examens

Examen final d'automatique

Durée : 2 heures.
Aucun document autorisé.

Le sujet de cet examen est la modélisation à l'aide d'une représentation d'état d'une micro pince. Dans un deuxième temps, nous développerons une commande par retour d'état du système fondée sur un observateur complet de Luenberger. Enfin nous développerons un observateur de l'effort exercé par la pince sur l'objet.

Quelques conseils

— La justification des choix est aussi importante que les calculs effectués.
— Ne restez pas bloqué sur une question, revenez-y par la suite.

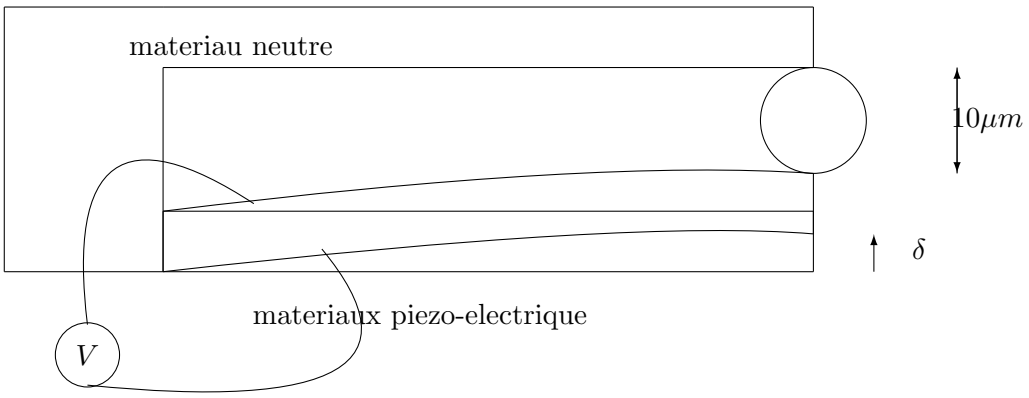


FIGURE 9.1 – Présentation de la pince étudiée.

Cette pince est en matériau piézo-électrique, la tension V appliqué sur les électrodes provoque le même effet qu'une force αV appliquée au bout de la mâchoire de la pince.
En première approximation le comportement dynamique de la pince est le même que celui du modèle mécanique présenté figure 9.2 :
La masse M est supposée ponctuelle, toutes les forces sont appliquées à une longueur L du centre de rotation.

Modélisation

1. En appliquant la relation fondamentale de la dynamique sur modèle mécanique précédent, déterminer l'équation différentielle reliant le mouvement de rotation aux forces appliquées.
2. En faisant l'hypothèse que $\theta << 1$ et que les force de gravité sont négligéables vis-à-vis des autres forces, déterminez l'équation différentielle linéarisée reliant $\ddot{\delta}$ aux autres paramètres du système.
3. Montrez que l'on obtient une équation différentielle de la forme :

$$m\ddot{\delta} + f\dot{\delta} + k\delta = -F_{ext} + \alpha V$$

Dans tout ce qui suit, l'équation différentielle du mouvement utilisée sera celle-ci !

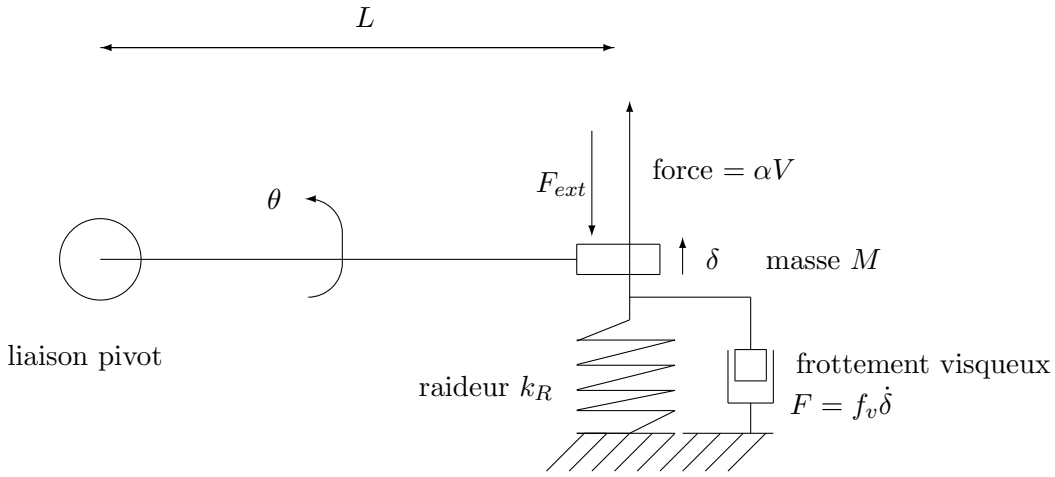


FIGURE 9.2 – Modèle mécanique équivalent.

4. En supposant que la force F_{ext} exercée par l'objet sur la pince est nulle. Déduisez de ce qui précède une représentation d'état de la forme :

$$\underline{\dot{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \quad (9.1)$$
$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \quad (9.2)$$

Applications numériques :
$\alpha = 0.002, \quad m = 10^{-3}, \quad f = 11, \quad k = 1^5$

Analyse

1. Le système est-il stable en boucle ouverte ?
2. Le système est-il commandable ?
3. Le système est-il observable ?

Commande

1. Donnez la valeur numérique des pôles en boucle ouverte.
2. Au vu des pôles en boucle ouverte, de votre expérience et du système proposez des pôles en boucle fermée.
3. Justifiez-les en 4 lignes maximum (+ un dessin si besoin).
4. Calculer un régulateur d'état qui permet d'obtenir la dynamique que vous vous êtes fixée en boucle fermée.
5. Par le calcul de la fonction de transfert, vérifiez que la dynamique souhaitée est bien obtenue.
6. A l'aide de cette fonction de transfert, calculer la matrice (ici un réel) de préfiltre.
7. Proposez une autre méthode pour obtenir une erreur statique nulle.

Observation

1. Donnez le schéma-bloc du système sous sa forme de représentation d'état avec son observateur de Luenberger complet.
2. Proposez et justifiez les pôles de l'observateur.
3. Calculer, par la méthode de votre choix, la matrice de "retour" **G**

Observation de la perturbation

Dans cette deuxième partie, la force F_{ext} exercée par l'objet sur la pince est non nulle, par contre on supposera de faibles variations de force, sa dérivé $\dot{F}_{ext}$ est nulle.

1. A partir des équations établies lors de la modélisation et avec les hypothèses précédentes, donnez une représentation d'état du système incluant la force F_{ext} exercée par l'objet sur la pince.

$$\dot{\underline{z}} = \mathbf{A}_a \underline{z} + \mathbf{B}_a u \tag{9.3}$$
$$\delta = \mathbf{C}_a \underline{z} + \mathbf{D}_a u \tag{9.4}$$

   avec : $\underline{z} = \begin{bmatrix} \underline{x} \\ F_{ext} \end{bmatrix}$

2. Donnez le schéma bloc du système complet incluant la pince, son observateur augmenté, le point d'application de la perturbation, le retour d'état utilisant le vecteur estimé, la force estimée $\hat{F}_{ext}$.
3. Montrez que lorsque $t \to \infty$ et en l'absence d'erreurs de modèle, $\hat{F}_{ext} \to F_{ext}$
4. Calculez la matrice de retour $\mathbf{G}$ pour ce nouvel observateur.

Examen de septembre d'automatique

Durée : 2 heures.
Aucun document autorisé.

Le sujet de cet examen est la modélisation à l'aide d'une représentation d'état d'un hélicoptère. Dans un deuxième temps, nous développerons une commande par retour d'état du système fondée sur un observateur complet de Luenberger. Enfin nous développerons un nouvel asservissement permettant de garantir une erreur statique nulle

Quelques conseils

— La justification des choix est aussi importante que les calculs effectués.
— Ne restez pas bloqué sur une question, revenez-y par la suite.

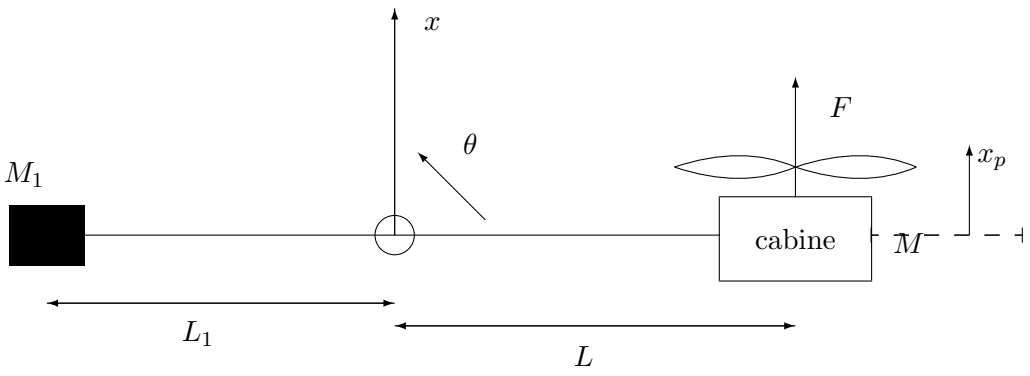


FIGURE 9.3 – Présentation de l'hélicoptère ($\theta = \pi/2$).

Nous supposerons que le problème est plan. La position instantanée de l'hélicoptère peut donc être décrite sans ambiguïté par l'angle du bras avec la verticale θ.
La cabine de masse M est surmontée d'une hélice, le rotor, qui provoque une force F qui tend à faire monter la cabine. Un contre-poids de masse M_1 est situé à une distance L_1 du centre de rotation. La liaison pivot présente un couple de frottement visqueux de la forme $f_v\dot{\theta}$ ainsi qu'un couple de rappel de la forme $k_R\theta$, ce couple est nul à la position d'équilibre en $\theta = \pi/2rad$.

Modélisation

1. Du modèle mécanique précédent, déterminer l'équation différentielle reliant le mouvement de l'hélicoptère à la force appliquée par le rotor F.
2. En faisant l'hypothèse que $\theta << 1$, déterminez l'équation différentielle linéarisée reliant $\ddot{x}_p$ aux autres paramètres du système.
3. Montrez que l'on obtient une équation différentielle de la forme :

$$m\ddot{x}_p + f\dot{x}_p + kx_p = F$$

Dans tout ce qui suit, l'équation différentielle du mouvement utilisée sera celle-ci !

4. Déduisez de ce qui précède une représentation d'état de la forme :

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \quad (9.5)$$
$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \quad (9.6)$$

avec $y = x_p$
Applications numériques :
$m = 2, \quad f = 9, \quad k = 9$

Analyse

1. Le système est-il stable en boucle ouverte ?
2. Le système est-il commandable ?
3. Le système est-il observable ?

Commande

1. Donnez la valeur numérique des pôles en boucle ouverte.
2. Au vu des pôles en boucle ouverte, de votre expérience et du système proposez des pôles en boucle fermée.
3. Justifiez-les en 4 lignes maximum (+ un dessin si besoin).
4. Calculer un régulateur d'état qui permet d'obtenir la dynamique que vous vous êtes fixée en boucle fermée.
5. Par le calcul de la fonction de transfert, vérifiez que la dynamique souhaitée est bien obtenue.
6. A l'aide de cette fonction de transfert, calculer la matrice (ici un réel) de préfiltre.
7. Proposez une autre méthode pour obtenir une erreur statique nulle.

Observation

1. Donnez le schéma-bloc du système sous sa forme de représentation d'état avec son observateur de Luenberger complet.
2. Proposez et justifiez les pôles de l'observateur.
3. Calculer, par la méthode de votre choix, la matrice de "retour" $\mathbf{G}$

Obtention de l'erreur statique nulle

1. Donnez le schéma complet du système incluant :
   — le système lui-même,
   — un intégrateur pur pour obtenir une erreur statique nulle.
   On considérera que la matrice d'anticipation est nulle.
2. Déduisez de ce qui précède une représentation d'état de la forme :

$$\dot{\underline{z}} = \mathbf{A}_a\underline{z} + \mathbf{B}_a y_c \quad (9.7)$$
$$y = \mathbf{C}_a\underline{z} + \mathbf{D}_a y_c \quad (9.8)$$

   avec : $\underline{z} = \begin{bmatrix} \underline{x} \\ \nu \end{bmatrix}$
   où ν représenta la sortie de l'intégrateur pur.
3. calculez un nouveau correcteur tel que ce nouveau système présente à peu prés les mêmes caractéristiques que celle que vous avez précédemment choisies.

Examen final d'automatique avancée

Durée : 2 heures.
Tous documents autorisés.

1 Premier problème : Commande des systèmes non commandables

Un système linéaire continu est donné par sa représentation d'état :

$$\dot{\underline{x}} = \mathbf{A}\underline{x} + \mathbf{B}\underline{u} \tag{9.9}$$
$$\underline{y} = \mathbf{C}\underline{x} + \mathbf{D}\underline{u} \tag{9.10}$$

avec : $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 1 & -2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$, $\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\mathbf{D} = 0$

1.1 Analyse

1. Montrer que le système est instable en boucle ouverte.
   Instable : Critère de Routh
2. Montrer que le système est non commandable.

$$Q_s = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 3 \\ -1 & 3 & -5 \end{bmatrix} \qquad \det Q_s = 0 \Longrightarrow \textit{Non commandable}$$

3. Le système est-il observable ?

$$Q_b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \det Q_s = 1 \Longrightarrow \textit{Observable}$$

1.2 Réduction du modèle

Dans cette partie on "enlève" les parties non commandables et/ou non observables

1. Déterminer la fonction de transfert du système $H(p) = \frac{Y(p)}{U(p)}$.

$$H(p) = \frac{2\,p + 3}{(p-1)\,(2+p)}$$

2. A partir de la fonction de transfert $H(p)$ déduisez-en une représentation d'état sous forme canonique de commandabilité.

$$\dot{\underline{x_r}} = \mathbf{A_r}\underline{x_r} + \mathbf{B_r}\underline{u} \tag{9.11}$$
$$\underline{y} = \mathbf{C_r}\underline{x_r} + \mathbf{D_r}\underline{u} \tag{9.12}$$

avec :

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{D} = 0$$

3. Ce nouveau système est-il commandable et observable ?
   oui, car il provient de la fonction de transfert. Pôles en -1 et -2. Commandable

$$Q_s = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \qquad \det Q_s = 0 \Longrightarrow \textit{Commandable}$$

   Observable ?

$$Q_b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \det Q_s = 1 \Longrightarrow \textit{Observable}$$

4. Vérifier que le pôle instable est toujours présent.
   Pôles en 1 et -2

1.3 Commande du modèle réduit

1. Calculer le régulateur $\mathbf{L}$ tel que les pôles du système en boucle fermée soient :

$$p_{s_{1,2}} = -3 \pm 3i$$

$$Q_s = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{donc} \quad Q_s^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{et} \quad q_s^T = [1, 0]$$

$$P(p) = \left(p + 3 - 3\sqrt{-1}\right)\left(p + 3 + 3\sqrt{-1}\right) = p^2 + 6\,p + 18$$

$$\mathbf{L} = [20, 5]$$

2. Expliquez brièvement le comportement en boucle fermée que l'on aura avec ces pôles.
   Sans doute pas celui voulu à cause du zéro de la fonction de transfert.
3. Calculer la matrice (ici un réel) de préfiltre $\mathbf{S}$.

$$\mathbf{S} = -\Big(\mathbf{C}(\mathbf{A} - \mathbf{BL})^{-1}\mathbf{B}\Big)^{-1} = \frac{-1}{18}$$

1.4 Observation

1. Donnez le schéma-bloc du système sous sa forme de représentation d'état avec son observateur de Luenberger complet.
2. Calculer, par la méthode de votre choix, la matrice de "retour" $\mathbf{G}$, telle que les pôles de l'observateur soient placés en

$$p_{o_{1,2}} = -10 \pm 10i$$

$$Q_b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{donc} \quad Q_b^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{et} \quad q_b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$P(p) = \left(p + 10 - 10\sqrt{-1}\right)\left(p + 10 + 10\sqrt{-1}\right) = p^2 + 20\,p + 200$$

$$\mathbf{G} = \begin{bmatrix} 38 \\ 183 \end{bmatrix}$$

1.5 Erreur statique nulle

On se propose d'ajouter un intégrateur pur tel que l'erreur statique devienne nulle, y compris en présence de perturbations ou d'erreurs de modélisation.

1. Le schéma-bloc du système est donné en figure 9.4. Déterminer la représentation d'état du système augmenté :

$$\underline{\dot{x_a}} = \mathbf{A_a}\underline{x_a} + \mathbf{B_a}\underline{y_c} \quad (9.13)$$
$$\underline{y} = \mathbf{C_a}\underline{x_a} + \mathbf{D_a}\underline{y_c} \quad (9.14)$$

avec $\underline{x_a} = \begin{bmatrix} \underline{x} \\ \eta \end{bmatrix}$

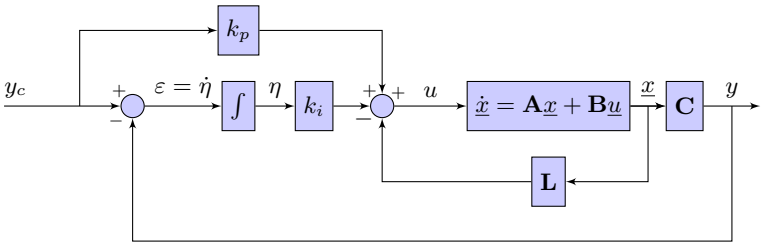


FIGURE 9.4 – Schéma d'une commande intégrale.

$$\mathbf{A_a} = \begin{bmatrix} 0 & 1 & 0 \\ 2 - l0 & -1 - l1 & ki \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{B_a} = \begin{bmatrix} 0 \\ kp \\ 1 \end{bmatrix}, \quad \mathbf{C_a} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{D_a} = 0$$

2. En posant $\mathbf{L} = \begin{bmatrix} l_0 & l_1 \end{bmatrix}$, calculer l_0, l_1 et k_i tels que les pôles du système soient placés en :

$$p_{s_{1,2,3}} = -3 \quad \text{et} \quad -3 \pm 3i$$

par identification :

$$\det \begin{bmatrix} p & -1 & 0 \\ -2 + l0 & 1 + l1 + p & -ki \\ 1 & 0 & p - 1 \end{bmatrix} = (p+3)\,(p+3-3\,i)\,(p+3+3i)$$

$$p^3 + l1\,p^2 + (l0 - 3 - l1)\,p + 2 - l0 + ki = p^3 + 9\,p^2 + 36\,p + 54$$
$$\Longrightarrow \{ki = 100, l1 = 9, l0 = 48\}$$

1.6 Observateur numérique

L'observateur du système précédent est en fait implanté sous forme numérique.

1. transformer le système

$$\underline{\dot{x_r}} = \mathbf{A_r}\underline{x} + \mathbf{B_r}\underline{u} \quad (9.15)$$
$$\underline{y} = \mathbf{C_r}\underline{x_r} + \mathbf{D_r}\underline{u} \quad (9.16)$$

en un système échantillonné de la forme :

$$\underline{x}_{k+1} = \mathbf{\Phi}\underline{x}_k + \mathbf{\Gamma}u_k. \quad (9.17)$$
$$\underline{y}_k = \mathbf{C}\underline{x}_k \quad (9.18)$$

sachant que en prenant $T_e = 1$s, $\mathbf{\Phi} = \begin{bmatrix} 1.86 & 0.86 \\ 1.72 & 0.99 \end{bmatrix}$

$$\mathbf{\Gamma} = \mathbf{A}^{-1}(\mathbf{\Phi} - \mathbf{I})\mathbf{B} = \begin{bmatrix} 0.43 \\ 0.86 \end{bmatrix}$$

2. Déterminer un observateur complet pour ce système numérique. Les pôles de cet observateur complet seront placés en 0.

$$Q_b = \begin{bmatrix} 1 & 0 \\ 1.86 & 0.86 \end{bmatrix} \quad \text{donc} \quad Q_b^{-1} = \begin{bmatrix} 1.0 & 0 \\ -2.16 & 1.16 \end{bmatrix} \quad \text{et} \quad q_b = \begin{bmatrix} 0 \\ 1.16 \end{bmatrix}$$

$$P(z) = z^2$$

$$\mathbf{G} = \begin{bmatrix} -2.32 \\ 3.48 \end{bmatrix}$$

Chapitre 10

Travaux dirigés

1 TD1 - Représentation d'état

1.1 A partir d'un système

Soit un moteur à courant continu commandé par l'inducteur.

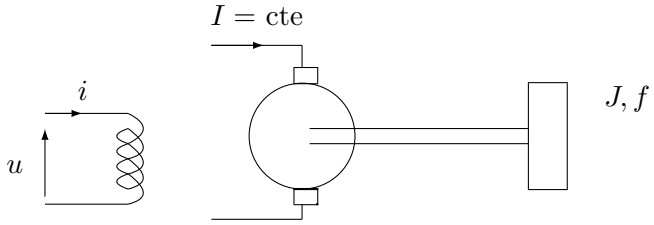


FIGURE 10.1 – Moteur à courant continu commandé par l'inducteur

— commande : u
— sortie : ω

Les équations fondamentales de ce système sont données ci-après :

$$U = Ri + L\frac{di}{dt} \tag{10.1}$$

$$J\frac{d\omega}{dt} = -f\omega + \gamma \tag{10.2}$$

$$\gamma = ki \tag{10.3}$$

$$\tag{10.4}$$

1. Déduisez-en l'équation différentielle qui régit la vitesse du rotor en fonction de la tension. Déduisez-en la fonction de transfert du système, puis la réponse impulsionnelle du système.
2. Déduisez-en un schéma-bloc du système.
3. En posant :

$$x_1 = \omega, \qquad x_2 = \frac{d\omega}{dt} = \dot{\omega}, \tag{10.5}$$

donnez l'équation différentielle du système sous la forme matricielle suivante :
Equation d'état :

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} . & . \\ . & . \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} . \\ . \end{bmatrix} u \tag{10.6}$$

Equation de sortie :

$$\omega = \begin{bmatrix} . & . \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{10.7}$$

1.2 A partir d'un schéma bloc

Soit le système représenté figure 10.2

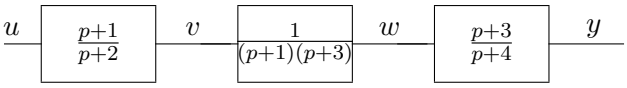


FIGURE 10.2 – Système étudié

Montrez que ce système peut être mis sous la forme suivante (fig. 10.3)

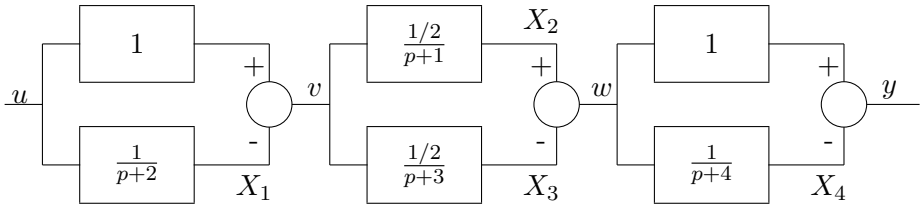


FIGURE 10.3 – Nouvelle forme du système

Mise en équation du système

Mettre le système en équation avec les trois méthodes proposées.

1. la représentation d'état
2. la fonction de transfert
3. l'équation différentielle

Comparez l'information présente dans les trois mises en équation.

1.3 Pour l'entraînement

Etudiez le système suivant (figure 10.4)

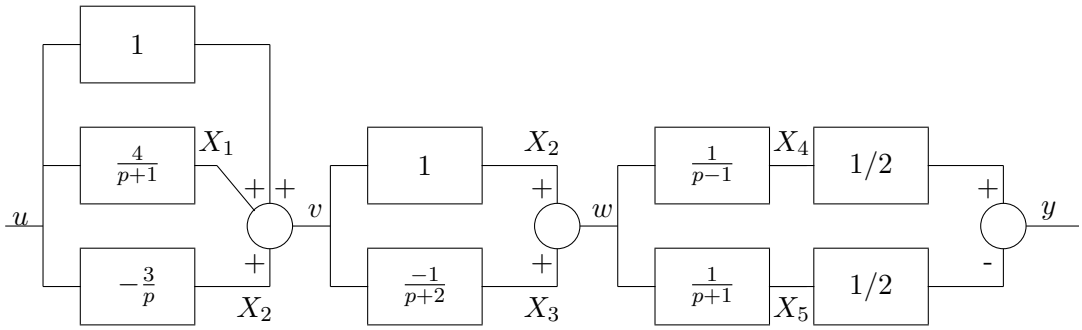


FIGURE 10.4 – Système étudié

2 Commandabilité et observabilité

soit le système représenté figure 10.5

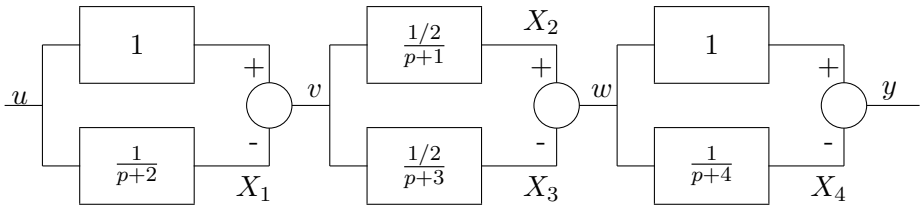


FIGURE 10.5 – Système étudié

2.1 Mise en équation du système

Mettre le système en équation avec les trois méthodes proposées.

1. la représentation d'état
2. la fonction de transfert
3. l'équation différentielle

Comparez l'information présente dans les trois mises en équation.

2.2 Mise sous la forme canonique de Jordan

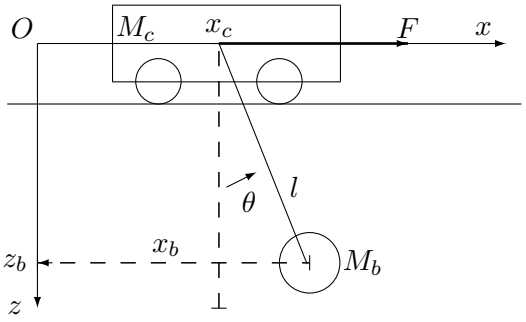
1. Résoudre l'équation
$$\det(p\mathbf{I} - \mathbf{A}) = 0$$
2. Calculez le vecteur propre associé à chaque valeur propre
3. Déterminez la matrice de transformation $\mathbf{T}$
4. Vérifiez que $\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{\Lambda}$
5. Calculez les vecteurs $\widehat{\mathbf{C}} = \mathbf{C}\mathbf{T}$ et $\widehat{\mathbf{B}} = \mathbf{T}^{-1}\mathbf{B}$
6. Donnez la représentation d'état complète du système sous la forme canonique de Jordan.

2.3 Commandabilité et observabilité

1. Vérifiez la commandabilité et l'observabilité du système sur la représentation initiale.
2. Vérifiez la commandabilité et l'observabilité sur la représentation sous la forme canonique de Jordan.
3. Tracez le graphe de fluence du système sous la forme canonique de Jordan.

3 TD asservissement dans l'espace d'état d'un pont roulant

Le rôle d'un pont roulant est de s'emparer d'une charge à un endroit déterminé, de la déplacer sur une distance déterminée, puis de la déposer à un autre endroit donné. Le problème qui surgit lors de l'automatisation de ce processus est que la charge entre en oscillation à la mise en marche et au freinage du chariot. Ces oscillations sont nuisibles au fonctionnement puisqu'elles retardent la prise et le dépôt de la charge.
Etant donné que les performances d'un pont roulant résident principalement dans la vitesse à laquelle il est capable d'exécuter un cycle de travail, il est nécessaire pour améliorer ses performances d'empêcher la naissance de ces oscillations ou, tout du moins de les réduire à une valeur acceptable.
Nous supposerons que le problème est plan. La position instantanée de la benne peut donc être décrite sans ambiguïté à l'aide de deux coordonnées, qui peuvent être, par exemple, l'abscisse du chariot x_c et soit l'abscisse de la benne x_b soit l'angle du filin avec la verticale θ.



avec :

x_c	:	position du chariot
M_c	:	masse du chariot
F	:	force de traction exercée sur le chariot
x_b, z_b	:	position de la benne
M_b	:	masse de la benne
l	:	longueur du filin
θ	:	angle du filin

FIGURE 10.6 – Schéma du chariot et de la benne

3.1 Mise en équation

Mettez en équations ce problème. Linéarisez ensuite dans l'hypothèse, généralement vérifié dans la réalité, d'excusions angulaires faibles de la benne (θ petit).

3.2 Représentation d'état

Etablir les équations d'état de ce système en choisissant les variables suivantes :
x_1 : position du chariot
x_2 : vitesse du chariot
x_3 : angle du filin
x_4 : vitesse angulaire du filin

L'équation de sortie sera, pour l'instant, ignorée.
Note : Je vous suggère d'utiliser les notations suivantes :

$$\alpha = \frac{M_b}{M_c} g,\ \beta = -\left(1 + \frac{M_b}{M_c}\right)\frac{g}{l},\ b_2 = \frac{1}{M_c},\ b_4 = -\frac{1}{M_c l}$$

3.3 Simulation Matlab

Simulez le système sur un calculateur. Relevez et interprétez les courbes de $x_c,\ \dot{x_c},\ \theta$ en réponse à un échelon de force appliquée $F = 1\text{kN}$.
Valeurs numériques :

$$M_c = 1000\text{kg},\ M_b = 4000\text{kg},\ l = 10\text{m},\ g = 9,81\text{ms}^{-2}$$

3.4 Commandabilité et observabilité

Déterminez la commandabilité en calculant la matrice de commandabilité puis interprétez physiquement le résultat obtenu.
Déterminez l'observabilité du système en se plaçant dans les deux hypothèses suivantes :

1. mesure de la position du chariot seule,
2. mesure de l'angle du filin seul.

3.5 Commande par retour d'état

Effectuez la synthèse de l'asservissement de ce système par retour d'état par la méthode du placement de pôles en respectant le cahier des charges suivant :

1. en régime transitoire,
   (a) assez bien amorti, dépassement faible,
   (b) rapide, le temps d'établissement à 2% ne doit pas dépasser $t_{2\%} = 20$s,
2. en régime permanent, erreur statique nulle

Indications :

1. Commencez par déterminer les pôles du système à asservir,
2. Choisir les pôles du système bouclé en fonction du cahier des charges de la façon suivante :
   (a) 1 paire de pôles complexes dominants,
   (b) 1 pôle réel unique d'ordre ? ? en $p = -1$.
3. Calculez les paramètres du régulateur d'état.
4. Calculez la matrice de préfiltre.
5. Tracez le schéma fonctionnel du système bouclé, en représentant le système à asservir par un seul bloc.

3.6 Observateur complet

Déterminez un observateur de Luenberger complet, dans le cas de la mesure de la position du chariot, ayant une valeur propre unique d'ordre ? ?, en $p = -2$. Justifiez ce choix. Calculez les paramètres de cet observateur numériquement et tracez son schéma fonctionnel.

3.7 Observateur réduit

Déterminez un observateur de Luenberger réduit, dans le cas de la mesure de la position du chariot, ayant une valeur propre unique d'ordre ? ?, en $p = -2$. Justifiez ce choix. Calculez les paramètres de cet observateur numériquement et tracez son schéma fonctionnel.

4 Commande du tangage d'un avion

Les équations du mouvement d'un avion sont particulièrement complexes puis qu'il s'agit de 6 équations différentielles non linéaires couplées. Moyennant certaines hypothèses, celles-ci peuvent être réduites à des équations différentielles découplées et linéarisées. Dans cet exercice on se propose d'étudier un pilote automatique contrôlant le tangage de l'avion. Attention en cas d'erreur, les conséquences seraient très graves !
La complexité des calculs est telle que l'utilisation du logiciel Matlab s'impose.

4.1 Modélisation

Les équation du mouvement de tangage sont les suivantes :

$$\dot{\alpha} = \mu\Omega\left[-(C_L + C_D)\alpha + (1/\mu - C_L)q - (C_w \sin(\gamma_e))\theta + C_L\right] \quad (10.8)$$

$$\dot{q} = \frac{\mu\Omega}{2i_{yy}}\left\{\left[C_M - \nu(C_L + C_D)\right]\alpha + \left[C_M + \sigma C_M(1 - \mu C_L)\right]q + (\nu C_W \sin(\gamma_e))\delta_e\right\} \quad (10.9)$$

$$\dot{\theta} = \Omega q \quad (10.10)$$

Définition de quelques variables :
α = angle d'attaque
q = variation d'angle de tangage
θ = angle de tangage
δ_e = angle des volets de l'avion
ρ_e = densité de l'air ambiant
S = surface de l'aile
m = masse de l'avion
U = vitesse de l'avion
Déduisez des équations précédentes la forme générale de la représentation d'état du système.
Pour la suite, les valeurs numériques suivantes seront utilisées.

$$\dot{\alpha} = -0,313\alpha + 56,7q + 0.232\delta_e \quad (10.11)$$

$$\dot{q} = -0,0139\alpha - 0,426q + 0.0203\delta_e \quad (10.12)$$

$$\dot{\theta} = 56,7q \quad (10.13)$$

4.2 Cahier des charges

Les critères de confort des passagers font apparaître le cahier des charges suivant :
dépassement de moins de 10%
temps de montée de moins de 2 secondes
régime permanent en moins de 10 secondes
erreur statique de moins de 2%

4.3 Etude en boucle ouverte - fonction de transfert

1. Déterminez la fonction de transfert du système.
   entrée : δ_e, sortie : θ
2. le système est-il stable en boucle ouverte.
3. le système est-il contrôlable, est-il observable ?

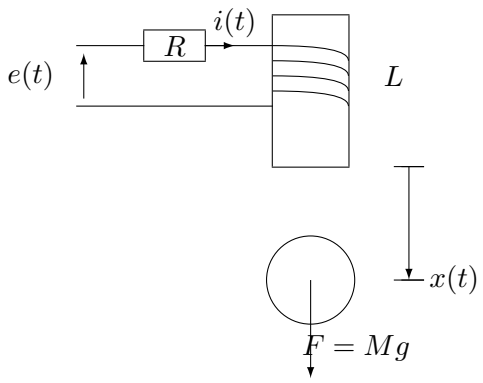


FIGURE 10.7 – Suspension magnétique

4.4 Commande du système

Placement de pôles

1. Déterminez les pôles complexes conjugués répondant au cahier des charges. Les autres pôles seront choisis tous identiques et négligeables devant les pôles dominants.
2. Mettez en place ce correcteur et vérifiez que le cahier des charges est bien respecté.

Commande optimale

1. Calculez le correcteur optimal avec les matrices de pondération suivantes : $\mathbf{R} = 1$ et $\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Contrôle numérique

Après avoir déterminé la représentation d'état échantillonnée du système, reprenez les deux méthodes de correction du système précédentes et appliquez-les au cas discret.

4.5 Synthèse d'observateurs

Synthétisez un observateur complet de Luenberger sur le système précédent.

5 Asservissement d'une suspension magnétique

Principe : L'électro-aimant d'inductance L variable est parcouru par un courant $i(t)$ délivré par un générateur de tension $e(t)$ et de résistance interne R. Cet électro-aimant exerce une force d'attraction sur la bille, celle ci se rapproche de l'électro-aimant et donc l'inductance augmente. La loi de variation de L est :

$$L(x) = \frac{L_0}{x}$$

5.1 Modélisation

1. A l'aide des équations de Lagrange, modélisez le système. Les coordonnées généralisées choisies seront :
   q_1 = la charge électrique , $\dot{q}_1$ = le courant dans l'inductance,
   q_2 = la position de la bille, $\dot{q}_2$ = la vitesse de la bille

2. Ce système étant non linéaire, la première étape consiste donc à le linéariser autour d'un point de fonctionnement par un développement limité au premier ordre ($a = a_0 + \hat{a}$).
   Ainsi, si :

$$\ddot{q}_1 = F_1(q_1, q_2, \dot{q}_1, \dot{q}_2, e)$$

   alors, au point de fonctionnement

$$\ddot{\hat{q}_1} = \left.\frac{\partial F_1}{\partial q_1}\right|_E \hat{q_1} + \left.\frac{\partial F_1}{\partial q_2}\right|_E \hat{q_2} + \left.\frac{\partial F_1}{\partial \dot{q}_1}\right|_E \dot{\hat{q}_1} + \left.\frac{\partial F_1}{\partial \dot{q}_2}\right|_E \dot{\hat{q}_2} + \left.\frac{\partial F_1}{\partial e}\right|_E \hat{e}$$

3. En supposant qu'au point de fonctionnement la vitesse de la bille est nulle et que la tension moyenne appliquée est $e_0 = Ri_0$, montrez que la représentation d'état obtenue est :

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{Rq_{2_0}}{L} & 0 & \frac{\dot{q}_{1_0}}{q_{2_0}} \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{L\dot{q}_{1_0}}{{q_{2_0}}^2 M} & \frac{L\dot{q}_{1_0}^2}{{q_{2_0}}^3 M} & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ \frac{q_{2_0}}{L} \\ 0 \\ 0 \end{bmatrix}$$

4. Déterminez l'équation différentielle reliant la position et ses dérivées successives à la tension d'entrée.
5. Déterminez la fonction de transfert du système.

5.2 Analyse du système

1. Le système est-il stable ?
2. Le système est-il observable ? Est-il commandable ?

5.3 Correction du système

1. Donnez la représentation d'état du système sous forme diagonale, sous la forme canonique de commandabilité, sous la forme canonique d'observabilité.
2. Afin de stabiliser le système, proposez un choix de pôles en boucle fermée. Calculez le correcteur et la matrice de préfiltre.

5.4 Observation du système

1. Le système étant non observable, le correcteur précédent est inapplicable, proposez alors une autre solution.

5.5 Réduction du système

1. A partir de la fonction de transfert, déterminer une représentation d'état sous la forme canonique de commandabilité du système.
2. Le modèle ainsi obtenu est-il stable, commandable, observable ?

5.6 Synthèse de la commande

Applications numériques : $M = 0.1kg$, $i_0 = 1A$, $x_0 = 0,03m$, $; L = 1H$, $R = 0.1\Omega$.

1. Avec les applications numériques précédentes, le système présente les pôles suivants :

$$+10.35, -5.18 - 8.97 * I, -5.18 + 8.97 * I$$

2. Choisissez les pôles en boucle fermée, calculez le correcteur.

5.7 Synthèse de l'observateur

1. Synthétisez un observateur complet pour ce système.

6 Commande numérique d'un AMF

Le système se compose d'un fil en Alliage à Mémoire de Forme (AMF) qui présente la propriété de se contracter lorsqu'il est chauffé. Ce fil de $150\mu m$ de diamètre est relié mécaniquement à un ressort de traction. Le principe de chauffage est l'effet Joule.
Les modèles, simplifiés à l'extrême sont donnés ci-après.

$$mC_p\dot{T} + hS(T - T_a) = Ri^2$$

$$\dot{x} + bx = aT$$

	T_a	=	300 K	température ambiante
	h	=	38 J m^{-2}	coefficient de convection
	ϕ	=	150.10^{-6} m	diamètre du fil
	L	=	0.10 m	longueur du fil
avec :	R	=	5Ω	résistance électrique
	ρ	=	6450 kg/m^{-3}	densité volumique
	C_p	=	1886 $K.kg^{-1}$	chaleur spécifique
	a	=	$5.10^{-6}m.K^{-1}$	
	b	=	0.1 s^{-1}	

6.1 Modélisation et analyse

1. Donnez la représentation d'état du système.
2. Déduisez-en la représentation d'état échantillonnée du système.
3. Le système est-il stable, commandable, observable ?

6.2 Première commande

1. Au vu des pôles en boucle ouverte du système, proposez des pôles en boucle fermée.
2. Calculez le correcteur pour obtenir le comportement dynamique souhaité, ainsi que la matrice de préfiltre.
3. Dans quelles conditions, la matrice de préfiltre garantie-t-elle une erreur statique nulle ?

6.3 Amélioration de la commande

Afin de garantir une erreur statique nulle, y compris en cas de variation de la température ambiante, introduisons un intégrateur pur dans le système en amont du point d'application des perturbations.

1. Déterminez la nouvelle représentation d'état en z du système.
2. Calculez un correcteur pour ce nouveau système, présentant approximativement les mêmes caractéristiques que le précédent.

6.4 Synthèse de l'observateur

1. Calculez un observateur de Luenberger complet.
2. Après avoir déterminé les variables non accessibles à la mesure, déduisez-en un observateur d'ordre réduit.

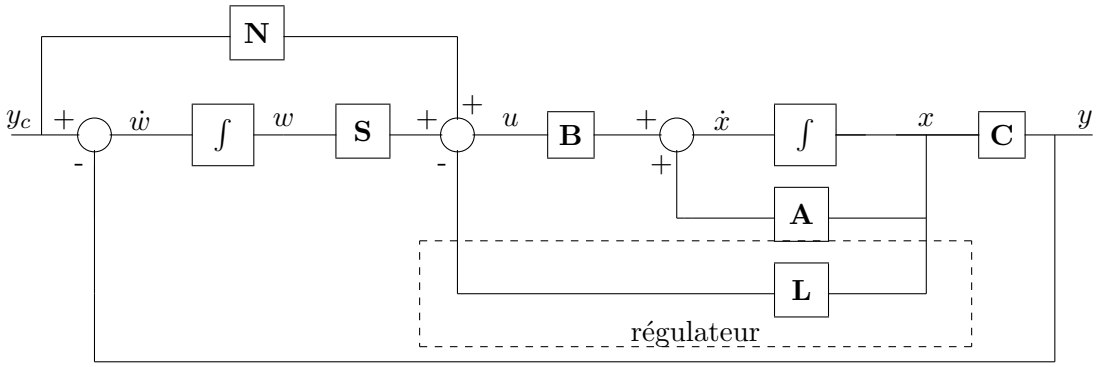


FIGURE 10.8 – Introduction d'un intégrateur pur.

6.5 Création du code C

1. Donnez la fonction de commande du système précédent dont la définition est donnée ci-après. Nous supposerons que la mise à jour des tableaux est faite par ailleurs. Ainsi, à l'instant k, la sortie du système de l'instant $k-1$ est stockée dans sortie[0], $y(k-2)$ est dans sortie[1] ...

```
float regulateur(float* commande, float* sortie)
{



return(commande)
}
```

6.6 Prise en compte d'un retard pur

1. La mise en place du code précédent montre que le temps de calcul est grand par rapport à la période d'échantillonnage, l'hypothèse de simultanéité temporelle de l'échantillonnage et de l'envoi de la commande est donc fausse.

   Afin de prendre en compte ce phénomène, introduisez un retard pur (z^{-1}) dans le système.

2. Reprenez la conception du correcteur et de l'observateur d'ordre réduit.

7 Bille sur un rail

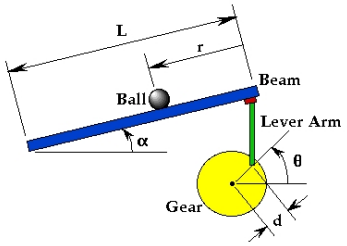


FIGURE 10.9 – Bille sur un rail

7.1 Modélisation

M	masse de la bille	0.11kg
R	rayon de la bille	0.015 m
d	rayon du réducteur	0.03 m
g	accélération terrestre	9.8 m/s^2
L	longueur du rail	1.0 m
J	moment d'inertie de la bille	$1e-5$ kg.m^2
r	position de la bille	m
α	angle du rail	rad.
θ	angle du servomoteur	rad.

1. Avec les hypothèses suivantes :
   — la commande u du système commande directement l'accélération su servomoteur $u = \ddot{\alpha}$,
   — l'angle α est faible, on considère que $\sin\alpha = 0$,
   montrez que la représentation d'état est de la forme :

$$\begin{pmatrix} \dot{r} \\ \ddot{r} \\ \dot{\alpha} \\ \ddot{\alpha} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-mg}{\frac{J}{R^2}+m} & 0 \\ 0 & & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} r \\ \dot{r} \\ \alpha \\ \dot{\alpha} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u \tag{10.14}$$

$$y = (1,0,0,0) \begin{pmatrix} r \\ \dot{r} \\ \alpha \\ \dot{\alpha} \end{pmatrix} \tag{10.15}$$

2. De la représentation d'état continue, déduisez la représentation d'état discrète.

7.2 Analyse

1. Le système est-il stable ?
2. Le système est-il observable ? Est-il commandable ?
3. Que faire ? Dans quelle mesure les hypothèses formulées sont-elles sources de problèmes.

7.3 Commande

1. Calculer un régulateur numérique pour le système précédent.

7.4 Observation

1. Calculer un observateur numérique pour le système précédent.

8 Pendule inverse : Traité avec Matlab

8.1 Modélisation

Description et valeurs numériques :

M	masse du chariot	0.5 kg
m	masse du pendule	0.5kg
b	coefficient de frottement du chariot	0.1N.m^{-1}.s
l	longueur du pendule	0.3 m
I	inertie du pendule	0.006 kg*m^2
F	force appliquée sur le chariot	N
x	position du chariot	m
θ	angle du pendule	rad.

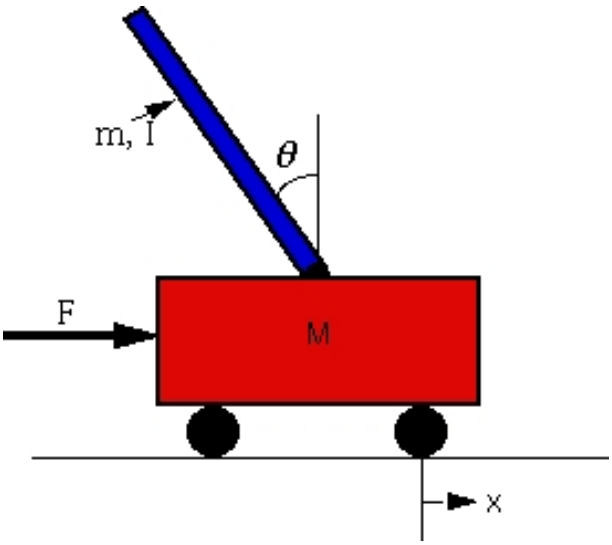


FIGURE 10.10 – Pendule inverse

1. Déterminez les équations différentielles qui régissent le mouvement des deux corps.

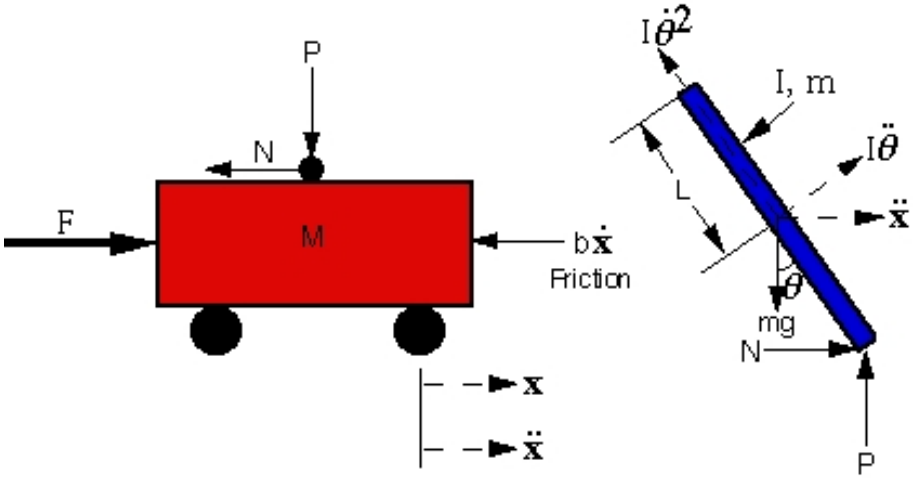


FIGURE 10.11 – Mise en équation pendule inverse

2. Après linéarisation des équations précédentes ($\theta \simeq \pi + \phi$ avec ϕ petit), déterminez la fonction de transfert du système.
3. Calculer la représentation d'état du système.
4. Vérifier que la représentation obtenue est :

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\Phi} \\ \ddot{\Phi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{-(I+ml^2)b}{I(M+m)+Mml^2} & \frac{m^2gl^2}{I(M+m)+Mml^2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{-mlb}{I(M+m)+Mml^2} & \frac{mgl(M+m)}{I(M+m)+Mml^2} & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \Phi \\ \dot{\Phi} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{I+ml^2}{I(M+m)+Mml^2} \\ 0 \\ \frac{ml}{I(M+m)+Mml^2} \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \Phi \\ \dot{\Phi} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} u$$

FIGURE 10.12 – Représentation d'état du pendule inverse

8.2 Analyse

1. Le système est-il stable ? Est-il observable ?
2. Quelles variables faut-il mesurer pour que le système soit observable.

8.3 Commande

1. Calculer un régulateur par la méthode du placement de pôles qui respecte le cahier des charges suivant :

   — Temps de réponse sur θ inférieur à 5 s.
   — Temps de montée de x inférieur à 1 s.
   — Dépassement de moins de 0.035 radians.
   — Erreur statique nulle.
2. Comparer les réponses temporelles obtenues en boucle fermée par rapport à celle obtenues par une commande optimale obtenue avec :

$$\mathbf{R} = 5000, \quad \mathbf{Q} = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

9 Exercices divers

9.1 Placement de pôles

Soit la fonction de transfert :

$$F(p) = \frac{1+2p}{(p-1)((p+1)^2+2)}$$

1. Déterminez directement la représentation d'état de cette fonction de transfert sous la forme de commandabilité.
2. On désire que le système en boucle fermée présente un pôle triple en -1, calculez le correcteur.
3. Par ailleurs, on désire que le système présente une erreur nulle pour une entrée en échelon, déterminer la matrice de préfiltre.

9.2 Passage de l'équation récurrente - représentation d'état - fonction de transfert

soit le système décrit par les équation récurrentes suivantes :

$$\begin{aligned} x_1(k+1) &= x_2(k), && (10.16)\\ x_2(k+1) &= x_3(k), && (10.17)\\ x_3(k+1) &= -4x_2(k) - 5x_3(k) + u(k), && (10.18)\\ y(k) &= 4x_1(k) + 4x_2(k) + x_3(k). && (10.19) \end{aligned}$$

1. Déterminez directement la représentation d'état du système décrit par les équation récurrentes précédentes.
2. Calculez la fonction de transfert du système.

Première partie

Annexes

Annexe A

Quelques publications originales

A New Approach to Linear Filtering and Prediction Problems[1]

R. E. KALMAN
Research Institute for Advanced Study,[2] Baltimore, Md.

The classical filtering and prediction problem is re-examined using the Bode-Shannon representation of random processes and the "state transition" method of analysis of dynamic systems. New results are:

(1) The formulation and methods of solution of the problem apply without modification to stationary and nonstationary statistics and to growing-memory and infinite-memory filters.

(2) A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. From the solution of this equation the coefficients of the difference (or differential) equation of the optimal linear filter are obtained without further calculations.

(3) The filtering problem is shown to be the dual of the noise-free regulator problem.

The new method developed here is applied to two well-known problems, confirming and extending earlier results.

The discussion is largely self-contained and proceeds from first principles; basic concepts of the theory of random processes are reviewed in the Appendix.

Introduction

AN IMPORTANT class of theoretical and practical problems in communication and control is of a statistical nature. Such problems are: (i) Prediction of random signals; (ii) separation of random signals from random noise; (iii) detection of signals of known form (pulses, sinusoids) in the presence of random noise.

In his pioneering work, Wiener [1][3] showed that problems (i) and (ii) lead to the so-called Wiener-Hopf integral equation; he also gave a method (spectral factorization) for the solution of this integral equation in the practically important special case of stationary statistics and rational spectra.

Many extensions and generalizations followed Wiener's basic work. Zadeh and Ragazzini solved the finite-memory case [2]. Concurrently and independently of Bode and Shannon [3], they also gave a simplified method [2] of solution. Booton discussed the nonstationary Wiener-Hopf equation [4]. These results are now in standard texts [5-6]. A somewhat different approach along these main lines has been given recently by Darlington [7]. For extensions to sampled signals, see, e.g., Franklin [8], Lees [9]. Another approach based on the eigenfunctions of the Wiener-Hopf equation (which applies also to nonstationary problems whereas the preceding methods in general don't), has been pioneered by Davis [10] and applied by many others, e.g., Shinbrot [11], Blum [12], Pugachev [13], Solodovnikov [14].

In all these works, the objective is to obtain the specification of a linear dynamic system (Wiener filter) which accomplishes the prediction, separation, or detection of a random signal.[4]

Present methods for solving the Wiener problem are subject to a number of limitations which seriously curtail their practical usefulness:

(1) The optimal filter is specified by its impulse response. It is not a simple task to synthesize the filter from such data.

(2) Numerical determination of the optimal impulse response is often quite involved and poorly suited to machine computation. The situation gets rapidly worse with increasing complexity of the problem.

(3) Important generalizations (e.g., growing-memory filters, nonstationary prediction) require new derivations, frequently of considerable difficulty to the nonspecialist.

(4) The mathematics of the derivations are not transparent. Fundamental assumptions and their consequences tend to be obscured.

This paper introduces a new look at this whole assemblage of problems, sidestepping the difficulties just mentioned. The following are the highlights of the paper:

(5) *Optimal Estimates and Orthogonal Projections.* The Wiener problem is approached from the point of view of conditional distributions and expectations. In this way, basic facts of the Wiener theory are quickly obtained; the scope of the results and the fundamental assumptions appear clearly. It is seen that all statistical calculations and results are based on first and second order averages; no other statistical data are needed. Thus difficulty (4) is eliminated. This method is well known in probability theory (see pp. 75–78 and 148–155 of Doob [15] and pp. 455–464 of Loève [16]) but has not yet been used extensively in engineering.

(6) *Models for Random Processes.* Following, in particular, Bode and Shannon [3], arbitrary random signals are represented (up to second order average statistical properties) as the output of a linear dynamic system excited by independent or uncorrelated random signals ("white noise"). This is a standard trick in the engineering applications of the Wiener theory [2–7]. The approach taken here differs from the conventional one only in the way in which linear dynamic systems are described. We shall emphasize the concepts of *state* and *state transition;* in other words, linear systems will be specified by systems of first-order difference (or differential) equations. This point of view is

[1] This research was supported in part by the U. S. Air Force Office of Scientific Research under Contract AF 49 (638)-382.

[2] 7212 Bellona Ave.

[3] Numbers in brackets designate References at end of paper.

[4] Of course, in general these tasks may be done better by nonlinear filters. At present, however, little or nothing is known about how to obtain (both theoretically and practically) these nonlinear filters.

Contributed by the Instruments and Regulators Division and presented at the Instruments and Regulators Conference, March 29– Apri1 2, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, February 24, 1959. Paper No. 59—IRD-11.

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

natural and also necessary in order to take advantage of the simplifications mentioned under (5).

(7) *Solution of the Wiener Problem.* With the state-transition method, a single derivation covers a large variety of problems: growing and infinite memory filters, stationary and nonstationary statistics, etc.; difficulty (3) disappears. Having guessed the "state" of the estimation (i.e., filtering or prediction) problem correctly, one is led to a nonlinear difference (or differential) equation for the covariance matrix of the optimal estimation error. This is vaguely analogous to the Wiener-Hopf equation. Solution of the equation for the covariance matrix starts at the time t_0 when the first observation is taken; at each later time t the solution of the equation represents the covariance of the optimal prediction error given observations in the interval (t_0, t). From the covariance matrix at time t we obtain at once, without further calculations, the coefficients (in general, time-varying) characterizing the optimal linear filter.

(8) *The Dual Problem.* The new formulation of the Wiener problem brings it into contact with the growing new theory of control systems based on the "state" point of view [17–24]. It turns out, *surprisingly*, that the Wiener problem is the *dual* of the noise-free optimal regulator problem, which has been solved previously by the author, using the state-transition method to great advantage [18, 23, 24]. The mathematical background of the two problems is identical—this has been suspected all along, but until now the analogies have never been made explicit.

(9) *Applications.* The power of the new method is most apparent in theoretical investigations and in numerical answers to complex practical problems. In the latter case, it is best to resort to machine computation. Examples of this type will be discussed later. To provide some feel for applications, two standard examples from nonstationary prediction are included; in these cases the solution of the nonlinear difference equation mentioned under (7) above can be obtained even in closed form.

For easy reference, the main results are displayed in the form of theorems. Only Theorems 3 and 4 are original. The next section and the Appendix serve mainly to review well-known material in a form suitable for the present purposes.

Notation Conventions

Throughout the paper, we shall deal mainly with *discrete* (or *sampled)* dynamic systems; in other words, signals will be observed at equally spaced points in time *(sampling instants).* By suitable choice of the time scale, the constant intervals between successive sampling instants *(sampling periods)* may be chosen as unity. Thus variables referring to time, such as t, t_0, τ, T will always be integers. The restriction to discrete dynamic systems is not at all essential (at least from the engineering point of view); by using the discreteness, however, we can keep the mathematics rigorous and yet elementary. Vectors will be denoted by small bold-face letters: **a**, **b**, ..., **u**, **x**, **y**, ... A *vector* or more precisely an *n-vector* is a set of n numbers $x_1, \ldots x_n$; the x_i are the *co-ordinates* or *components* of the vector **x**.

Matrices will be denoted by capital bold-face letters: **A**, **B**, **Q**, **Φ**, **Ψ**, …; they are $m \times n$ arrays of elements a_{ij}, b_{ij}, q_{ij},... The *transpose* (interchanging rows and columns) of a matrix will be denoted by the prime. In manipulating formulas, it will be convenient to regard a vector as a matrix with a single column.

Using the conventional definition of matrix multiplication, we write the *scalar product* of two n-vectors **x**, **y** as

$$\mathbf{x'y} = \sum_{i=1}^{n} x_i y_i = \mathbf{y'x}$$

The scalar product is clearly a scalar, i.e., not a vector, quantity. Similarly, the quadratic form associated with the $n \times n$ matrix **Q** is,

$$\mathbf{x'Qx} = \sum_{i,j=1}^{n} x_i q_{ij} x_j$$

We define the expression $\mathbf{xy'}$ where $\mathbf{x'}$ is an m-vector and **y** is an n-vector to be the $m \times n$ matrix with elements $x_i y_j$.

We write $E(\mathbf{x}) = E\mathbf{x}$ for the expected value of the random vector **x** (see Appendix). It is usually convenient to omit the brackets after E. This does not result in confusion in simple cases since constants and the operator E commute. Thus $E\mathbf{xy'}$ = matrix with elements $E(x_i y_j)$; $E\mathbf{x}E\mathbf{y'}$ = matrix with elements $E(x_i)E(y_j)$.

For ease of reference, a list of the principal symbols used is given below.

Optimal Estimates

t	time in general, present time.
t_0	time at which observations start.
$x_1(t)$, $x_2(t)$	basic random variables.
$y(t)$	observed random variable.
$x_1^*(t_1\|t)$	optimal estimate of $x_1(t_1)$ given $y(t_0)$, …, $y(t)$.
L	loss function (non random function of its argument).
ε	estimation error (random variable).

Orthogonal Projections

$\mathcal{Y}(t)$	linear manifold generated by the random variables $y(t_0)$, …, $y(t)$.
$\bar{x}(t_1\|t)$	orthogonal projection of $x(t_1)$ on $\mathcal{Y}(t)$.
$\tilde{x}(t_1\|t)$	component of $x(t_1)$ orthogonal to $\mathcal{Y}(t)$.

Models for Random Processes

$\mathbf{\Phi}(t+1; t)$	transition matrix
$\mathbf{Q}(t)$	covariance of random excitation

Solution of the Wiener Problem

$\mathbf{x}(t)$	basic random variable.
$\mathbf{y}(t)$	observed random variable.
$\mathcal{Y}(t)$	linear manifold generated by $\mathbf{y}(t_0)$, …, $\mathbf{y}(t)$.
$\mathcal{Z}(t)$	linear manifold generated by $\mathbf{y}(t\|t-1)$.
$\mathbf{x}^*(t_1\|t)$	optimal estimate of $\mathbf{x}(t_1)$ given $\mathcal{Y}(t)$.
$\tilde{\mathbf{x}}(t_1\|t)$	error in optimal estimate of $\mathbf{x}(t_1)$ given $\mathcal{Y}(t)$.

Optimal Estimates

To have a concrete description or the type of problems to be studied, consider the following situation. We are given signal $x_1(t)$ and noise $x_2(t)$. Only the sum $y(t) = x_1(t) + x_2(t)$ can be observed. Suppose we have observed and know exactly the values of $y(t_0)$, ..., $y(t)$. What can we infer from this knowledge in regard to the (unobservable) value of the signal at $t = t_1$, where t_1 may be less than, equal to, or greater than t? If $t_1 < t$, this is a *data-smoothing (interpolation)* problem. If $t_1 = t$, this is called *filtering.* If $t_1 > t$, we have a *prediction* problem. Since our treatment will be general enough to include these and similar problems, we shall use hereafter the collective term *estimation.*

As was pointed out by Wiener [1], the natural setting of the estimation problem belongs to the realm of probability theory and statistics. Thus signal, noise, and their sum will be random variables, and consequently they may be regarded as random processes. From the probabilistic description of the random processes we can determine the probability with which a particular sample of the signal and noise will occur. For any given set of measured values $\eta(t_0)$, ..., $\eta(t)$ of the random variable $y(t)$ one can then also determine, in principle, the probability of simultaneous occurrence of various values $\xi_1(t)$ of the random variable $x_1(t_1)$. This is the conditional probability distribution function

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

$$Pr[x_1(t_1) \le \xi_1 | y(t_0) = \eta(t_0), \ldots, y(t) = \eta(t)] = F(\xi_1) \qquad (1)$$

Evidently, $F(\xi_1)$ represents all the information which the measurement of the random variables $y(t_0)$, ..., $y(t)$ has conveyed about the random variable $x_1(t_1)$. Any statistical estimate of the random variable $x_1(t_1)$ will be some function of this distribution and therefore a (nonrandom) function of the random variables $y(t_0)$, ..., $y(t)$. This statistical estimate is denoted by $X_1(t_1|t)$, or by just $X_1(t_1)$ or X_1 when the set of observed random variables or the time at which the estimate is required are clear from context.

Suppose now that X_1 is given as a fixed function of the random variables $y(t_0)$, ..., $y(t)$. Then X_1 is itself a random variable and its actual value is known whenever the actual values of $y(t_0)$, ..., $y(t)$ are known. In general, the actual value of $X_1(t_1)$ will be different from the (unknown) actual value of $x_1(t_1)$. To arrive at a rational way of determining X_1, it is natural to assign a *penalty* or *loss* for incorrect estimates. Clearly, the loss should be a (i) positive, (ii) nondecreasing function of the *estimation error* $\varepsilon = x_1(t_1) - X_1(t_1)$. Thus we define a *loss function* by

$$L(0) = 0$$

$$L(\varepsilon_2) \ge L(\varepsilon_1) \ge 0 \quad \text{when} \quad \varepsilon_2 \ge \varepsilon_1 \ge 0 \qquad (2)$$

$$L(\varepsilon) = L(-\varepsilon)$$

Some common examples of loss functions are: $L(\varepsilon) = a\varepsilon^2$, $a\varepsilon^4$, $a|\varepsilon|$, $a[1 - \exp(-\varepsilon^2)]$, etc., where a is a positive constant.

One (but by no means the only) natural way of choosing the random variable X_1 is to require that this choice should minimize the average loss or risk

$$E\{L[x_1(t_1) - X_1(t_1)]\} = E[E\{L[x(t_1) - X_1(t_1)]|y(t_0), \ldots, y(t)\}] \qquad (3)$$

Since the first expectation on the right-hand side of (3) does not depend on the choice of X_1 but only on $y(t_0)$, ..., $y(t)$, it is clear that minimizing (3) is equivalent to minimizing

$$E\{L[x_1(t_1) - X_1(t_1)]|y(t_0), \ldots, y(t)\} \qquad (4)$$

Under just slight additional assumptions, optimal estimates can be characterized in a simple way.

Theorem 1. *Assume that L is of type* (2) *and that the conditional distribution function* $F(\xi)$ *defined by* (1) *is:*

(*A*) *symmetric about the mean* $\bar{\xi}$:

$$F(\xi - \bar{\xi}) = 1 - F(\bar{\xi} - \xi)$$

(*B*) *convex for* $\xi \le \bar{\xi}$:

$$F(\lambda\xi_1 + (1 - \lambda)\xi_2) \le \lambda F(\xi_1) + (1 - \lambda)F(\xi_2)$$

for all $\xi_1, \xi_2 \le \bar{\xi}$ *and* $0 \le \lambda \le 1$

Then the random variable $x_1^*(t_1|t)$ *which minimizes the average loss* (3) *is the conditional expectation*

$$x_1^*(t_1|t) = E[x_1(t_1)|y(t_0), \ldots, y(t)] \qquad (5)$$

Proof: As pointed out recently by Sherman [25], this theorem follows immediately from a well-known lemma in probability theory.

Corollary. *If the random processes* $\{x_1(t)\}$, $\{x_2(t)\}$, and $\{y(t)\}$ *are gaussian, Theorem 1 holds.*

Proof: By Theorem 5, (*A*) (see Appendix), conditional distributions on a gaussian random process are gaussian. Hence the requirements of Theorem 1 are always satisfied.

In the control system literature, this theorem appears sometimes in a form which is more restrictive in one way and more general in another way:

Theorem I-a. *If* $L(\varepsilon) = \varepsilon^2$, *then Theorem 1 is true without assumptions* (*A*) and (*B*).

Proof: Expand the conditional expectation (4):

$$E[x_1^2(t_1)|y(t_0), \ldots, y(t)] - 2X_1(t_1)E[x_1(t_1)|y(t_0), \ldots, y(t)] + X_1^2(t_1)$$

and differentiate with respect to $X_1(t_1)$. This is not a completely rigorous argument; for a simple rigorous proof see Doob [15], pp. 77–78.

Remarks. (*a*) As far as the author is aware, it is not known what is the most general class of random processes $\{x_1(t)\}$, $\{x_2(t)\}$ for which the conditional distribution function satisfies the requirements of Theorem 1.

(*b*) Aside from the note of Sherman, Theorem 1 apparently has never been stated explicitly in the control systems literature. In fact, one finds many statements to the effect that loss functions of the general type (2) cannot be conveniently handled mathematically.

(*c*) In the sequel, we shall be dealing mainly with vector-valued random variables. In that case, the estimation problem is stated as: Given a vector-valued random process $\{\mathbf{x}(t)\}$ and observed random variables $\mathbf{y}(t_0)$, ..., $\mathbf{y}(t)$, where $\mathbf{y}(t) = \mathbf{M}\mathbf{x}(t)$ ($\mathbf{M}$ being a singular matrix; in other words, not all co-ordinates of $\mathbf{x}(t)$ can be observed), find an estimate $\mathbf{X}(t_1)$ which minimizes the expected loss $E[L(\|\mathbf{x}(t_1) - \mathbf{X}(t_1)\|)]$, $\| \ \|$ being the norm of a vector.

Theorem 1 remains true in the vector case also, provided we re- quire that the conditional distribution function of the n coordi- nates of the vector $\mathbf{x}(t_1)$,

$$Pr[x_1(t_1) \le \xi_1, \ldots, x_n(t_1) \le \xi_n | \mathbf{y}(t_0), \ldots, \mathbf{y}(t)] = F(\xi_1, \ldots, \xi_n)$$

be symmetric with respect to the n variables $\xi_1 - \bar{\xi}_1, \ldots, \xi_n - \bar{\xi}_n$ and convex in the region where all of these variables are negative.

Orthogonal Projections

The explicit calculation of the optimal estimate as a function of the observed variables is, in general, impossible. There is an important exception: The processes $\{x_1(t)\}$, $\{x_2(t)\}$ are gaussian.

On the other hand, if we attempt to get an optimal estimate under the restriction $L(\varepsilon) = \varepsilon^2$ and the additional requirement that the estimate be a linear function of the observed random variables, we get an estimate which is identical with the optimal estimate in the gaussian case, without the assumption of linearity or quadratic loss function. This shows that results obtainable by linear estimation can be bettered by nonlinear estimation only when (i) the random processes are nongaussian and even then (in view of Theorem 5, (*C*)) only (ii) by considering at least third-order probability distribution functions.

In the special cases just mentioned, the explicit solution of the estimation problem is most easily understood with the help of a geometric picture. This is the subject of the present section.

Consider the (real-valued) random variables $y(t_0)$, …, $y(t)$. The set of all linear combinations of these random variables with real coefficients

$$\sum_{i=t_0}^{t} a_i y(i) \qquad (6)$$

forms a *vector space (linear manifold)* which we denote by $\mathcal{Y}(t)$. We regard, abstractly, any expression of the form (6) as "point" or "vector" in $\mathcal{Y}(t)$; this use of the word "vector" should not be confused, of course, with "vector-valued" random variables, etc. Since we do not want to fix the value of t (i.e., the total number of possible observations), $\mathcal{Y}(t)$ should be regarded as a finite-dimensional subspace of the space of all possible observations.

Given any two vectors u, v in $\mathcal{Y}(t)$ (i.e., random variables expressible in the form (6)), we say that u and v are *orthogonal* if $Euv = 0$. Using the Schmidt orthogonalization procedure, as described for instance by Doob [15], p. 151, or by Loève [16], p. 459, it is easy to select an *orthonormal basis* in $\mathcal{Y}(t)$. By this is meant a set of vectors $e_{t_0}, \ldots, e_t$ in $\mathcal{Y}(t)$ such that any vector in $\mathcal{Y}(t)$ can be expressed as a unique linear combination of $e_{t_0}, \ldots, e_t$ and

$$\left.\begin{aligned} Ee_ie_j = \delta_{ij} = 1 \quad &\text{if} \quad i = j \\ = 0 \quad &\text{if} \quad i \neq j \end{aligned}\right\} \quad (i, j = t_0, \ldots, t) \tag{7}$$

Thus any vector $\bar{x}$ in $\mathcal{Y}(t)$. is given by

$$\bar{x} = \sum_{i=t_0}^{t} a_i e_i$$

and so the coefficients a_i can be immediately determined with the aid of (7):

$$E\bar{x}e_j = E\left(\sum_{i=t_0}^{t} a_i e_i\right) e_j = \sum_{i=t_0}^{t} a_i Ee_ie_j = \sum_{i=t_0}^{t} a_i \delta_{ij} = a_j \tag{8}$$

It follows further that any random variable x (not necessarily in $\mathcal{Y}(t)$) can be uniquely decomposed into two parts: a part $\bar{x}$ in $\mathcal{Y}(t)$ and a part $\tilde{x}$ orthogonal to $\mathcal{Y}(t)$ (i.e., orthogonal to every vector in $\mathcal{Y}(t)$). In fact, we can write

$$x = \bar{x} + \tilde{x} = \sum_{i=t_0}^{t} (Exe_i)e_i + \tilde{x} \tag{9}$$

Thus $\bar{x}$ is uniquely determined by equation (9) and is obviously a vector in $\mathcal{Y}(t)$. Therefore $\tilde{x}$ is also uniquely determined; it remains to check that it is orthogonal to $\mathcal{Y}(t)$:

$$E\tilde{x}e_i = E(x - \bar{x})e_i = Exe_i - E\bar{x}e_i$$

Now the co-ordinates of $\bar{x}$ with respect to the basis $e_{t_0}, \ldots, e_t$ are given either in the form $E\bar{x}e_i$ (as in (8)) or in the form Exe_i (as in (9)). Since the co-ordinates are unique, $Exe_i = E\bar{x}e_i$ $(i = t_0, \ldots, t)$; hence $E\tilde{x}e_i = 0$ and $\tilde{x}$ is orthogonal to every base vector e_i; and therefore to $\mathcal{Y}(t)$. We call $\bar{x}$ the *orthogonal projection* of x on $\mathcal{Y}(t)$.

There is another way in which the orthogonal projection can be characterized: $\bar{x}$ is that vector in $\mathcal{Y}(t)$ (i.e., that *linear* function of the random variables $y(t_0), \ldots, y(t)$) which minimizes the quadratic loss function. In fact, if $\bar{w}$ is any other vector in $\mathcal{Y}(t)$, we have

$$E(x - \bar{w})^2 = E(\tilde{x} + \bar{x} - \bar{w})^2 = E[(x - \bar{x}) + (\bar{x} - \bar{w})]^2$$

Since $\tilde{x}$ is orthogonal to every vector in $\mathcal{Y}(t)$ and in particular to $\bar{x} - \bar{w}$ we have

$$E(x - \bar{w})^2 = E(x - \bar{x})^2 + E(\bar{x} - \bar{w})^2 \geq E(x - \bar{x})^2 \tag{10}$$

This shows that, if $\bar{w}$ also minimizes the quadratic loss, we must have $E(\bar{x} - \bar{w})^2 = 0$ which means that the random variables $\bar{x}$ and $\bar{w}$ are equal (except possibly for a set of events whose probability is zero).

These results may be summarized as follows:

Theorem 2. Let $\{x(t)\}$, $\{y(t)\}$ *random processes with zero mean* (*i.e.*, $Ex(t) = Ey(t) = 0$ *for all* t). *We observe* $y(t_0), \ldots, y(t)$.

If either

(*A*) *the random processes* $\{x(t)\}$, $\{y(t)\}$ *are gaussian; or*

(*B*) *the optimal estimate is restricted to be a linear function of the observed random variables and* $L(\varepsilon) = \varepsilon^2$;

then

$$x^*(t_1|t) = \textit{optimal estimate of } x(t_1) \text{ given } y(t_0), \ldots, y(t)$$
$$= \textit{orthogonal projection } \bar{x}(t_1|t) \textit{ of } x(t_1) \text{ on } \mathcal{Y}(t). \tag{11}$$

These results are well-known though not easily accessible in the control systems literature. See Doob [15], pp. 75–78, or Pugachev [26]. It is sometimes convenient to denote the orthogonal projection by

$$\bar{x}(t_1 \mid t) \equiv x^*(t_1 \mid t) = \hat{E}[x(t_1) \mid \mathcal{Y}(t)]$$

The notation $\hat{E}$ is motivated by part (*b*) of the theorem: If the stochastic processes in question are gaussian, then orthogonal projection is actually identical with conditional expectation.

Proof. (*A*) This is a direct consequence of the remarks in connection with (10).

(*B*) Since $x(t)$, $y(t)$ are random variables with zero mean, it is clear from formula (9) that the orthogonal part $\tilde{x}(t_1|t)$ of $x(t_1)$ with respect to the linear manifold $\mathcal{Y}(t)$ is also a random variable with zero mean. Orthogonal random variables with zero mean are uncorrelated; if they are also gaussian then (by Theorem 5 (*B*)) they are independent. Thus

$$\begin{aligned} 0 = E\,\tilde{x}(t_1|t) \quad &= E[\tilde{x}(t_1|t)|y(t_0), \ldots, y(t)] \\ &= E[x(t_1) - \bar{x}(t_1|t)|y(t_0), \ldots, y(t)] \\ &= E[x(t_1)|y(t_0), \ldots, y(t)] - \bar{x}(t_1|t) = 0 \end{aligned}$$

Remarks. (*d*) A rigorous formulation of the contents of this section as $t \to \infty$ requires some elementary notions from the theory of Hilbert space. See Doob [15] and Loève [16].

(*e*) The physical interpretation of Theorem 2 is largely a matter of taste. If we are not worried about the assumption of gaussianness, part (*A*) shows that the orthogonal projection is the optimal estimate for all reasonable loss functions. If we do worry about gaussianness, even if we are resigned to consider only linear estimates, we know that orthogonal projections are *not* the optimal estimate for many reasonable loss functions. Since in practice it is difficult to ascertain to what degree of approximation a random process of physical origin is gaussian, it is hard to decide whether Theorem 2 has very broad or very limited significance.

(*f*) Theorem 2 is immediately generalized for the case of vector-valued random variables. In fact, we define the linear manifold $\mathcal{Y}(t)$ generated by $\mathbf{y}(t_0), \ldots, \mathbf{y}(t)$ to be the set of all linear combinations

$$\sum_{i=t_0}^{t} \sum_{j=1}^{m} a_{ij} y_j(i)$$

of all m co-ordinates of each of the random vectors $\mathbf{y}(t_0), \ldots, \mathbf{y}(t)$. The rest of the story proceeds as before.

(*g*) Theorem 2 states in effect that the optimal estimate under conditions (*A*) or (*B*) is a linear combination of all previous observations. In other words, the optimal estimate can be regarded as the output of a linear filter, with the input being the actually occurring values of the observable random variables; Theorem 2 gives a way of computing the impulse response of the optimal filter. As pointed out before, knowledge of this impulse response is not a complete solution of the problem; for this reason, no explicit formulas for the calculation of the impulse response will be given.

Models for Random Processes

In dealing with physical phenomena, it is not sufficient to give an empirical description but one must have also some idea of the underlying causes. Without being able to separate in some sense causes and effects, i.e., without the assumption of causality, one can hardly hope for useful results.

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

It is a fairly generally accepted fact that primary macroscopic sources of random phenomena are independent gaussian processes.[5] A well-known example is the noise voltage produced in a resistor due to thermal agitation. In most cases, *observed* random phenomena are not describable by independent random variables. The statistical dependence (correlation) between random signals observed at different times is usually explained by the presence of a dynamic system between the primary random source and the observer. *Thus a random function of time may be thought of as the output of a dynamic system excited by an independent gaussian random process.*

An important property of gaussian random signals is that they remain gaussian after passing through a linear system (Theorem 5 (*A*)). Assuming independent gaussian primary random sources, if the observed random signal is also gaussian, we may assume that the dynamic system between the observer and the primary source is *linear*. This conclusion may be forced on us also because of lack of detailed knowledge of the statistical properties of the observed random signal: Given any random process with known first and second-order averages, we can find a gaussian random process with the same properties (Theorem 5 (*C*)). Thus gaussian distributions and linear dynamics are natural, mutually plausible assumptions particularly when the statistical data are scant.

How is a dynamic system (linear or nonlinear) described? The fundamental concept is the notion of the *state*. By this is meant, intuitively, some quantitative information (a set of numbers, a function, etc.) which is the least amount of data one has to know about the past behavior of the system in order to predict its future behavior. The dynamics is then described in terms of *state transitions*, i.e., one must specify how one state is transformed into another as time passes.

A *linear* dynamic system may be described in general by the vector differential equation

$$\left.\begin{aligned} d\mathbf{x}/dt &= \mathbf{F}(t)\mathbf{x} + \mathbf{D}(t)\mathbf{u}(t) \\ \text{and} \qquad & \\ \mathbf{y}(t) &= \mathbf{M}(t)\mathbf{x}(t) \end{aligned}\right\} \tag{12}$$

where $\mathbf{x}$ is an n-vector, the *state* of the system (the components x_i of $\mathbf{x}$ are called *state variables*); $\mathbf{u}(t)$ is an m-vector ($m \le n$) representing the *inputs* to the system; $\mathbf{F}(t)$ and $\mathbf{D}(t)$ are $n \times n$, respectively, $n \times m$ matrices. If all coefficients of $\mathbf{F}(t)$, $\mathbf{D}(t)$, $\mathbf{M}(t)$ are constants, we say that the dynamic system (12) is *time-invariant* or *stationary*. Finally, $\mathbf{y}(t)$ is a *p-vector* denoting the outputs of the system; $\mathbf{M}(t)$ is an $n \times p$ matrix; $p \le n$

The physical interpretation of (12) has been discussed in detail elsewhere [18, 20, 23]. A look at the block diagram in Fig. 1 may be helpful. This is not an ordinary but a matrix block diagram (as revealed by the fat lines indicating signal flow). The integrator in

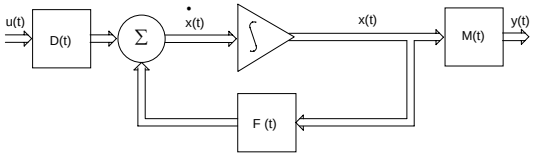


Fig 1. Matrix block diagram of the general linear continuous-dynamic system

[5] The probability distributions will be gaussian because macroscopic random effects may be thought of as the superposition of very many microscopic random effects; under very general conditions, such aggregate effects tend to be gaussian, regardless of the statistical properties of the microscopic effects. The assumption of independence in this context is motivated by the fact that microscopic phenomena tend to take place much more rapidly than macroscopic phenomena; thus primary random sources would appear to be independent on a macroscopic time scale.

Fig. 1 actually stands for n integrators such that the output of each is a state variable; $\mathbf{F}(t)$ indicates how the outputs of the integrators are fed back to the inputs of the integrators. Thus $f_{ij}(t)$ is the coefficient with which the output of the jth integrator is fed back to the input of the ith integrator. It is not hard to relate this formalism to more conventional methods of linear system analysis.

If we assume that the system (12) is stationary and that $\mathbf{u}(t)$ is constant during each sampling period, that is

$$\mathbf{u}(t + \tau) = \mathbf{u}(t); \quad 0 \le \tau < 1, \quad t = 0, 1, \ldots \tag{13}$$

then (12) can be readily transformed into the more convenient discrete form.

$$\mathbf{x}(t + 1) = \mathbf{\Phi}(1)\mathbf{x}(t) + \mathbf{\Delta}(1)\mathbf{u}(t); \quad t = 0, 1, \ldots$$

where [18, 20]

$$\mathbf{\Phi}(1) = \exp \mathbf{F} = \sum_{i=0}^{\infty} \mathbf{F}^i/i! \quad (\mathbf{F}^0 = \text{unit matrix})$$

and

$$\mathbf{\Delta}(1) = \left(\int_0^1 \exp \mathbf{F}\tau d\tau \right) \mathbf{D}$$

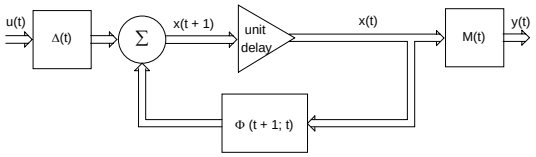


Fig 2. Matrix block diagram of the general linear discrete-dynamic system

See Fig. 2. One could also express exp $\mathbf{F}\tau$ in closed form using Laplace transform methods [18, 20, 22, 24]. If $\mathbf{u}(t)$ satisfies (13) but the system (12) is nonstationary, we can write analogously

$$\left.\begin{aligned} \mathbf{x}(t + 1) &= \mathbf{\Phi}(t + 1; t) + \mathbf{\Delta}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{M}(t)\mathbf{x}(t) \end{aligned}\right\} \quad t = 0, 1, \ldots \tag{14}$$

but of course now $\mathbf{\Phi}(t + 1; t)$, $\mathbf{\Delta}(t)$ cannot be expressed in general in closed form. Equations of type (14) are encountered frequently also in the study of complicated sampled-data systems [22]. See Fig. 2

$\mathbf{\Phi}(t + 1; t)$ is the *transition matrix* of the system (12) or (14). The notation $\mathbf{\Phi}(t_2; t_1)$ (t_2, t_1 = integers) indicates transition from time t_1 to time t_2. Evidently $\mathbf{\Phi}(t; t) = \mathbf{I}$ = unit matrix. If the system (12) is stationary then $\mathbf{\Phi}(t + 1; t) = \mathbf{\Phi}(t + 1 - t) = \mathbf{\Phi}(1) = \text{const}$. Note also the product rule: $\mathbf{\Phi}(t; s)\mathbf{\Phi}(s; r) = \mathbf{\Phi}(t; r)$ and the inverse rule $\mathbf{\Phi}^{-1}(t; s) = \mathbf{\Phi}(s; t)$, where t, s, r are integers. In a stationary system, $\mathbf{\Phi}(t; \tau) = \exp \mathbf{F}(t - \tau)$.

As a result of the preceding discussion, we shall represent random phenomena by the model

$$\mathbf{x}(t + 1) = \mathbf{\Phi}(t + 1; t)\mathbf{x}(t) + \mathbf{u}(t) \tag{15}$$

where $\{\mathbf{u}(t)\}$ is a vector-valued, independent, gaussian random process, with zero mean, which is completely described by (in view of Theorem 5 (*C*))

$$E\mathbf{u}(t) = \mathbf{0} \quad \text{for all } t;$$

$$E\mathbf{u}(t)\mathbf{u}'(s) = \mathbf{0} \quad \text{if } t \ne s$$

$$E\mathbf{u}(t)\mathbf{u}'(t) = \mathbf{G}(t).$$

Of course (Theorem 5 (*A*)), $\mathbf{x}(t)$ is then also a gaussian random process with zero mean, but it is no longer independent. In fact, if we consider (15) in the steady state (assuming it is a stable system), in other words, if we neglect the initial state $\mathbf{x}(t_0)$, then

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

$$x(t) = \sum_{r=-\infty}^{t-1} \mathbf{\Phi}(t; r+1)\mathbf{u}(r).$$

Therefore if $t \geq s$ we have

$$E\mathbf{x}(t)\mathbf{x}'(s) = \sum_{r=-\infty}^{s-1} \mathbf{\Phi}(t; r+1)\mathbf{Q}(r)\,\mathbf{\Phi}'(s; r+1).$$

Thus if we assume a linear dynamic model and know the statistical properties of the gaussian random excitation, it is easy to find the corresponding statistical properties of the gaussian random process $\{\mathbf{x}(t)\}$.

In real life, however, the situation is usually reversed. One is given the covariance matrix $E\mathbf{x}(t)\mathbf{x}'(s)$ (or rather, one attempts to estimate the matrix from limited statistical data) and the problem is to get (15) and the statistical properties of $\mathbf{u}(t)$. This is a subtle and presently largely unsolved problem in experimentation and data reduction. As in the vast majority of the engineering literature on the Wiener problem, we shall find it convenient to start with the model (15) and regard the problem of obtaining the model itself as a separate question. To be sure, the two problems *should* be optimized jointly if possible; the author is not aware, however, of any study of the *joint* optimization problem.

In summary, the following assumptions are made about random processes:

Physical random phenomena may be thought of as *due to primary random sources exciting dynamic systems. The primary sources are assumed to be independent gaussian random processes with zero mean; the dynamic systems will be linear. The random processes are therefore described by models such as* (15). *The question of how the numbers specifying the model are obtained from experimental data will not be considered.*

Solution of the Wiener problem

Let us now define the principal problem of the paper.

Problem I. *Consider the dynamic model*

$$\mathbf{x}(t+1) = \mathbf{\Phi}(t+1; t)\mathbf{x}(t) + \mathbf{u}(t) \tag{16}$$

$$\mathbf{y}(t) = \mathbf{M}(t)\mathbf{x}(t) \tag{17}$$

where $\mathbf{u}(t)$ is *an independent gaussian random process of n-vectors with zero mean,* $\mathbf{x}(t)$ is *an n-vector,* $\mathbf{y}(t)$ is *a p-vector* ($p \leq n$), $\mathbf{\Phi}(t+1; t)$, $\mathbf{M}(t)$ *are* $n \times n$, *resp.* $p \times n$, *matrices whose elements are nonrandom functions of time.*

Given the observed values of $\mathbf{y}(t_0)$, ..., $\mathbf{y}(t)$ *find an estimate* $\mathbf{x}^*(t_1|t)$ *of* $\mathbf{x}(t_1)$ *which minimizes the expected loss.* (See Fig. 2, where $\mathbf{\Delta}(t) = \mathbf{I}$.)

This problem includes as a special case the problems of filtering, prediction, and data smoothing mentioned earlier. It includes also the problem of reconstructing all the state variables of a linear dynamic system from noisy observations of some of the state variables ($p < n$!).

From Theorem 2-*a* we know that the solution of Problem I is simply the orthogonal projection of $\mathbf{x}(t_1)$ on the linear manifold $\mathcal{Y}(t)$ generated by the observed random variables. As remarked in the Introduction, this is to be accomplished by means of a linear (not necessarily stationary!) dynamic system of the general form (14). With this in mind, we proceed as follows.

Assume that $\mathbf{y}(t_0)$, ..., $\mathbf{y}(t-1)$ have been measured, i.e., that $\mathcal{Y}(t-1)$ is known. Next, at time t, the random variable $\mathbf{y}(t)$ is measured. As before let $\tilde{\mathbf{y}}(t|t-1)$ be the component of $\mathbf{y}(t)$ orthogonal to $\mathcal{Y}(t-1)$. If $\tilde{\mathbf{y}}(t|t-1) \equiv 0$, which means that the values of all components of this random vector are zero for almost every possible event, then $\mathcal{Y}(t)$ is obviously the same as $\mathcal{Y}(t-1)$ and therefore the measurement of $\mathbf{y}(t)$ does not convey any additional information. This is not likely to happen in a physically meaningful situation. In any case, $\tilde{\mathbf{y}}(t|t-1)$ generates a linear manifold (possibly 0) which we denote by $\mathcal{Z}(t)$. By definition, $\mathcal{Y}(t-1)$ and $\mathcal{Z}(t)$ taken together are the same manifold as $\mathcal{Y}(t)$, and every vector in $\mathcal{Z}(t)$ is orthogonal to every vector in $\mathcal{Y}(t-1)$.

Assuming by induction that $\mathbf{x}^*(t_1 - 1|t-1)$ is known, we can write:

$$\begin{aligned}\mathbf{x}^*(t_1|t) &= \hat{E}\,[\mathbf{x}(t_1)|\mathcal{Y}(t)] = \hat{E}\,[\mathbf{x}(t_1)|\mathcal{Y}(t-1)] + \hat{E}\,[\mathbf{x}(t_1)|\mathcal{Z}(t)] \\ &= \mathbf{\Phi}(t+1; t)\,\mathbf{x}^*(t_1 - 1|t-1) + \hat{E}\,[\mathbf{u}(t_1-1)|\mathcal{Y}(t-1)] \\ &\quad + \hat{E}\,[\mathbf{x}(t_1)|\mathcal{Z}(t)] \end{aligned} \tag{18}$$

where the last line is obtained using (16).

Let $t_1 = t + s$, where s is any integer. If $s \geq 0$, then $\mathbf{u}(t_1 - 1)$ is independent of $\mathcal{Y}(t-1)$. This is because $\mathbf{u}(t_1 - 1) = \mathbf{u}(t + s - 1)$ is then independent of $\mathbf{u}(t-2)$, $\mathbf{u}(t-3)$, ... and therefore by (16–17), independent of $\mathbf{y}(t_0)$, ..., $\mathbf{y}(t-1)$, hence independent of $\mathcal{Y}(t-1)$. Since, for all t, $\mathbf{u}(t_0)$ has zero mean by assumption, it follows that $\mathbf{u}(t_1 - 1)$ ($s \geq 0$) is orthogonal to $\mathcal{Y}(t-1)$. Thus if $s \geq 0$, the second term on the right-hand side of (18) vanishes; if $s < 0$, considerable complications result in evaluating this term. We shall consider only the case $t_1 \geq t$. Furthermore, it will suffice to consider in detail only the case $t_1 = t + 1$ since the other cases can be easily reduced to this one.

The last term in (18) must be a linear operation on the random variable $\tilde{\mathbf{y}}(t|t-1)$:

$$\hat{E}\,[\mathbf{x}(t+1)|\mathcal{Z}(t)] = \mathbf{\Delta}^*(t)\,\tilde{\mathbf{y}}(t|t-1) \tag{19}$$

where $\mathbf{\Delta}^*(t)$ is an $n \times p$ matrix, and the star refers to "optimal filtering".

The component of $\mathbf{y}(t)$ lying in $\mathcal{Y}(t-1)$ is $\bar{\mathbf{y}}(t|t-1) = \mathbf{M}(t)\mathbf{x}^*(t|t-1)$. Hence

$$\tilde{\mathbf{y}}(t|t-1) = \mathbf{y}(t) - \bar{\mathbf{y}}(t|t-1) = \mathbf{y}(t) - \mathbf{M}(t)\mathbf{x}^*(t|t-1). \tag{20}$$

Combining (18-20) (see Fig. 3) we obtain

$$\mathbf{x}^*(t+1|t) = \mathbf{\Phi}^*(t+1; t)\mathbf{x}^*(t|t-1) + \mathbf{\Delta}^*(t)\mathbf{y}(t) \tag{21}$$

where

$$\mathbf{\Phi}^*(t+1; t) = \mathbf{\Phi}(t+1; t) - \mathbf{\Delta}^*(t)\mathbf{M}(t) \tag{22}$$

Thus optimal estimation is performed by a linear dynamic system of the same form as (14). The state of the estimator is the previous estimate, the input is the last measured value of the observable random variable $\mathbf{y}(t)$, the transition matrix is given by (22). Notice that physical realization of the optimal filter requires only (*i*) the model of the random process (*ii*) the operator $\mathbf{\Delta}^*(t)$.

The estimation error is also governed by a linear dynamic system. In fact,

$$\begin{aligned}\tilde{\mathbf{x}}(t+1|t) &= \mathbf{x}(t+1) - \mathbf{x}^*(t+1|t) \\ &= \mathbf{\Phi}(t+1; t)\mathbf{x}(t) + \mathbf{u}(t) - \mathbf{\Phi}^*(t+1; t)\mathbf{x}^*(t|t-1) \\ &\quad - \mathbf{\Delta}^*(t)\mathbf{M}(t)\mathbf{x}(t)\end{aligned}$$

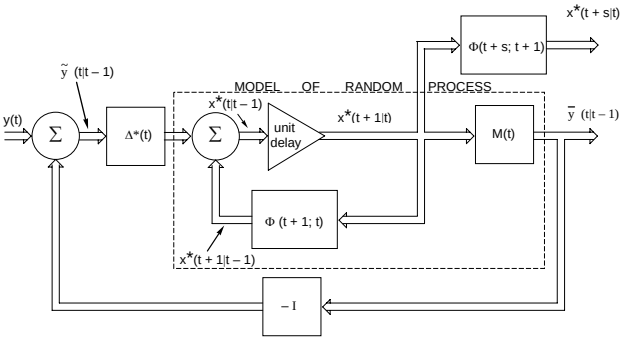


Fig. 3 Matrix block diagram of optimal filter

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

$$= \mathbf{\Phi}^*(t+1;\, t)\, \tilde{\mathbf{x}}\,(t|t-1) + \mathbf{u}(t) \tag{23}$$

Thus $\mathbf{\Phi}^*$ is also the transition matrix of the linear dynamic system governing the error.

From (23) we obtain at once a recursion relation for the covariance matrix $\mathbf{P}^*(t)$ of the optimal error $\tilde{\mathbf{x}}\,(t|t-1)$. Noting that $\mathbf{u}(t)$ is independent of $\mathbf{x}(t)$ and therefore of $\tilde{\mathbf{x}}\,(t|t-1)$ we get

$$\begin{aligned}\mathbf{P}^*(t+1) &= E\,\tilde{\mathbf{x}}\,(t+1|t)\,\tilde{\mathbf{x}}\,'(t+1|t)\\ &= \mathbf{\Phi}^*(t+1;\,t)E\,\tilde{\mathbf{x}}\,(t|t-1)\,\tilde{\mathbf{x}}\,'(t|t-1)\mathbf{\Phi}^{*\prime}\,(t+1;\,t) + \mathbf{Q}(t)\\ &= \mathbf{\Phi}^*(t+1;\,t)E\,\tilde{\mathbf{x}}\,(t|t-1)\,\tilde{\mathbf{x}}\,'(t|t-1)\mathbf{\Phi}'(t+1;\,t) + \mathbf{Q}(t)\\ &= \mathbf{\Phi}^*(t+1;\,t)\mathbf{P}^*(t)\mathbf{\Phi}'(t+1;\,t) + \mathbf{Q}(t)\end{aligned} \tag{24}$$

where $\mathbf{Q}(t) = E\mathbf{u}(t)\mathbf{u}'(t)$.

There remains the problem of obtaining an explicit formula for $\mathbf{\Delta}^*$ (and thus also for $\mathbf{\Phi}^*$). Since,

$$\tilde{\mathbf{x}}\,(t+1)|Z(t)) = \mathbf{x}(t+1) - \hat{E}\,[\mathbf{x}(t+1)|Z(t)]$$

is orthogonal to $\tilde{\mathbf{y}}\,(t\,|t-1)$, it follows that by (19) that

$$\begin{aligned}0 &= E[\mathbf{x}(t+1) - \mathbf{\Delta}^*(t)\,\tilde{\mathbf{y}}\,(t|t-1)]\,\tilde{\mathbf{y}}\,'(t|t-1)\\ &= E\mathbf{x}(t+1)\,\tilde{\mathbf{y}}\,'(t|t-1) - \mathbf{\Delta}^*(t)E\,\tilde{\mathbf{y}}\,(t|t-1)\,\tilde{\mathbf{y}}\,'(t|t-1).\end{aligned}$$

Noting that $\bar{\mathbf{x}}\,(t+1|t-1)$ is orthogonal to $Z(t)$, the definition of $\mathbf{P}(t)$ given earlier, and (17), it follows further

$$\begin{aligned}0 &= E\,\tilde{\mathbf{x}}\,(t+1|t-1)\,\tilde{\mathbf{y}}\,'(t|t-1) - \mathbf{\Delta}^*(t)\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t)\\ &= E[\mathbf{\Phi}(t+1;\,t)\,\tilde{\mathbf{x}}\,(t|t-1) + \mathbf{u}(t|t-1)]\,\tilde{\mathbf{x}}\,'(t|t-1)\mathbf{M}'(t)\\ &\qquad - \mathbf{\Delta}^*(t)\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t).\end{aligned}$$

Finally, since $\mathbf{u}(t)$ is independent of $\mathbf{x}(t)$,

$$0 = \mathbf{\Phi}(t+1;\,t)\mathbf{P}^*(t)\mathbf{M}'(t) - \mathbf{\Delta}^*(t)\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t).$$

Now the matrix $\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t)$will be positive definite and hence invertible whenever $\mathbf{P}^*(t)$ is positive definite, provided that none of the rows of $\mathbf{M}(t)$ are linearly dependent at any time, in other words, that none of the observed scalar random variables $y_1(t)$, ..., $y_m(t)$, is a linear combination of the others. Under these circumstances we get finally:

$$\mathbf{\Delta}^*(t) = \mathbf{\Phi}(t+1;\,t)\mathbf{P}^*(t)\mathbf{M}'(t)[\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t)]^{-1} \tag{25}$$

Since observations start at t_0, $\tilde{\mathbf{x}}\,(t_0|t_0-1) = \mathbf{x}(t_0)$; to begin the iterative evaluation of $\mathbf{P}^*(t)$ by means of equation (24), we must obviously specify $\mathbf{P}^*(t_0) = E\mathbf{x}(t_0)\mathbf{x}'(t_0)$. Assuming this matrix is positive definite, equation (25) then yields $\mathbf{\Delta}^*(t_0)$; equation (22) $\mathbf{\Phi}^*(t_0+1;\,t_0)$, and equation (24) $\mathbf{P}^*(t_0+1)$, completing the cycle. If now $\mathbf{Q}(t)$ is positive definite, then all the $\mathbf{P}^*(t)$ will be positive definite and the requirements in deriving (25) will be satisfied at each step.

Now we remove the restriction that $t_1 = t+1$. Since $\mathbf{u}(t)$ is orthogonal to $\mathcal{Y}(t)$, we have

$$\mathbf{x}^*(t+1|t) = \hat{E}\,[\mathbf{\Phi}(t+1;\,t)\mathbf{x}(t) + \mathbf{u}(t)|\mathcal{Y}(t)] = \mathbf{\Phi}(t+1;\,t)\mathbf{x}^*(t|t)$$

Hence if $\mathbf{\Phi}(t+1;\,t)$ has an inverse $\mathbf{\Phi}(t;\,t+1)$ (which is always the case when $\mathbf{\Phi}$ is the transition matrix of a dynamic system describable by a differential equation) we have

$$\mathbf{x}^*(t|t) = \mathbf{\Phi}(t;\,t+1)\mathbf{x}^*(t+1|t)$$

If $t_1 \geq t+1$, we first observe by repeated application of (16) that

$$\mathbf{x}(t+s) = \mathbf{\Phi}(t+s;\,t+1)\mathbf{x}(t+1) + \sum_{r1}^{s-1}\ \mathbf{\Phi}(t+s;\,t+r)\mathbf{u}(t+r) \qquad (s \geq 1)$$

Since $\mathbf{u}(t+s-1)$, …, $\mathbf{u}(t+1)$ are all orthogonal to $\mathcal{Y}(t)$,

$$\begin{aligned}\mathbf{x}^*(t+s|t) &= \hat{E}\,[\mathbf{x}(t+s)|\mathcal{Y}(t)]\\ &= \hat{E}\,[\mathbf{\Phi}(t+s;\,t+1)\mathbf{x}(t+1)|\mathcal{Y}(t)]\\ &= \mathbf{\Phi}(t+s;\,t+1)\mathbf{x}^*(t+1|t) \qquad (s \geq 1)\end{aligned}$$

If $s < 0$, the results are similar, but $\mathbf{x}^*(t-s|t)$ will have $(1-s)(n-p)$ co-ordinates.

The results of this section may be summarized as follows:

Theorem 3. (*Solution of the Wiener Problem*)

Consider Problem I. The optimal estimate $\mathbf{x}^*(t+1|t)$ *of* $\mathbf{x}(t+1)$ *given* $\mathbf{y}(t_0)$, ..., $\mathbf{y}(t)$ is *generated by the linear dynamic system*

$$\mathbf{x}^*(t+1|t) = \mathbf{\Phi}^*(t+1;\,t)\mathbf{x}^*(t|t-1) + \mathbf{\Delta}^*(t)\mathbf{y}(t) \tag{21}$$

The estimation error is given by

$$\tilde{\mathbf{x}}\,(t+1|t) = \mathbf{\Phi}^*(t+1;\,t)\,\tilde{\mathbf{x}}\,(t|t-1) + \mathbf{u}(t) \tag{23}$$

The covariance matrix of the estimation error is

$$\operatorname{cov}\,\tilde{\mathbf{x}}\,(t|t-1) = E\,\tilde{\mathbf{x}}\,(t|t-1)\,\tilde{\mathbf{x}}\,'(t|t-1) = \mathbf{P}^*(t) \tag{26}$$

The expected quadratic loss is

$$\sum_{i=1}^{n} E\tilde{x}_i^{\,2}(t\,|\,t-1)\ =\ \operatorname{trace}\,\mathbf{P}^*(t) \tag{27}$$

The matrices $\mathbf{\Delta}^*(t)$, $\mathbf{\Phi}^*(t+1;\,t)$, $\mathbf{P}^*(t)$ *are generated by the recursion relations*

$$\left.\begin{aligned}\mathbf{\Delta}^*(t) &= \mathbf{\Phi}(t+1;\,t)\mathbf{P}^*(t)\mathbf{M}'(t)[\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t)]^{-1} &(28)\\ \mathbf{\Phi}^*(t+1;\,t) &= \mathbf{\Phi}(t+1;\,t) - \mathbf{\Delta}^*(t)\mathbf{M}(t) &(29)\\ \mathbf{P}^*(t+1) &= \mathbf{\Phi}^*(t+1;\,t)\mathbf{P}^*(t)\mathbf{\Phi}'(t+1;\,t) + \mathbf{Q}(t) &(30)\end{aligned}\right\}\ t \geq t_0$$

In order to carry out the iterations, one must specify the covariance $\mathbf{P}^*(t_0)$ *of* $\mathbf{x}(t_0)$ *and the covariance* $\mathbf{Q}(t)$ *of* $\mathbf{u}(t)$. *Finally, for any* $s \geq 0$, *if* $\mathbf{\Phi}$ is *invertible*

$$\begin{aligned}\mathbf{x}^*(t+s|t) &= \mathbf{\Phi}(t+s;\,t+1)\mathbf{x}^*(t+1|t)\\ &= \mathbf{\Phi}(t+s;\,t+1)\mathbf{\Phi}^*(t+1;\,t)\mathbf{\Phi}(t;\,t+s-1)\\ &\qquad \times \mathbf{x}^*(t+s-1|t-1)\\ &\quad + \mathbf{\Phi}(t+s;\,t+1)\mathbf{\Delta}^*(t)\mathbf{y}(t)\end{aligned} \tag{31}$$

so that the estimate $\mathbf{x}^*(t+s|t)$ $(s \geq 0)$ is *also given by a linear dynamic system of the type* (21).

Remarks. (*h*) Eliminating $\mathbf{\Delta}^*$ and $\mathbf{\Phi}^*$ from (28–30), a nonlinear difference equation is obtained for $\mathbf{P}^*(t)$:

$$\begin{aligned}\mathbf{P}^*(t+1) &= \mathbf{\Phi}(t+1;\,t)\{\mathbf{P}^*(t) - \mathbf{P}^*(t)\mathbf{M}'(t)[\mathbf{M}(t)\mathbf{P}^*(t)\mathbf{M}'(t)]^{-1}\\ &\qquad \times \mathbf{P}^*(t)\mathbf{M}(t)\}\mathbf{\Phi}'(t+1;\,t) + \mathbf{Q}(t) \qquad t \geq t_0\end{aligned} \tag{32}$$

This equation is linear only if $\mathbf{M}(t)$ is invertible but then the problem is trivial since all components of the random vector $\mathbf{x}(t)$ are observable $\mathbf{P}^*(t+1) = \mathbf{Q}(t)$. Observe that equation (32) plays a role in the present theory analogous to that of the Wiener-Hopf equation in the conventional theory.

Once $\mathbf{P}^*(t)$ has been computed via (32) starting at $t = t_0$, the explicit specification of the optimal linear filter is immediately available from formulas (29-30). Of course, the solution of Equation (32), or of its differential-equation equivalent, is a much simpler task than solution of the Wiener-Hopf equation.

(*i*) The results stated in Theorem 3 do not resolve completely Problem I. Little has been said, for instance, about the physical significance of the assumptions needed to obtain equation (25), the convergence and stability of the nonlinear difference equation (32), the stability of the optimal filter (21), etc. This can actually be done in a completely satisfactory way, but must be left to a future paper. In this connection, the principal guide and

tool turns out to be the duality theorem mentioned briefly in the next section. See [29].

(*j*) By letting the sampling period (equal to one so far) approach zero, the method can be used to obtain the specification of a differential equation for the optimal filter. To do this, i.e., to pass from equation (14) to equation (12), requires computing the logarithm $\mathbf{F}^*$ of the matrix $\mathbf{\Phi}^*$. But this can be done only if $\mathbf{\Phi}^*$ is nonsingular—which is easily seen *not* to be the case. This is because it is sufficient for the optimal filter to have $n - p$ state variables, rather than n, as the formalism of equation (22) would seem to imply. By appropriate modifications, therefore, equation (22) can be reduced to an equivalent set of only $n - p$ equations whose transition matrix is nonsingular. Details of this type will be covered in later publications.

(*k*) The dynamic system (21) is, in general, nonstationary. This is due to two things: (1) The time dependence of $\mathbf{\Phi}(t+1;\, t)$ and $\mathbf{M}(t)$; (2) the fact that the estimation starts at $t = t_0$ and improves as more data are accumulated. If $\mathbf{\Phi}$, $\mathbf{M}$ are constants, it can be shown that (21) becomes a stationary dynamic system in the limit $t \to \infty$. This is the case treated by the classical Wiener theory.

(*l*) It is noteworthy that the derivations given are not affected by the nonstationarity of the model for $\mathbf{x}(t)$ or the finiteness of available data. In fact, as far as the author is aware, the only explicit recursion relations given before for the growing-memory filter are due to Blum [12]. However, his results are much more complicated than ours.

(*m*) By inspection of Fig. 3 we see that the optimal filter is a feedback system, and that the signal after the first summer is white noise since $\tilde{\mathbf{y}}(t|t-1)$ is obviously an orthogonal random process. This corresponds to some well-known results in Wiener filtering, see, e.g., Smith [28], Chapter 6, Fig. 6–4. However, this is apparently the first *rigorous* proof that every Wiener filter is realizable by means of a feedback system. Moreover, it will be shown in another paper that such a filter is always *stable*, under very mild assumptions on the model (16–17). See [29].

The Dual Problem

Let us now consider another problem which is conceptually very different from optimal estimation, namely, the noise-free regulator problem. In the simplest cases, this is:

Problem II. *Consider the dynamic system*

$$\mathbf{x}(t+1) = \hat{\mathbf{\Phi}}(t+1;\, t)\mathbf{x}(t) + \hat{\mathbf{M}}(t)\mathbf{u}(t) \tag{33}$$

where $\mathbf{x}(t)$ *is an n-vector,* $\mathbf{u}(t)$ *is an m-vector* ($m \le n$), $\hat{\mathbf{\Phi}}$, $\hat{\mathbf{M}}$ *are* $n \times n$ *resp.* $n \times m$ *matrices whose elements are nonrandom functions of time. Given any state* $\mathbf{x}(t)$ *at time t, we are to find a sequence* $\mathbf{u}(t)$, ..., $\mathbf{u}(T)$ *of control vectors which minimizes the performance index*

$$V[\mathbf{x}(t)] = \sum_{\tau=t}^{T+1} \mathbf{x}'(\tau)\mathbf{Q}(\tau)\mathbf{x}(\tau)$$

Where $\hat{\mathbf{Q}}(t)$ *is a positive definite matrix whose elements are nonrandom functions of time.* See Fig. 2, where $\mathbf{\Delta} = \hat{\mathbf{M}}$ and $\mathbf{M} = \mathbf{I}$.

Probabilistic considerations play no part in Problem II; it is implicitly assumed that every state variable can be measured exactly at each instant t, $t+1$, ..., T. It is customary to call $T \ge t$ the *terminal time* (it may be infinity).

The first general solution of the noise-free regulator problem is due to the author [18]. The main result is that the optimal control vectors $\mathbf{u}^*(t)$ are nonstationary linear functions of $\mathbf{x}(t)$. After a change in notation, the formulas of the Appendix, Reference [18] (see also Reference [23]) are as follows:

$$\mathbf{u}^*(t) = -\hat{\mathbf{\Delta}}^*(t)\mathbf{x}(t) \tag{34}$$

Under optimal control as given by (34), the "closed-loop" equations for the system are (see Fig. 4)

$$\mathbf{x}(t+1) = \hat{\mathbf{\Phi}}^*(t+1;\, t)\mathbf{x}(t)$$

and the minimum performance index at time t is given by

$$V^*[\mathbf{x}(t)] = \mathbf{x}'(t)\mathbf{P}^*(t-1)\mathbf{x}(t)$$

The matrices $\hat{\mathbf{\Delta}}^*(t)$, $\hat{\mathbf{\Phi}}^*(t+1;\, t)$, $\hat{\mathbf{P}}^*(t)$ are determined by the recursion relations:

$$\left.\begin{aligned}
\hat{\mathbf{\Delta}}^*(t) &= [\hat{\mathbf{M}}'(t)\,\hat{\mathbf{P}}^*(t)\,\hat{\mathbf{M}}(t)]^{-1}\hat{\mathbf{M}}'(t)\hat{\mathbf{P}}^*(t)\hat{\mathbf{\Phi}}(t+1;\, t) && (35)\\
\hat{\mathbf{\Phi}}^*(t+1;\, t) &= \hat{\mathbf{\Phi}}(t+1;\, t) - \hat{\mathbf{M}}(t)\,\hat{\mathbf{\Delta}}^*(t) && (36)\\
\hat{\mathbf{P}}^*(t-1) &= \hat{\mathbf{\Phi}}'(t+1;\, t)\hat{\mathbf{P}}^*(t)\,\hat{\mathbf{\Phi}}^*(t+1;\, t) + \hat{\mathbf{Q}}(t) && (37)
\end{aligned}\right\} \quad t \le T$$

Initially we must set $\hat{\mathbf{P}}^*(T) = \hat{\mathbf{Q}}(T+1)$.

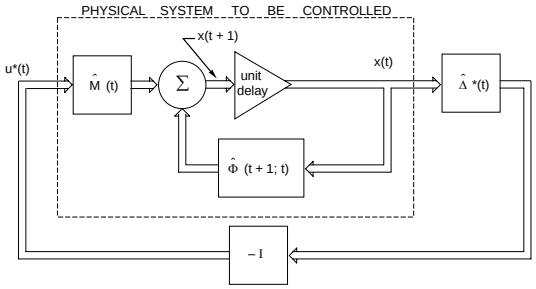


Fig. 4 Matrix block diagram of optimal controller

Comparing equations (35–37) with (28–30) and Fig. 3 with Fig. 4 we notice some interesting things which are expressed precisely by

Theorem 4. (*Duality Theorem*) *Problem I and Problem II are duals of each other in the following sense:*

Let $\tau \ge 0$. *Replace every matrix* $\mathbf{X}(t) = \mathbf{X}(t_0 + \tau)$ *in* (28–30) *by* $\hat{\mathbf{X}}'(t) = \hat{\mathbf{X}}'(T - \tau)$. *Then One has* (35–37). *Conversely, replace every matrix* $\hat{\mathbf{X}}(T - \tau)$ *in* (35–37) *by* $\mathbf{X}'(t_0 + \tau)$. *Then one has* (28–30).

Proof. Carry out the substitutions. For ease of reference, the dualities between the two problems are given in detail in Table 1.

Table 1

	Problem I	Problem II
1	$\mathbf{x}(t)$ (unobservable) state variables of random process.	$\mathbf{x}(t)$ (observable) state variables of plant to be regulated.
2	$\mathbf{y}(t)$ observed random variables.	$\mathbf{u}(t)$ control variables
3	t_0 first observation.	T last control action.
4	$\mathbf{\Phi}(t_0 + \tau + 1;\, t_0 + \tau)$ transition matrix.	$\hat{\mathbf{\Phi}}(T - \tau + 1;\, T - \tau)$ transition matrix.
5	$\mathbf{P}^*(t_0 + \tau)$ covariance of optimized estimation error.	$\hat{\mathbf{P}}^*(T - \tau)$ matrix of quadratic form for performance index under optimal regulation.
6	$\mathbf{\Delta}^*(t_0 + \tau)$ weighting of observation for optimal estimation.	$\hat{\mathbf{\Delta}}^*(T - \tau)$ weighting of state for optimal control.
7	$\mathbf{\Phi}^*(t_0 + \tau + 1;\, t_0 + \tau)$ transition matrix for optimal estimation error.	$\hat{\mathbf{\Phi}}^*(T - \tau + 1;\, T - \tau)$ transition matrix under optimal regulation.
8	$\mathbf{M}(t_0 + \tau)$ effect of state on observation.	$\hat{\mathbf{M}}(T - \tau)$ effect of control vectors on state.
9	$\mathbf{Q}(t_0 + \tau)$ covariance of random excitation.	$\hat{\mathbf{Q}}(T - \tau)$ matrix of quadratic form defining error criterion.

Remarks. (*n*) The *mathematical* significance of the duality between Problem I and Problem II is that both problems reduce to the solution of the Wiener-Hopf-like equation (32).

(*o*) The *physical* significance of the duality is intriguing. Why are observations and control dual quantities?

Recent research [29] has shown that the essence of the Duality Theorem lies in the duality of constraints at the output (represented by the matrix $\hat{\mathbf{M}}(t)$ in Problem I) and constraints at the input (represented by the matrix $\hat{\mathbf{M}}(t)$ in Problem II).

(*p*) Applications of Wiener's methods to the solution of noise-free regulator problem have been known for a long time; see the recent textbook of Newton, Gould, and Kaiser [27]. However, the connections between the two problems, and in particular the duality, have apparently never been stated precisely before.

(*q*) The duality theorem offers a powerful tool for developing more deeply the theory (as opposed to the computation) of Wiener filters, as mentioned in Remark (*i*). This will be published elsewhere [29].

Applications

The power of the new approach to the Wiener problem, as expressed by Theorem 3, is most obvious when the data of the problem are given in numerical form. In that case, one simply performs the numerical computations required by (28–30). Results of such calculations, in some cases of practical engineering interest, will be published elsewhere.

When the answers are desired in closed analytic form, the iterations (28–30) may lead to very unwieldy expressions. In a few cases, $\mathbf{\Delta}^*$ and $\mathbf{\Phi}^*$ can be put into "closed form." Without discussing here how (if at all) such closed forms can be obtained, we now give two examples indicative of the type of results to be expected.

Example 1. Consider the problem mentioned under "Optimal Estimates." Let $x_1(t)$ be the signal and $x_2(t)$ the noise. We assume the model:

$$x_1(t+1) = \phi_{11}(t+1;\, t)x_1(t) + u_1(t)$$

$$x_2(t+1) = u_2(t)$$

$$y_1(t) = x_1(t) + x_2(t)$$

The specific data for which we desire a solution of the estimation problem are as follows:

1 $t_1 = t + 1$; $t_0 = 0$
2 $Ex_1^2(0) = 0$, i.e., $x_1(0) = 0$
3 $Eu_1^2(t) = a^2$, $Eu^2(t) = b^2$, $Eu_1(t)\, u_2(t) = 0$ (for all t)
4 $\phi_{11}(t+1;\, t) = \phi_{11} = \text{const.}$

A simple calculation shows that the following matrices satisfy the difference equations (28–30), for all $t \geq t_0$:

$$\mathbf{\Delta}^*(t) = \begin{bmatrix} \phi_{11}C(t) \\ 0 \end{bmatrix}$$

$$\mathbf{\Phi}^*(t+1;\, t) = \begin{bmatrix} \phi_{11}[1-C(t)] & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{P}^*(t+1) = \begin{bmatrix} a^2 + \phi_{11}^2 b^2 C(t) & 0 \\ 0 & b^2 \end{bmatrix}$$

where
$$C(t+1) = 1 - \frac{b^2}{a^2 + b^2 + \phi_{11}^2 b^2 C(t)} \qquad t \geq 0 \qquad (38)$$

Since it was assumed that $x_1(0) = 0$, neither $x_1(1)$ nor $x_2(1)$ can be predicted from the measurement of $y_1(0)$. Hence the measurement at time $t = 0$ is useless, which shows that we should set $C(0) = 0$. This fact, with the iterations (38), completely determines the function $C(t)$. The nonlinear difference equation (38) plays the role of the Wiener-Hopf equation.

If $b^2/a^2 \ll 1$, then $C(t) \approx 1$ which is essentially pure prediction. If $b^2/a^2 \gg 1$, then $C(t) \approx 0$, and we depend mainly on $x_1^*(t|t-1)$ for the estimation of $x_1^*(t+1|t)$ and assign only very small weight to the measurement $y_1(t)$; this is what one would expect when the measured data are very noisy.

In any case, $x_2^*(t|t-1) = 0$ at all times; one cannot predict independent noise! This means that ϕ^*_{12} can be set equal to zero. The optimal predictor is a first-order dynamic system. See Remark (*j*).

To find the stationary Wiener filter, let $t = \infty$ on both sides of (38), solve the resulting quadratic equation in $C(\infty)$, etc.

Example 2. A number or particles leave the origin at time $t_0 = 0$ with random velocities; after $t = 0$, each particle moves with a constant (unknown) velocity. Suppose that the position of one of these particles is measured, the data being contaminated by stationary, additive, correlated noise. What is the optimal estimate of the position and velocity of the particle at the time of the last measurement?

Let $x_1(t)$ be the position and $x_2(t)$ the velocity of the particle; $x_3(t)$ is the noise. The problem is then represented by the model,

$$x_1(t+1) = x_1(t) + x_2(t)$$
$$x_2(t+1) = x_2(t)$$
$$x_3(t+1) = \phi_{33}(t+1;\, t)x_3(t) + u_3(t)$$
$$y_1(t) = x_1(t) + x_3(t)$$

and the additional conditions

1 $t_1 = t$; $t_0 = 0$
2 $Ex_1^2(0) = Ex_2(0) = 0$, $Ex_2^2(0) = a^2 > 0$;
3 $Eu_3(t) = 0$, $Eu_3^2(t) = b^2$.
4 $\phi_{33}(t+1;\, t) = \phi_{33} = \text{const.}$

According to Theorem 3, $\mathbf{x}^*(t|t)$ is calculated using the dynamic system (31).

First we solve the problem of predicting the position and velocity of the particle one step ahead. Simple considerations show that

$$\mathbf{P}^*(1) = \begin{bmatrix} a^2 & a^2 & 0 \\ a^2 & a^2 & 0 \\ 0 & 0 & b^2 \end{bmatrix} \quad \text{and} \quad \mathbf{\Delta}^*(0) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

It is then easy to check by substitution into equations (28–30) that

$$\mathbf{P}^*(t) = \frac{b^2}{C_1(t-1)} \times \begin{bmatrix} t^2 & t & -\phi_{33}t(t-1) \\ t & 1 & -\phi_{33}(t-1) \\ -\phi_{33}t(t-1) & -\phi_{33}(t-1) & \phi_{33}^2(t-1)^2 + C_1(t-1) \end{bmatrix}$$

is the correct expression for the covariance matrix of the prediction error $\tilde{\mathbf{x}}(t|t-1)$ for all $t \geq 1$, provided that we define

$$C_1(0) = b^2/a^2$$
$$C_1(t) = C_1(t-1) + [t - \phi_{33}(t-1)]^2, \quad t \geq 1$$

It is interesting to note that the results just obtained are valid also when ϕ_{33} depends on t. This is true also in Example 1. In conventional treatments of such problems there *seems* to be an essential difference between the cases of stationary and nonstationary noise. This misleading impression created by the conventional theory is due to the very special *methods* used in solving the Wiener-Hopf equation.

Introducing the abbreviation

$$C_2(0) = 0$$
$$C_2(t) = t - \phi_{33}(t-1), \quad t \geq 1$$

and observing that

$$\text{cov}\ \tilde{\mathbf{x}}(t+1|t) = \mathbf{P}^*(t+1)$$
$$= \mathbf{\Phi}(t+1;\, t)[\text{cov}\ \tilde{\mathbf{x}}(t|t)]\mathbf{\Phi}'(t+1;\, t) + \mathbf{Q}(t)$$

the matrices occurring in equation (31) and the covariance matrix of $\tilde{\mathbf{x}}(t|t)$ are found after simple calculations. We have, for all $t \geq 0$,

$$\mathbf{\Phi}(t;\, t+1)\mathbf{\Delta}^*(t) = \frac{1}{C_1(t)}\begin{bmatrix} tC_2(t) \\ C_2(t) \\ C_1(t) - tC_2(t) \end{bmatrix}$$

$$\mathbf{\Phi}(t;\, t+1)\mathbf{\Phi}^*(t+1;\, t)\mathbf{\Phi}(t+1;\, t)$$

$$= \frac{1}{C_1(t)}\begin{bmatrix} C_1(t) - tC_2(t) & C_1(t) - tC_3(t) & -\phi_{33}tC_2(t) \\ -C_2(t) & C_1(t) - C_2(t) & -\phi_{33}C_2(t) \\ -C_1(t) + tC_2(t) & -C_1(t) + tC_2(t) & +\phi_{33}tC_2(t) \end{bmatrix}$$

and

$$\operatorname{cov}\,\tilde{\mathbf{x}}(t|t) = E\,\tilde{\mathbf{x}}(t|t)\,\tilde{\mathbf{x}}'(t|t) = \frac{b^2}{C_1(t)}\begin{bmatrix} t^2 & t & -t^2 \\ t & 1 & -t \\ -t^2 & -t & t^2 \end{bmatrix}$$

To gain some insight into the behavior of this system, let us examine the limiting case $t \to \infty$ of a large number of observations. Then $C_1(t)$ obeys approximately the differential equation

$$dC_1(t)/dt \approx C_2^2(t) \qquad (t >> 1)$$

from which we find

$$C_1(t) \approx (1-\phi_{33})^2 t^3/3 + \phi_{33}(1-\phi_{33})t^2 + \phi_{33}^2 t + b^2/a^2 \qquad (t >> 1) \qquad (39)$$

Using (39), we get further,

$$\mathbf{\Phi}^{-1}\mathbf{\Phi}^*\mathbf{\Phi} \approx \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \end{bmatrix} \text{ and } \mathbf{\Phi}^{-1}\mathbf{\Delta}^* \approx \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad (t >> 1)$$

Thus as the number of observations becomes large, we depend almost exclusively on $x_1^*(t|t)$ and $x_2^*(t|t)$ to estimate $x_1^*(t+1|t+1)$ and $x_2^*(t+1|t+1)$. Current observations are used almost exclusively to estimate the noise

$$x_3^*(t|t) \approx y_1^*(t) - x_1^*(t|t) \qquad (t >> 1)$$

One would of course expect something like this since the problem is analogous to fitting a straight line to an increasing number of points.

As a second check on the reasonableness of the results given, observe that the case $t >> 1$ is essentially the same as prediction based on continuous observations. Setting $\phi_{33} = 0$, we have

$$E\,\tilde{x}_1^{\,2}(t\,|\,t) \approx \frac{a^2b^2t^2}{b^2 + a^2t^3/3} \qquad (t >> 1;\ \phi_{33} = 0)$$

which is identical with the result obtained by Shinbrot [11], Example 1, and Solodovnikov [14], Example 2, in their treatment of the Wiener problem in the finite-length, continuous-data case, using an approach entirely different from ours.

Conclusions

This paper formulates and solves the Wiener problem from the "state" point of view. On the one hand, this leads to a very general treatment including cases which cause difficulties when attacked by other methods. On the other hand, the Wiener problem is shown to be closely connected with other problems in the theory of control. Much remains to be done to exploit these connections.

References

1 N. Wiener, "The Extrapolation, Interpolation and Smoothing of Stationary Time Series," John Wiley & Sons, Inc., New York, N.Y., 1949.

2 L. A. Zadeh and J. R. Ragazzini, "An Extension of Wiener's Theory of Prediction," *Journal of Applied Physics,* vol. 21, 1950, pp. 645–655.

3 H. W. Bode and C. E. Shannon, "A Simplified Derivation of Linear Least-Squares Smoothing and Prediction Theory," *Proceedinqs IRE,* vol. 38, 1950, pp. 417–425.

4 R. C. Booton, "An Optimization Theory for Time-Varying Linear Systems With Nonstationary Statistical Inputs," *Proceedings IRE,* vol. 40, 1952, pp. 977–981.

5 J. H. Laning and R. H. Battin, "Random Processes in Automatic Control," McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

6 W. B. Davenport, Jr., and W. L. Root, "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

7 S. Darlington, "Linear Least-Squares Smoothing and Prediction, With Applications," *Bell System Tech. Journal,* vol. 37, 1958, pp.1221–1294.

8 G. Franklin, "The Optimum Synthesis of Sampled-Data Systems" Doctoral dissertation, Dept. of Elect. Engr., Columbia University, 1955.

9 A. B. Lees, "Interpolation and Extrapolation of Sampled Data," *Trans. IRE Prof. Grou.p on Information Theory,* IT-2, 1956, pp. 173–175.

10 R. C. Davis. "On the Theory of Prediction of Nonstationary Stochastic Processes," *Journal of Applied Physics,* vol. 23. 1952, pp. 1047–1053.

11 M. Shinbrot, "Optimization of Tirne-Varying Linear Systems With Nonstationary Inputs," TRANS. ASME, vol. 80, 1958. pp. 457–462.

12 M. Blum, "Recursion Formulas for Growing Memory Digital Filters," *Trans. IRE Prof. Group on Information Theory.* IT-4. 1958, pp. 24–30.

13 V. S. Pugachev, "The Use of Canonical Expansions of Random Functions in Determining an Optimum Linear System," *Automatics and Remote Control* (USSR), vol. 17, 1956, pp. 489–499; translation pp. 545–556.

14 V. V. Solodovnikov and A. M. Batkov. "On the Theory of Self-Optimizing Systems (in German and Russian)," Proc. Heidelberg Conference on Automatic Control, 1956. pp. 308–323.

15 J. L. Doob, "Stochastic Processes," John Wiley & Sons, Inc., New York, N. Y., 1955.

16 M. Loève, "Probability Theory," Van Nostrand Company, Inc., New York, N. Y., 1955.

17 R. E. Bellman, I. Glicksberg, and 0. A. Gross, "Some Aspects of the Mathematical Theory of Control Processes," RAND Report R-313, 1958, 244 pp.

18 R. E. Kalman and R. W. Koepcke, "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes," TRANS. ASME, vol. 80, 1958, pp. 1820–1826.

19 J. E. Bertram, "Effect of Quantization in Sampled-Feedback Systems," *Trans. AlEE,* vol. 77, II, 1958, pp. 177–182.

20 R. E. Kalman and J. E. Bertram, "General Synthesis Procedure for Computer Control of Single and Multi-Loop Linear Systems" *Trans. AlEE,* vol. 77, II, 1958, pp. 602–609.

21 C. W. Merriam, III, "A Class of Optimum Control Systems," *Journal of the Franklin Institute,* vol. 267, 1959, pp. 267–281.

22 R. E. Kalman and J. E. Bertram, "A Unified Approach to the Theory of Sampling Systems," *Journal of the Franklin Institute,* vol. 267, 1959, pp. 405–436.

23 R. E. Kalman and R. W. Koepcke, "The Role of Digital Computers in the Dynamic Optimization of Chemical Reactors," Proc. Western Joint Computer Conference, 1959, pp. 107–116.

24 R. E. Kalman, "Dynamic Optimization of Linear Control Systems, I. Theory," to appear.

25 S. Sherman, "Non-Mean-Square Error Criteria," *Trans. IRE Prof. Group on Information Theory,* IT-4, 1958, pp. 125–126.

26 V. S. Pugachev, "On a Possible General Solution of the Problem of Determining Optimum Dynamic Systems," *Automatics and Remote Control* (USSR), vol. 17, 1956, pp. 585–589.

27 G. C. Newton, Jr., L. A. Gould, and J. F. Kaiser, "Analytical Design of Linear Feedback Controls," John Wiley & Sons, Inc., New York, N. Y., 1957.

28 0. J. M. Smith, "Feedback Control Systems," McGraw-Hill Book Company, Inc., New York, N. Y., 1958.

29 R. E. Kalman, "On the General Theory of Control Systems," Proceedings First International Conference on Automatic Control, Moscow, USSR, 1960.

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45.

APPENDIX
RANDOM PROCESSES: BASIC CONCEPTS

For convenience of the reader, we review here some elementary definitions and facts about probability and random processes. Everything is presented with the utmost possible simplicity; for greater depth and breadth, consult Laning and Battin [5] or Doob [15].

A *random variable* is a function whose values depend on the outcome of a chance event. The *values* of a random variable may be any convenient mathematical entities; real or complex numbers, vectors, etc. For simplicity, we shall consider here only real-valued random variables, but this is no real restriction. Random variables will be denoted by *x, y,* ... and their values by ξ, η, …. Sums, products, and functions of random variables are also random variables.

A random variable x can be explicitly defined by stating the probability that x is less than or equal to some real constant ξ. This is expressed symbolically by writing

$$Pr(x \le \xi) = F_x(\xi);\ F_x(-\infty) = 0,\ F_x(+\infty) = 1$$

$F_x(\xi)$ is called the *probability distribution function* of the random variable x. When $F_x(\xi)$ is differentiable with respect to ξ, then $f_x(\xi) = dF_x(\xi)/d\xi$ is called the *probability density function* of x.

The *expected value* (*mathematical expectation, statistical average, ensemble average, mean,* etc., are commonly used synonyms) of any nonrandom function $g(x)$ of a random variable x is defined by

$$Eg(x) = E[g(x)] = \int_{-\infty}^{\infty} g(\xi)dF_x(\xi) = \int_{-\infty}^{\infty} g(\xi)\,f_x(\xi)d\xi \qquad (40)$$

As indicated, it is often convenient to omit the brackets after the symbol E. A sequence of random variables (finite or infinite)

$$\{x(t)\} = \ldots, x(-1), x(0), x(1), \ldots \qquad (41)$$

is called a *discrete* (or *discrete-parameter*) *random* (or *stochastic*) *process.* One particular set of observed values of the random process (41)

$$\ldots, \xi(-1), \xi(0), \xi(1), \ldots$$

is called a *realization* (or a *sample function)* of the process. Intuitively, a random process is simply a set of random variables which are indexed in such a way as to bring the notion of time into the picture.

A random process is *uncorrelated* if

$$Ex(t)x(s) = Ex(t)Ex(s) \qquad (t \ne s)$$

If, furthermore,

$$Ex(t)x(s) = 0 \qquad (t \ne s)$$

then the random process is *orthogonal.* Any uncorrelated random process can be changed into orthogonal random process by replacing $x(t)$ by $x'(t) = x(t) - Ex(t)$ since then

$$Ex'(t)x'(s) = E[x(t) - Ex(t)]\cdot[x(s) - Ex(s)]$$
$$= Ex(t)x(s) - Ex(t)Ex(s) = 0$$

It is useful to remember that, if a random process is orthogonal, then

$$E[x(t_1) + x(t_2) + \ldots]^2 = Ex^2(t_1) + Ex^2(t_2) + \ldots \quad (t_1 \ne t_2 \ne \ldots)$$

If $\mathbf{x}$ is a vector-valued random variable with components $x_1, \ldots, x_n$ (which are of course random variables), the matrix

$$[E(x_i - Ex_i)(x_j - Ex_j)] = E(\mathbf{x} - E\mathbf{x})(\mathbf{x}' - E\mathbf{x}')$$
$$= \operatorname{cov} \mathbf{x} \qquad (42)$$

is called the *covariance matrix of x.*

A random process may be specified explicitly by stating the probability of simultaneous occurrence of any finite number of events of the type
$x(t_1) \le \xi_1, \ldots, x(t_n) \le \xi_n;\ (t_1 \ne \ldots \ne t_n)$, i.e.,

$$Pr[(x(t_1) \le \xi_1, \ldots, x(t_n) \le \xi_n)] = F_{x(t_1),\ \ldots,\ x(t_n)}(\xi_1, \ldots, \xi_n) \qquad (43)$$

where $F_{x(t_1),\ \ldots,\ x(t_n)}$ is called the *joint probability distribution function* of the random variables $x(t_1), \ldots, x(t_n)$. The *joint probability density function* is then

$$f_{x(t_1),\ \ldots,\ x(t_n)}(\xi_1, \ldots, \xi_n) = \partial^n F_{n(t_1),\ \ldots,\ x(t_n)}/\partial\xi_1, \ldots, \partial\xi_n$$

provided the required derivatives exist. The expected value $Eg[x(t_1), \ldots, x(t_n)]$ of any nonrandom function of n random variables is defined by an n-fold integral analogous to (40).

A random process is *independent* if for any finite $t_1 \ne \ldots \ne t_n$, (43) is equal to the product of the first-order distributions

$$Pr[x(t_1) \le \xi_1] \ldots Pr[x(t_n) \le \xi_n]$$

If a set of random variables is independent, then they are obviously also uncorrelated. The converse is not true in general. For a set of more than 2 random variables to be independent, it is not sufficient that any pair of random variables be independent.

Frequently it is of interest to consider the probability distribution of a random variable $x(t_{n+1})$ of a random process given the actual values $\xi(t_1), \ldots, \xi(t_n)$ with which the random variables $x(t_1), \ldots, x(t_n)$ have occurred. This is denoted by

$$Pr[x(t_{n+1}) \le \xi_{n+1} | x(t_1) = \xi_1, \ldots, x(t_n) = \xi_n]$$

$$= \frac{\int_{-\infty}^{\xi_{n+1}} f_{x(t_1),\ldots,x(t_{n+1})}(\xi_1,\ldots,\xi_{n+1})d\xi_{n+1}}{f_{x(t_1),\ldots,x(t_n)}(\xi_1,\ldots,\xi_n)} \qquad (44)$$

which is called the *conditional probability distribution function* of $x(t_{n+1})$ given $x(t_1), \ldots, x(t_n)$. The *conditional expectation*

$$E\{g[x(t_{n+1})] | x(t_1), \ldots, x(t_n)\}$$

is defined analogously to (40). The conditional expectation is a random variable; it follows that

$$E[E\{g[x(t_{n+1})] | x(t_1), \ldots, x(t_n)\}] = E\{g[x(t_{n+1})]\}$$

In all cases of interest in this paper, integrals of the type (40) or (44) need never be evaluated explicitly, only the *concept* of the expected value is needed.

A random variable x is *gaussian* (or *normally distributed)* if

$$f_x(\xi) = \frac{1}{[2\pi E(x - Ex)^2]^{1/2}} \exp\left[-\frac{1}{2}\frac{(\xi - Ex)^2}{E(x - Ex)^2}\right]$$

which is the well-known bell-shaped curve. Similarly, a random vector $\mathbf{x}$ is *gaussian* if

$$f_x(\boldsymbol{\xi}) = \frac{1}{(2\pi)^{n/2}(det\ \mathbf{C})^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\xi} - E\mathbf{x})'\mathbf{C}^{-1}(\boldsymbol{\xi} - E\mathbf{x})\right]$$

where $\mathbf{C}^{-1}$ is the inverse of the covariance matrix (42) of $\mathbf{x}$. A *gaussian random process* is defined similarly.

The importance of gaussian random variables and processes is largely due to the following facts:

Theorem 5. (*A*) *Linear functions* (*and therefore conditional expectations*) *on a gaussian random process are gaussian random variables.*

(*B*) *Orthogonal gaussian random variables are independent.*

(*C*) *Given any random process with means* $Ex(t)$ *and covariances* $Ex(t)x(s)$, *there exists a unique gaussian random process with the same means and covariances.*

Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45. Copyright © 1960 by ASME

Explanation of this transcription, John Lukesh, 20 January 2002.

Using a photo copy of R. E. Kalman's 1960 paper from an original of the ASME "Journal of Basic Engineering", March 1960 issue, I did my best to make an accurate version of this rather significant piece, in an up-to-date computer file format. For this I was able to choose page formatting and type font spacings that resulted in a document that is a close match to the original. (All pages start and stop at about the same point, for example; even most individual lines of text do.) I used a recent version of Word for Windows and a recent Hewlett Packard scanner with OCR (optical character recognition) software. The OCR software is very good on plain text, even distinguishing between italic versus regular characters quite reliably, but it does not do well with subscripts, superscripts, and special fonts, which were quite prevalent in the original paper. And I found there was no point in trying to work from the OCR results for equations. A lot of manual labor was involved.

Since I wanted to make a faithful reproduction of the original, I did not make any changes to correct (what I believed were) mistakes in it. For example, equation (32) has a $\mathbf{P}^*(t)\mathbf{M}(t)$ product that should be reversed, I think. I left this, and some other things that I thought were mistakes in the original, as is. (I didn't find very many other problems with the original.) There may, *of course*, be problems with my transcription. The plain text OCR results, which didn't require much editing, are pretty accurate I think. But the subscripts etc and the equations which I copied essentially manually, are suspect. I've reviewed the resulting document quite carefully, several times finding mistakes in what I did each time. The last time there were five, four cosmetic and one fairly inconsequential. There are probably more. I would be very pleased to know about it if any reader of this finds some of them; jlukesh@deltanet.com.