

Faculté de technologie - Département d'Electronique

Niveau: Master 1

Filière: Automatique et Systèmes

## Chapitre II: Identification par estimation statistique

### 1. Introduction:

L'identification consiste à chercher les paramètres de modèles mathématiques d'un système, à partir de données expérimentales et de connaissances disponibles a priori. Ces paramètres peuvent avoir une signification physique, ou ne pas en avoir, comme c'est le cas pour les modèles de comportement. Dans les deux cas ils doivent fournir une approximation fidèle des comportements du système physique, dans la mesure où leurs paramètres sont ajustés à partir de données expérimentales. L'objectif cherché est de rendre identiques les réponses du processus et du modèle, pour des séquences d'entrée données.

### 2. Identification paramétrique:

Il existe deux analyses possibles de l'identification selon que l'on considère la nature aléatoire du bruit comme négligeable et donc  $y$  est de nature déterministe, ou alors le bruit comme un signal aléatoire important et  $y$  est de nature statistique:

Approche déterministe: La sortie d'un système déterministe est parfaitement déterminée à partir de ses entrées passées et présentes. Les erreurs de nature aléatoires sont ignorées et les paramètres estimés sont alors déterministes et peuvent être obtenus à partir d'un nombre réduit de mesures.

Approches stochastique: Dans un système stochastique, le signal de sortie est de type aléatoire et change donc de manière imprévisible. Les paramètres estimés, sont aussi des variables aléatoires qu'il convient de caractériser statistiquement. Le nombre de mesures nécessaires est en général plus élevé que pour une approche déterministe.

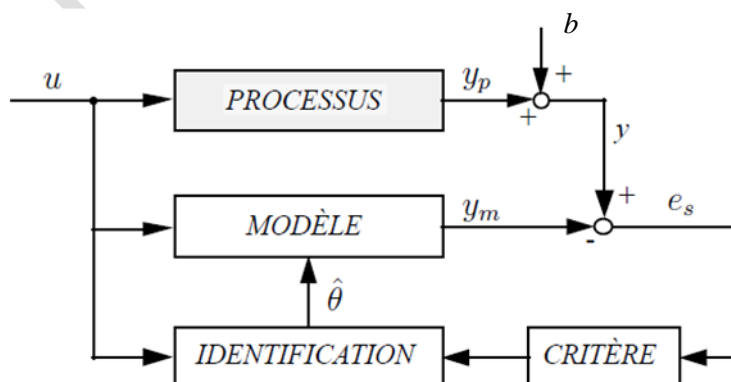


Figure II.1 Identification par représentation paramétrique

Le principe de ces techniques est très simple, comme illustré sur la figure II.1, le système réel et le modèle mathématique à identifier sont excités par la même entrée  $u(t)$ . La sortie du procédé  $y_p(t)$  et du modèle  $y_m(t)$  sont comparées pour créer le vecteur erreur  $e(t)$ . Puis, un algorithme d'optimisation est mis en place pour chercher les valeurs des paramètres du modèle  $\theta = [a_1, \dots, a_n, b_0, b_1, \dots, b_m]^T$  permettant de minimiser un critère de performance  $J(e)$ . Le critère est généralement choisi de forme quadratique.  $\hat{\theta}$  : vecteur des paramètres estimés.

Comparer aux méthodes non paramétriques, les méthodes paramétriques présentent les avantages suivants:

- La structure du modèle à identifier est spécifiée a priori, donc, une représentation plus précise;
- Un critère de performance est utilisé pour déterminer les paramètres du modèle, qui constitue une référence de validation;
- Les algorithmes ne sont pas influencés par les perturbations et les bruits de mesure;
- La possibilité d'une implémentation temps réel, ce qui permet un ajustement du régulateur pendant le fonctionnement du processus.

Les modèles d'identification utilisent le plus souvent une représentation discrète, bien que les processus physiques réels soient généralement de nature continu. D'une part, l'identification est effectuée à travers un calculateur numérique et d'une autre part l'identification des paramètres et la simulation du système sont plus simples sous forme numérique que sous forme analogique.

Dans ce chapitre nous considérons uniquement le cas des systèmes dynamiques linéaires discrets mono-variables.

Nous avons:

$$y_m(k) + \sum_{i=1}^n a_i y_m(k-i) = \sum_{j=1}^m b_j u(k-d-j), \quad k = 0,1,2,3, \dots \quad \text{II.1}$$

Où  $k$  est l'instant d'échantillonnage,  $d$  exprime le retard du système qui est un multiple entier de la période d'échantillonnage  $T_e$ ,  $n$  et  $m$  représentent respectivement de degré de deux polynômes  $A(z^{-1})$  et  $B(z^{-1})$ , tels que:

$$H(z^{-1}) = \frac{Y_m(z)}{U(z)} = z^{-d} \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} = z^{-d} \cdot \frac{B(z^{-1})}{A(z^{-1})} \quad \text{II.2}$$

Où:  $a_i$  et  $b_i$  sont les paramètres à identifier.

Opérateur retard:

Dans le but de simplifier l'écriture des équations aux différences, on introduit l'opérateur retard  $q^{-1}$  défini comme suit:

$$q^{-i} y(k) = y(k-i)$$

L'équation s'écrit alors:

$$A(q^{-1}) \cdot y_m(k) = B(q^{-1}) \cdot u(k-d) = q^{-d} B(q^{-1}) \cdot u(k) \quad \text{II.3}$$

Avec:

$$A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_n q^{-n}$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m}$$

On peut ainsi définir l'opérateur de transfert temporel  $G(q^{-1}) = \frac{y_m(k)}{u(k)} = q^{-d} \cdot \frac{B(q^{-1})}{A(q^{-1})}$

### 3. Critères de Performance:

On utilise le plus souvent un critère quadratique basé sur une mesure de distance (ou erreur) entre les valeurs mesurées et celle prédites à l'aide du modèle. Trois critères sont considérés:

#### 3.1. Critère basée sur l'erreur de mesure (sortie):

L'erreur de sortie  $e_s(k)$  est la différence entre la sortie mesurée  $y(k)$  et la sortie du modèle  $y_m(k)$  qui résultent de la même excitation  $u(k)$ , comme illustré sur la figure II.2.

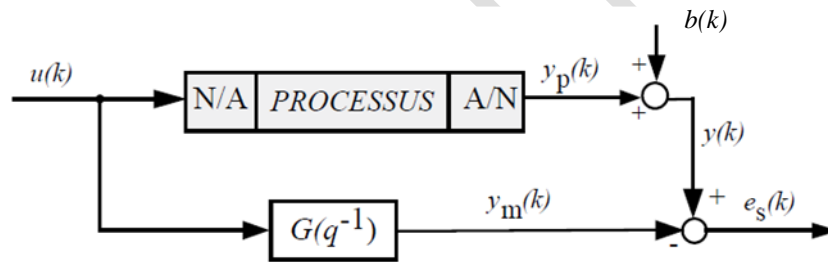


Figure II.2 Identification basée sur l'erreur de sortie

Nous avons:

$$e_s = y(k) - y_m(k) = y_p(k) + b(k) - y_m(k) \quad \text{II.4}$$

Pour le modèle:

$$G(q^{-1}) = \frac{y_m(k)}{u(k)} = q^{-d} \cdot \frac{B(q^{-1})}{A(q^{-1})} \quad \text{II.5}$$

L'équation aux différences correspondante s'écrit:

$$y_m(k) = -a_1 y_m(k-1) - \dots - a_n y_m(k-n) + b_0 u(k-d) + b_1 u(k-d-1) + \dots + b_m u(k-d-m), \quad k = 0,1,2,3, \dots \quad \text{II.6}$$

La sortie du modèle  $y_m(k)$  dépend des paramètres  $a_i$  et  $b_j$ , ( $i = 1,2, \dots, n; j = 0,1, \dots, m$ ). Comme  $y_m(k-1), \dots, y_m(k-n)$  dépendent également de ces mêmes paramètres,  $y_m(k)$  est une fonction non linéaire de ces paramètres. Il s'ensuit que l'erreur de sortie est une fonction non linéaire des paramètres à identifier. Considérons l'expression de l'erreur de sortie:

$$e_s(k) = y(k) - y_m(k) = y(k) - q^{-d} \cdot \frac{B(q^{-1})}{A(q^{-1})} \cdot u(k) \quad \text{II.7}$$

Les paramètres  $a_i$  et  $b_j$  interviennent de façon non linéaire dans cette expression.

L'identification, considère un critère basé sur la minimisation de l'erreur quadratique, par rapport à  $\theta = [a_1, \dots, a_n, b_0, \dots, b_m]^T$  telle que:

$$\min_{\theta} J(\theta) = \sum_{k=1}^N e_s^2(k) \quad \text{II.8}$$

Où :  $N$ : représente le nombre de mesures disponibles.

La difficulté majeure de cette approche réside dans le fait qu'il s'agit là d'un problème de régression non linéaire avec l'existence possible de minima locaux.

### 3.2. Critère basé sur l'erreur d'équation:

En général, la sortie mesurée  $y$  diffère de celle du modèle  $y_m$ . L'erreur d'équation  $e_e(k)$  représente l'erreur dans l'équation II.6 lorsque la sortie du modèle  $y_m$  est remplacée par la sortie mesurée  $y$ :

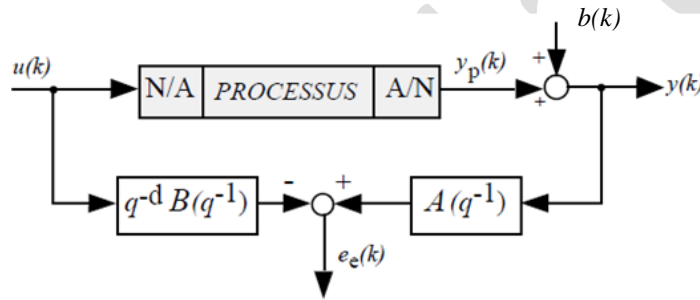


Figure II.3 Identification basée sur l'erreur d'équation

$$e_e(k) = y(k) + a_1 y(k-1) + \dots + a_n y(k-n) - b_0 u(k-d) - b_1 u(k-d-1) - \dots - b_m u(k-d-m), \quad k = 0,1,2,3, \dots \quad \text{II.9}$$

L'erreur d'équation peut également s'écrire comme suit:

$$e_e(k) = A(q^{-1}).y(k) - q^{-d}B(q^{-1})u(k) \quad \text{II.10}$$

Comme les échantillons  $u(k-d), \dots, u(k-d-m)$  sont donnés et les échantillons:  $y(k), \dots, y(k-n)$  sont mesurés, l'erreur d'équation  $e_e(k)$  est linéaire par rapport aux paramètres  $\theta$ , c'est à dire  $a_i$  et  $b_j$ , ( $i = 1,2, \dots, n; j = 0,1, \dots, m$ ).

Dans le cas d'un modèle correct, nous avons  $y_m(k) = y_p(k)$  et donc  $y(k) = y_m(k) + b(k)$ .

En remplaçant  $y(k)$  par  $y_m(k) + b(k)$  dans l'équation (II.10) et en utilisant l'équation II.5, on obtient:

$$e_e(k) = A(q^{-1}).b(k) \quad \text{II.11}$$

La minimisation de l'erreur d'équation quadratique s'écrit:

$$\min_{\theta} J(\theta) = \sum_{k=1}^N e_e^2(k) \quad \text{II.12}$$

Et constitue un problème de régression linéaire par rapport aux paramètres  $\theta = [a_1, \dots, a_n, b_0, \dots, b_m]^T$

### 3.3. Critère basé sur l'erreur de prédiction:

L'erreur de prédiction  $\varepsilon(k)$  est la différence entre la sortie mesurée  $y(k)$  et la sortie prédite  $\hat{y}(k)$ . La sortie prédite à l'instant  $k$  est calculée à partir de toutes les informations disponibles jusqu'à l'instant  $k - 1$ . Cette sortie peut être présentée telle que:

$$\hat{y}(k) = \mathcal{F}(\theta, u(k-1), u(k-2), \dots, y(k-1), y(k-2), \dots, \hat{y}(k-1), \hat{y}(k-2), \dots) \quad k = 0, 1, 2, 3, \dots \quad \text{II.13}$$

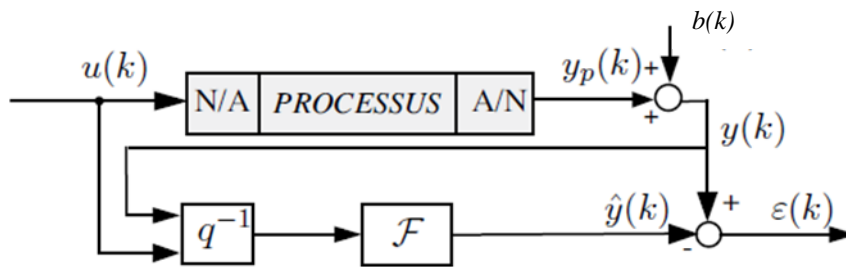


Figure II.4 Identification basée sur l'erreur de prédiction

Où  $\mathcal{F}$  est une fonction à définir. Il est à mentionner que l'erreur de sortie  $e_s(k)$  et l'erreur d'équation  $e_e(k)$  sont des cas particuliers de l'erreur de prédiction.

L'erreur d'équation peut être écrite comme une erreur de prédiction en choisissant:

$$\hat{y}(k) = -a_1 y(k-1) - \dots - a_n y(k-n) + b_0 u(k-d) + b_1 u(k-d-1) + \dots + b_m u(k-d-m), \quad k = 0, 1, 2, 3, \dots \quad \text{II.14}$$

Avec ce choix on obtient:

$$\varepsilon(k) = e_e(k) = y(k) - \hat{y}(k) \quad \text{II.15}$$

Même chose si on prend la sortie du modèle pour la sortie prédite  $\hat{y}(k) = y_m(k)$ , on obtient le critère de l'erreur de sortie.

L'ensemble des méthodes basées sur la minimisation de l'erreur de prédiction sont appelées méthodes des moindres. L'avantage de ces méthodes est d'être relativement simples à mettre en œuvre et de pouvoir être implantés en temps réel sur calculateur sous forme récursive.

## 4. Algorithme des Moindres Carrés:

### 4.1. Moindres Carrés Simples:

Considérons le cas de la minimisation du critère basé sur l'erreur d'équation en utilisant la méthode des moindres carrés. A partir de l'équation II.15, nous avons:

$$\varepsilon(k) = y(k) - \Phi^T(k) \cdot \theta \quad \text{II.16}$$

Avec:

$$\Phi^T(k) = [-y(k-1) \dots -y(k-n) \quad u(k-d) \dots u(k-d-m)] \quad \text{II.17}$$

$$\theta = [a_1 \dots a_n \quad b_0 \dots b_m]^T$$

Où:  $\Phi(k)$  représente le vecteur des variables explicatives.

Le terme  $\Phi^T(k) \cdot \theta$  représente la prédiction de  $y(k)$  obtenue à l'aide du modèle et des entrées et sorties mesurées. Le modèle est linéaire par rapport aux paramètres à identifier  $\theta$ .

L'algorithme des moindres carrés présenté ici est valable pour tout modèle qui peut s'écrire sous la forme II.16.

Si l'on effectue N mesures successives, de l'équation II.16 on obtient la relation matricielle:

$$\begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(N) \end{bmatrix} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} - \begin{bmatrix} \Phi^T(1) \\ \vdots \\ \Phi^T(N) \end{bmatrix} \cdot \theta \quad \text{II.17}$$

Que l'on note:

$$\varepsilon_N = Y_N - \Phi_N \cdot \theta \quad \text{II.18}$$

Avec les dimensions suivantes:

$\varepsilon_N$ : ( $N \times 1$ ) Vecteur d'erreur

$Y_N$ : ( $N \times 1$ ) Vecteur de mesures

$\Phi_N$ : ( $N \times p$ ) Matrice d'observation

$\theta$ : ( $p \times 1$ ) Vecteur de paramètres

La minimisation de l'erreur de prédiction quadratique s'écrit ainsi:

$$\min_{\theta} J(\varepsilon) = \min_{\theta} J(\varepsilon_N^T \varepsilon_N) \quad \text{II.19}$$

$$= \min_{\theta} ([Y_N - \Phi_N \theta]^T \cdot [Y_N - \Phi_N \theta]) = \min_{\theta} (Y_N^T Y_N - 2Y_N^T \Phi_N \theta + \theta^T \Phi_N^T \Phi_N \theta) \quad \text{II.20}$$

Le vecteur de paramètres  $\hat{\theta}$  qui minimise l'équation (II.20) annule le gradient de  $J$  par rapport à  $\theta$ :

$$\left. \frac{\partial J(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = -2Y_N^T \Phi_N + 2\Phi_N^T \Phi_N \hat{\theta} = 0 \quad \text{II.21}$$

D'où l'on tire:

$$\hat{\theta} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \quad \text{II.22}$$

Que l'on écrit également:

$$\hat{\theta} = \Phi_N^+ Y_N \quad \text{II.23}$$

Où:  $\Phi_N^+$  représente la matrice pseudo-inverse de  $\Phi_N$ .

Il est évident que la solution (II.22) existe si la matrice d'observation est de rang plein ( $\text{rang}(\Phi_N) = p$ ) et que le hessien de  $J$  par rapport à  $\theta$  évalué en  $\hat{\theta}$ , soit une matrice symétrique définie positive:

$$\left. \frac{\partial^2 J(\theta)}{\partial \theta \partial \theta^T} \right|_{\theta=\hat{\theta}} = 2\Phi_N^T \Phi_N > 0 \quad \text{II.24}$$

Ceci implique que  $\hat{\theta}$  corresponde bien à un minimum de  $J$ , qui peut être écrit sous une forme plus détaillée comme suit:

$$\hat{\theta} = [\sum_{k=1}^N \phi(k)\phi^T(k)]^{-1} [\sum_{k=1}^N \phi(k)y(k)] \quad \text{II.25}$$

De façon à obtenir une détermination unique de l'ensemble des paramètres, il est nécessaire d'effectuer un nombre de mesures  $N \geq 2n$

#### 4.2. Moindres Carrés Pondérés:

Une extension de cette méthode consiste à privilégier les mesures récentes ou les plus influencées par des facteurs d'oubli. Le critère de minimisation II.19 dans ce cas est modifié comme suit:

$$J_\lambda(\varepsilon) = \sum_{k=0}^N \lambda_k \varepsilon^2(k) \quad \text{II.26}$$

Où:  $\lambda$  est la matrice du facteur d'oubli, telle que:

$$\lambda^T \lambda = \begin{bmatrix} \lambda^{N-1} & & & 0 \\ 0 & \ddots & & \\ & & \lambda^1 & \\ 0 & & & \lambda^0 \end{bmatrix} \text{ avec } \lambda \leq 1 \text{ (souvent } 0.9 \leq \lambda \leq 0.99)$$

Le choix des coefficient élémentaires,  $\lambda_i$  ( $i = 1, \dots, N$ ), doit pondérer les mesures les plus anciennes qui possèdent une influence réduite dans le calcul des paramètres actuels. à travers l'introduction du facteur d'oubli  $\lambda$ .

L'erreur de prédiction sous forme matricielle devient:  $\varepsilon_{N\lambda} = \lambda[Y_N - \Phi_N \theta]$

Le vecteur des paramètres estimés devient alors:

$$\hat{\theta} = (\Phi_N^T \lambda^T \lambda \Phi_N)^{-1} \Phi_N^T \lambda^T \lambda Y_N \quad \text{II.27}$$

#### 4.3. Formulation récurrente de l'algorithme MCS (MCR):

Dans les méthodes précédentes, l'estimation de  $\hat{\theta}$  était réalisée à partir des mesures entrées-sorties:  $\{u(0), \dots, u(N), y(0), \dots, y(N)\}$ , et que pour  $N$  suffisamment grand, l'estimation devienne plus précise. Une formulation récurrente est proposée afin d'éviter les calculs redondants, permettant ainsi la simplification des calculs et l'implémentation sur calculateur.

En considérant les mesures jusqu'à l'instant  $kT$  et avec les notations suivantes similaires à celles de l'équation (II.17):

$$\varepsilon_k \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(k) \end{bmatrix}; \quad y_k = \begin{bmatrix} y(1) \\ \vdots \\ y(k) \end{bmatrix}; \quad \Phi_k = \begin{bmatrix} \phi^T(1) \\ \vdots \\ \phi^T(k) \end{bmatrix} \quad \text{II.28}$$

L'équation (II.25) donne:

$$\hat{\theta}_k = [\sum_{i=1}^k \phi(i)\phi^T(i)]^{-1} [\sum_{i=1}^k \phi(i)y(i)] \quad \text{II.29}$$

Où  $\hat{\theta}_k$  représente  $\theta$  à l'instant  $kT$  sur la base des  $k$  premières mesures ( $k$  premières équations). Lors de la connaissance de nouvelles mesures, le calcul de la nouvelle estime  $\hat{\theta}_{k+1}$  peut se faire par récurrence (sans inversion matricielle) à partir de (II.29) comme suit:

$$\hat{\theta}_k = P_k \sum_{i=1}^k \phi(i)y(i) \quad \text{II.30}$$

$$P_k = [\sum_{i=1}^k \phi(i)\phi^T(i)]^{-1} \quad \text{II.31}$$

Où  $P_k$  est l'inverse de la matrice d'information qui peut être recalculée de façon récurrente pour une nouvelle mesure par:

$$\begin{aligned} P_{k+1}^{-1} &= \sum_{i=1}^{k+1} \phi(i)\phi^T(i) = \sum_{i=1}^k \phi(i)\phi^T(i) + \phi(k+1)\phi^T(k+1) \\ &= P_k^{-1} + \phi(k+1)\phi^T(k+1) \end{aligned} \quad \text{II.32}$$

Donc le vecteur des paramètres  $\hat{\theta}_{k+1}$  peut s'écrire comme:

$$\hat{\theta}_{k+1} = P_{k+1} [\sum_{i=1}^{k+1} \phi(i)y(i)] = P_{k+1} [\sum_{i=1}^k \phi(i)y(i) + \phi(k+1)y(k+1)] \quad \text{II.33}$$

A partir de (II.30), on a :

$$\hat{\theta}_{k+1} = P_{k+1} [P_k^{-1}\hat{\theta}_k + \phi(k+1)y(k+1)] \quad \text{II.34}$$

En ajoutant et soustrayant le terme  $\phi(k+1)y(k+1)\hat{\theta}_k$  dans l'expression entre crochets de (II.34) et en tenant compte de (II.32), il devient:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + P_{k+1}\phi(k+1)[y(k+1) - \phi^T(k+1)\hat{\theta}_k] \quad \text{II.35}$$

Dans cette équation, le terme  $\varepsilon(k+1) = y(k+1) - \phi^T(k+1)\hat{\theta}_k$  représente l'erreur de prédiction pour la  $k+1$  mesure calculée à partir de la connaissance des paramètres  $\hat{\theta}_k$  (données par les  $k$  mesures), où le terme  $P_{k+1}\phi(k+1)$  est défini comme étant le gain d'adaptation.

### Procédure:

L'algorithme d'identification MCSR développée dans ce cas peut être résumé aux étapes suivantes:

- Mesurer les  $k$  valeurs de  $u(k)$  et  $y(k)$  et initialiser les paramètres  $\hat{\theta}_k$  et  $P_k$  suivant les équations (II.30) et (II.31);
- Mesurer la nouvelle valeur de  $u(k+1)$  et  $y(k+1)$  et calculer l'erreur de prédiction  $\varepsilon(k+1) = y(k+1) - \phi^T(k+1)\hat{\theta}_k$ ;
- Calculer  $P_{k+1}$  suivant l'équation (II.32), puis déduire le gain d'adaptation  $P_{k+1}\phi(k+1)$ ;



- d. Déterminer la nouvelle valeur estimée de  $\hat{\theta}_{k+1}$  suivant (II.35);
- e. Répéter les mêmes calculs pour l'itération suivante en allant à l'étape 2.

**Remarque 1:**

Pour cet algorithme, l'estimation des paramètres passe par l'inversion de la matrice  $P_{k+1}$  à chaque itération. Pour palier à cet inconvénient, une simplification est introduite en utilisant le lemme d'inversion matriciel, tel que:

$$P_{k+1} = P_k - \frac{P_k \phi(k+1) \phi^T(k+1) P_k}{1 + \phi^T(k+1) P_k \phi(k+1)} \quad \text{II.36}$$

Cette dernière équation a l'avantage de ne pas nécessiter l'inversion d'une matrice, mais uniquement celle d'un scalaire

**Avantages :**

Par rapport à l'algorithme MCS, qui traite les données par paquets, l'identification récurrente offre les avantages suivants:

- Estimation du modèle en temps réel;
- Compression importante des données, car l'algorithme récurrent traite à chaque instant une seule paire entrée-sortie au lieu de l'ensemble des données;
- N'exige pas des calculateurs à mémoire et processus de calcul puissant;
- Facilité d'implémentation en temps réel.

**Avantages :**

L'inconvénient majeur de cette méthode réside dans la décroissance de  $P_k$ ;  $(P_{k+1}^{-1} - P_k^{-1})$  : est une matrice définie positive dans (II.32), et qui peut devenir tellement petit. Ceci influence nettement sur la sensibilité du mécanisme d'identification (mécanisme d'adaptation des paramètres) pour l'ajustement du vecteur paramètres  $\hat{\theta}_{k+1}$  du modèle même si l'erreur de prédiction  $y(k+1) - \phi^T(k+1)\hat{\theta}_k$  est grande.

**Remarque 2:**

Le même principe peut être appliqué à l'algorithme des moindres carrés pondérés, ce qui aboutira à l'algorithme (MCPR).

**4.4. Limite de la méthode des moindres carrés:**

De façon générale, la méthode des moindres carrés basée sur l'erreur d'équation est simple à mettre en œuvre (régression linéaire) mais donne souvent des estimations biaisées car l'erreur d'équation est corrélée avec le régresseur.

L'estimation optimale des paramètres par la méthode d'identification MCS (sans facteur d'oubli), pour un processus représenté par le modèle réel suivant:

$$A_r(Z^{-1})y(k) = Z^{-d_r}B_r(Z^{-1})u(k) + e_r(k) \quad \text{II.37}$$

Où:  $A_r(Z^{-1})$  et  $B_r(Z^{-1})$  représentent les polynômes du modèle réel et  $e_r(k)$  est l'erreur d'équation due aux bruits de mesures.

Ainsi, la sortie de ce modèle peut s'écrire sous la forme régressive (II.18) comme suit:

$$\varepsilon_r = Y_N - \Phi_N \cdot \theta_r \quad \text{II.38}$$

Avec:

$$\varepsilon_r \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(N) \end{bmatrix}; \quad Y_k = \begin{bmatrix} y(1) \\ \vdots \\ y(kN) \end{bmatrix}; \quad \Phi_k = \begin{bmatrix} \phi^T(1) \\ \vdots \\ \phi^T(N) \end{bmatrix}$$

$$\text{et } \theta_r^T = [a_{r1}, \dots, a_{rn_r}, b_{r0}, \dots, b_{rm_r}]^T \quad \text{II.39}$$

Où:  $\theta_r$  est le vecteur des valeurs réels des paramètres du modèle.

Reprenons l'erreur de prédiction donnée (II.18) sous forme matricielle par:

$$\varepsilon_N = Y_N - \Phi_N \cdot \theta \quad \text{II.40}$$

Pour le cas d'une structure de modèle correcte ( $n = n_r; m = m_r$  et  $d = d_r$ ), et en combinant (II.38) et (II.40), on a :

$$\varepsilon_N = \Phi_N(\theta_r - \theta) + \varepsilon_r \quad \text{II.41}$$

La minimisation de cette erreur de prédiction quadratique permet d'obtenir:

$$\tilde{\theta} = \theta_r - \tilde{\theta} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T \varepsilon_r \quad .42$$

Comme  $\varepsilon_r$  est un vecteur de variables aléatoires, et le vecteur des paramètres estimés  $\hat{\theta}$  le sera aussi. Alors, l'espérance mathématique de l'erreur d'estimation peut être calculé par:

$$\begin{aligned} E\{\tilde{\theta}\} &= \theta_r + (E\{\Phi_N^T \Phi_N\})^{-1} E\{\Phi_N^T \varepsilon_r\} \\ &= \theta_r + \mathbf{R}_{\phi\phi}^{-1}(0) \mathbf{R}_{\phi\varepsilon_r}^{-1}(0) \end{aligned}$$

Ainsi, l'estimation donnée par la méthode des moindres carrés sera non biaisés si, d'une part  $\mathbf{R}_{\phi\phi}^{-1}(0)$  est régulière et d'une autre part  $\mathbf{R}_{\phi\varepsilon_r}^{-1}(0) = 0$ . La première condition est remplie si le système est suffisamment excité. La deuxième condition impose que le vecteur du bruit de mesures  $\varepsilon_r(k)$  du système ne soit pas corrélé avec le régresseur  $\Phi_N$ . Cette condition est vérifiée si  $e_r(k)$  est un bruit blanc, ce qui est très peu probable en réalité.