



# HUITIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 1<sup>er</sup> au 5 JUIN 1981

QUELQUES PROPRIETES IMPORTANTES DU MODELE AUTOREGRESSIF VECTORIEL

J.P. GAMBOTTO

E.T.C.A.  
16b, Av. Prieur de la Côte d'Or - 94114 ARCUEIL CEDEX -

## RESUME

Dans cet article nous étudions les méthodes d'estimation des moindres carrés et du maximum de vraisemblance pour le modèle autorégressif vectoriel. Plusieurs résultats nouveaux sont présentés ; nous montrons, en particulier, que comme dans le cas scalaire, les deux méthodes conduisent au même ensemble de paramètres estimés.

Nous démontrons ensuite une propriété intéressante de la formulation des moindres carrés et du maximum de vraisemblance : l'estimation du modèle autorégressif vectoriel est réalisée en minimisant l'erreur de prédiction pour chacune des composantes du signal vectoriel.

Plusieurs formes canoniques du modèle autorégressif vectoriel sont proposées, et nous étudions en détail, les propriétés algébriques et les propriétés de stabilité des algorithmes normalisés correspondants.

Nous présentons, ensuite, plusieurs critères de reconnaissance qui se déduisent des méthodes d'estimation précédentes.

## SUMMARY

In this paper the mean square and maximum likelihood estimation techniques for autoregressive vector models, are discussed. Some new results are presented ; in particular, it is shown that in the gaussian case both techniques give rise to the same minimization problem, leading to the same set of parameter estimates.

Another interesting property of the mean square and maximum likelihood formulation is the following : the fitted autoregressive vector model is obtained by minimizing the prediction error on each component of the vector signal.

We introduce several canonical forms for the autoregressive vector model, and study in detail the algebraic and stability properties of the corresponding normalized algorithms.

Several recognition criteria are then derived from the mean square and maximum likelihood estimation techniques.



## I Introduction

Le modèle vectoriel est utilisé lorsqu'il est nécessaire d'analyser simultanément plusieurs signaux, ayant ou non la même origine physique ; une meilleure modélisation est alors obtenue en utilisant l'information contenue dans l'intercorrélation des signaux.

Comme dans le cas scalaire une représentation simple du signal, est donnée par le modèle autorégressif (AR) :

$$A_{n,n} Y_m + A_{n,n-1} Y_{m-1} + \dots + A_{n,0} Y_{m-n} = U_m \quad (1)$$

où  $Y_m$  est l'échantillon d'un signal vectoriel stationnaire du second ordre de dimension  $r$  ; les matrices  $r \times r$   $A_{n,i}$  sont les paramètres du modèle AR vectoriel, et  $U_m$  est un bruit blanc de dimension  $r$  :

$$E \begin{bmatrix} U_m \\ U_m^T \end{bmatrix} = D_n \delta_{m\ell} \quad (2)$$

La matrice  $A_{n,n}$  est ici inversible ; cette restriction indique que l'on peut toujours calculer la meilleure estimée  $\hat{Y}_m$ . Une dégénérescence peut cependant se produire si  $Y_m$  n'est pas de rang complet (c'est-à-dire si la matrice  $D_n$  n'est pas de rang complet) ; cette possibilité de dégénérescence peut être importante en pratique ; en effet, le nombre de signaux à analyser simultanément est souvent arbitraire (cas d'un signal d'origine industrielle, par exemple, où le nombre de mesures en parallèle n'est pas limité) ;  $D_n$  peut donc être dégénérée si le vecteur  $Y_m$  a une dimension trop grande.

Malgré l'hypothèse précédente, la représentation (1) n'est pas unique ; il existe en effet plusieurs ensembles de paramètres  $\Theta = \{(A_{n,i} \ i = 1, \dots, n), D_n\}$  pour modéliser le signal  $Y_m$ .

Ce problème est particulièrement important pour l'identification des processus vectoriels [1]. Il est relativement compliqué pour le modèle autoregressif à moyenne mobile (ARMA) ; ce modèle se représente par sa fonction de transfert  $H(z) = B(z)/A(z)$  et plusieurs formes canoniques ont été trouvées en étudiant les propriétés de divisibilité de polynômes  $A(z)$  et  $B(z)$  [2] [3]. Dans le cas du modèle AR vectoriel (1) le problème est nettement plus simple, mais n'a pas été étudié de façon approfondie. Une forme canonique couramment utilisée se définit à partir de (1) en posant

$A_{n,n} = I$ . Nous présentons plusieurs autres normalisations pour le modèle AR vectoriel et étudions en détail les propriétés des algorithmes correspondants. Ces normalisations sont toutes équivalentes pour les problèmes d'estimation linéaire mais donnent en général des critères de reconnaissance différents. Nous étudions tout d'abord la normalisation  $A_{n,n} = I$ .

## II Les méthodes d'estimation

Plusieurs méthodes d'estimation peuvent être utilisées pour identifier le modèle vectoriel (1) : moindres carrés, maximum de vraisemblance, maximum d'entropie ... ; elles consistent toutes à minimiser un critère statistique sur l'ensemble des données ; ce critère s'exprime en fonction de la variance de l'erreur de prédiction  $\hat{U}_m = Y_m - \hat{Y}_m$  :

$$E \begin{bmatrix} \hat{U}_m \\ \hat{U}_m^T \end{bmatrix} \quad (3)$$

### II.1 La méthode des moindres carrés (MC)

Elle consiste à minimiser la trace de (3) ; dans le cas du modèle AR vectoriel (1), l'erreur de prédiction a pour variance :

$$Q = \sum_{i=0}^n \sum_{j=0}^n A_{n,n-i} E \begin{bmatrix} Y_{m-i} & Y_{m-j}^T \end{bmatrix} A_{n,n-j}^T \quad (4)$$

et en introduisant la matrice de corrélation :

$$R_{i-j} = E \begin{bmatrix} Y_{m-i} & Y_{m-j}^T \end{bmatrix} \quad (5)$$

les éléments diagonaux de la matrice  $Q$  s'écrivent :

$$Q_{kk} = \sum \sum (A_{n,n-i} R_{i-j} A_{n,n-j}^T)_{kk} \quad (6)$$

Remarquons que  $Q_{kk}$  ne dépend que des éléments  $\mathcal{Q}_k = \{(A_{n,i})_{ku} \ u = 1, \dots, r \ i = 0, \dots, n\}$  ; on en déduit que le minimum de la trace de (4) est obtenu en minimisant chaque élément diagonal  $Q_{kk}$  par rapport aux paramètres de l'ensemble  $\mathcal{Q}_k$  pour  $k = 1, \dots, r$ . On a donc :

$$\text{Min} [\text{Trace } Q] \Leftrightarrow \text{Min} [Q_{kk}] \quad k = 1, \dots, r \quad (7)$$

( $\mathcal{Q}_k \ k=1, \dots, r$ )                       $\mathcal{Q}_k$

L'équivalence (7) est une propriété remarquable ; elle montre que l'identification vectorielle est réalisée en minimisant l'erreur d'estimation sur chacune

des composantes du signal vectoriel, sans expliciter l'estimateur des autres composantes. Cette propriété permet d'utiliser le filtre AR vectoriel pour identifier les signaux bidimensionnels stationnaires du second ordre : les meilleures estimations de r modèles AR bidimensionnels (déduts du modèle AR vectoriel de dimension r) sont obtenues simultanément [4] .

A partir de (4) et (7) on obtient facilement le résultat bien connu : la méthode d'estimation par les MC est équivalente à la résolution du système d'équations de Yule Walker :

$$\underbrace{\begin{bmatrix} A_{n,n} & \dots & A_{n,0} \end{bmatrix}}_{\underline{A}_n} \underbrace{\begin{bmatrix} R_0 & R_1^T & \dots & R_n^T \\ R_1 & \dots & \dots & R_n^T \\ \vdots & \dots & \dots & \vdots \\ R_n & \dots & \dots & R_0 \end{bmatrix}}_{\underline{R}} = \underbrace{\begin{bmatrix} D_n & 0 & \dots & 0 \end{bmatrix}}_{\underline{D}} \quad (8)$$

où  $\underline{R}$  est une matrice Toeplitz par bloc, symétrique, formée à partir des matrices  $R_i$  ; la transposée par bloc de  $\underline{R}$  est définie par :

$$\underline{R}^v = \underline{L} \underline{R} \underline{L} \quad (9)$$

$\underline{L}$  est une matrice de permutation par bloc ( $\underline{L} \underline{L} = I$ ) de dimension  $(n+1) \cdot r$  ayant pour seuls éléments non nuls,  $(n+1)$  sous matrices identité de dimension r sur la seconde diagonale.

Le système d'équations (8) se résoud de façon récursive à l'aide de l'algorithme de Levinson généralisé LWR [5][6][7] ;  $Y_m$  est aussi la sortie du système linéaire rétrograde :

$$\underline{A}_{n,n}^* Y_m + \underline{A}_{n,n-1}^* Y_{m+1} + \dots + \underline{A}_{n,0}^* Y_{m+n} = \underline{U}_m^* \quad (10)$$

où les matrices  $r \times r$   $\underline{A}_{n,j}^*$  sont les paramètres du modèle rétrograde, et  $\underline{U}_m^*$  est un bruit blanc de dimension r :

$$E \begin{bmatrix} \underline{U}_m^* & \underline{U}_\ell^{*T} \end{bmatrix} = \underline{D}_n^* \delta_{m\ell} \quad (11)$$

A partir de (10) on obtient un système d'équations équivalent à (8) :

$$\begin{bmatrix} \underline{A}_{n,n}^* & \dots & \underline{A}_{n,0}^* \end{bmatrix} \underline{R}^v = \begin{bmatrix} \underline{D}_n^* & 0 & \dots & 0 \end{bmatrix} \quad (12)$$

et les récursions de l'algorithme LWR sont obtenues en combinant (8) et (12) :

$$\underline{A}_{n+1,i} = \underline{A}_{n,i-1} + K_n \underline{A}_{n,n-i}^* \quad (13)$$

$$\underline{A}_{n+1,i}^* = \underline{A}_{n,i-1}^* + \underline{K}_n \underline{A}_{n,n-i} \quad (14)$$

$$\underline{F}_n = \sum_{i=0}^n \underline{A}_{n,i} \underline{R}_{i+1}^T \quad (15)$$

$$\underline{K}_n = -\underline{F}_n \underline{D}_n^{*-1} \quad (16)$$

$$\underline{K}_n^* = -\underline{F}_n^T \underline{D}_n^{-1}$$

$$\underline{D}_{n+1} = \underline{D}_n + \underline{K}_n \underline{F}_n^T \quad (17)$$

$$\underline{D}_{n+1}^* = \underline{D}_n^* + \underline{K}_n^* \underline{F}_n$$

Les conditions initiales sont données par  $\underline{A}_{0,0} = \underline{A}_{0,0}^* = I$  et  $\underline{D}_0 = \underline{D}_0^* = R_0$

II.2 La méthode du maximum de vraisemblance

Considérons N observations d'un processus vectoriel  $Y_m$  stationnaire du second ordre de dimension r, et supposons que la densité de probabilité de l'observation  $Y_m$  par rapport aux observations précédentes et à  $\underline{Q}$  (où  $\underline{Q} = \{Q_i \ i = 1, \dots, r\}$ ) soit gaussienne. Le maximum de vraisemblance des N observations vectorielles par rapport à  $\theta$  est donné par [8] :

$$L(\theta) = -\frac{rN}{2} \log 2\pi - \frac{N}{2} \log (\det D) - \frac{1}{2} \sum_{i=1}^N U_i^T(\underline{Q}) D^{-1} U_i(\underline{Q}) \quad (17)$$

Nous avons supposé ici que (17) est vérifiée pour tout i, et que les variances des erreurs  $U_i(\underline{Q})$  sont toutes égales à D ; cette hypothèse signifie que l'estimateur  $\hat{Y}_i(\underline{Q})$  est stationnaire, c'est-à-dire que les conditions initiales sont rejetées à moins l'infini (hypothèse qui est en général valable pour des échantillons de grande dimension ; nous notons cependant que dans le cas scalaire des méthodes exactes d'identification par le MV ont été proposées [9] ). Le maximum de (17) par rapport aux paramètres  $\theta$  est obtenu en calculant :

$$\frac{\partial L(\theta)}{\partial D} = 0 \quad \text{et} \quad \frac{\partial L(\theta)}{\partial \underline{Q}} = 0 \quad (18)$$



La première relation donne directement la valeur de  $\hat{D}$  [10] :

$$\hat{D} = \frac{1}{N} \sum_{i=1}^N U_i (\mathbf{G}) U_i^T (\mathbf{G}) \quad (19)$$

et la deuxième équation est alors équivalente à minimiser la fonction :

$$J = \det \left( \frac{1}{N} \sum_{i=1}^N U_i U_i^T \right) \quad (22)$$

Pour un processus gaussien scalaire, les deux critères MC et MV sont donc équivalents : ils conduisent au même ensemble de paramètres  $\hat{\Theta}$  (obtenu en résolvant les équations de Yule Walker) ; cette propriété se généralise au cas vectoriel gaussien ; nous donnons ici les idées principales de la démonstration proposée par l'auteur ([11] page 145) :

Dans le cas d'un modèle AR vectoriel d'ordre  $n$ , on montre à partir de (4) que la matrice  $\hat{D}$  a la forme :

$$D = \underline{A}_n \underline{R} \underline{A}_n^T \quad (21)$$

On calcule ensuite la dérivée de  $J = \det \hat{D}$  par rapport à la sous matrice  $\underline{A}_n$  obtenue à partir de  $\underline{A}_n$  en supprimant la matrice  $\underline{A}_{n,n} = I$ . En utilisant les propriétés des dérivées de fonctions de matrices [12] il vient :

$$\frac{\partial J}{\partial \underline{A}_n} = \sum_{i,j} J (\hat{D}^{-T})_{ij} \frac{\partial D_{ij}}{\partial \underline{A}_n} \quad (22)$$

Le calcul des différents termes de cette équation matricielle donne :

$$\frac{\partial J}{\partial \underline{A}_n} = 2 J \hat{D}^{-T} \underline{A}_n \bar{\underline{R}} \quad (23)$$

où  $\bar{\underline{R}}$  est la sous-matrice de  $\underline{R}$ , obtenue en supprimant les  $r$  premières colonnes.  $\underline{Y}_m$  étant de rang complet,  $J$  n'est pas nul, et il vient :

$$\underline{A}_n \bar{\underline{R}} = 0 \quad (24)$$

(21) et (24) donnent alors le système d'équations de Yule Walker ; donc le résultat annoncé.

### III Formes canoniques pour le modèle AR vectoriel

Pour que le modèle AR vectoriel (1) soit identifiable, il est nécessaire de normaliser l'ensemble des paramètres  $\Theta$  de façon, à rendre unique la solution du système d'équations (8). La normalisation définie

par  $\underline{A}_{n,n} = I$  est la plus simple ;  $D_n$  est alors la variance du bruit d'entrée  $U_m$ . Nous présentons ici plusieurs autres normalisations qui se déduisent toutes de la précédente.

#### III.1 Introduction des normalisations

En écrivant les relations (8) et (12) pour  $n = 0, \dots, p$  on montre facilement que les récursions LWR sont équivalentes aux factorisations des matrices  $\underline{R}$  et  $\underline{R}^v$  sous la forme :

$$\underline{A} \underline{R} \underline{A}^T = \underline{D} \quad \underline{A}^v \underline{R}^v \underline{A}^{*T} = \underline{D}^* \quad (25)$$

où  $\underline{D} = \text{diag} (D_0, \dots, D_p)$   $\underline{D}^* = \text{diag} (\underline{D}_0^*, \dots, \underline{D}_p^*)$  et  $\underline{A}, \underline{A}^*$  sont des matrices triangulaires inférieures par bloc dont les bloc-lignes sont respectivement égales aux matrices  $\underline{A}_n$  et  $\underline{A}_n^*$ .

Les différentes normalisations se définissent alors à partir de la factorisation des matrices résidus :

$$D_n = H_n H_n^T \quad \underline{D}_n^* = \underline{H}_n^* \underline{H}_n^{*T} \quad n = 0, \dots, p \quad (26)$$

qui sont définies positives.

Les relations (25) deviennent :

$$\underline{C} \underline{R} \underline{C}^T = I \quad \underline{C}^v \underline{R}^v \underline{C}^{*T} = I \quad (27)$$

avec  $\underline{C} = \underline{H}^{-1} \underline{A}$ ,  $\underline{C}^* = \underline{H}^{-1} \underline{A}^*$  et  $\underline{H} = \text{diag} (H_0, \dots, H_p)$ ,  
 $\underline{H}^* = \text{diag} (\underline{H}_0^*, \dots, \underline{H}_p^*)$ .

Considérons la relation (16) :

$$\begin{aligned} D_{n+1} &= D_n + K_n F_n^T \\ &= H_n \left[ I + H_n^{-1} K_n F_n^T H_n^{-T} \right] H_n^T \end{aligned}$$

$$\text{en posant : } X_{n+1} = H_n^{-1} F_n \underline{H}_n^{*T} \quad (28)$$

et en utilisant (15), il vient :

$$D_{n+1} = H_n \left[ I - X_{n+1} X_{n+1}^T \right] H_n^T \quad (29)$$

et à partir de (16) on obtient de même :

$$\overset{*}{D}_{n+1} = \overset{*}{H}_n \left[ I - X_{n+1}^T X_{n+1} \right] \overset{*}{H}_n^T \quad (30)$$

$D_{n+1}$  et  $\overset{*}{D}_{n+1}$  étant 2 matrices définies positives il vient :

$$\begin{aligned} \left[ I - X_{n+1} X_{n+1}^T \right] &= Q_{n+1} Q_{n+1}^T \\ \left[ I - X_{n+1}^T X_{n+1} \right] &= \overset{*}{Q}_{n+1} \overset{*}{Q}_{n+1}^T \end{aligned} \quad (31)$$

$$\text{avec } Q_{n+1} = H_n^{-1} H_{n+1} \quad \text{et} \quad \overset{*}{Q}_{n+1} = \overset{*}{H}_n^{-1} \overset{*}{H}_{n+1} \quad (32)$$

Les relations (31) montrent que les matrices  $\left[ I - X_{n+1} X_{n+1}^T \right]$  et  $\left[ I - X_{n+1}^T X_{n+1} \right]$  sont définies positives (car  $Q_{n+1} Q_{n+1}^T$  et  $\overset{*}{Q}_{n+1} \overset{*}{Q}_{n+1}^T$  sont définies non négatives, et  $H_n, \overset{*}{H}_n$  n'ont pas de valeurs propres nulles). Les valeurs propres  $\lambda_j$  étant ordonnées par la relation :

$$\begin{aligned} |\lambda_n| &< |\lambda_{n-1}| < \dots < |\lambda_1|, \text{ il vient :} \\ \lambda_j \left( \left[ I - X_{n+1} X_{n+1}^T \right] \right) &= 1 - \lambda_j \left( X_{n+1} X_{n+1}^T \right) > 0 \\ j &= 1, \dots, r \end{aligned} \quad (33)$$

$X_{n+1} X_{n+1}^T$  étant une matrice définie non négative, on en déduit que les valeurs singulières de  $X_{n+1}$  sont inférieures à 1 :

$$0 \leq \lambda_j^{1/2} \left( X_{n+1} X_{n+1}^T \right) < 1 \quad (34)$$

### III.2 Les algorithmes LWR normalisés

Considérons maintenant les recursions de l'algorithme LWR et introduisons les nouveaux paramètres :

$$\begin{aligned} C_{n,i} &= H_n^{-1} A_{n,i} & \overset{*}{C}_{n,i} &= \overset{*}{H}_n^{-1} \overset{*}{A}_{n,i} \\ n &= 0, \dots, p & i &= 0, \dots, n \end{aligned} \quad (35)$$

en remplaçant ces paramètres dans les relations (13), et en remarquant que :

$$X_{n+1} = -H_n^{-1} K_n \overset{*}{H}_n \quad (36)$$

on obtient les nouvelles recursions :

$$C_{n+1,i} = Q_{n+1}^{-1} \left[ C_{n,i-1} - X_{n+1} \overset{*}{C}_{n,n-i} \right]$$

$$\overset{*}{C}_{n+1,i} = \overset{*}{Q}_{n+1}^{-1} \left[ \overset{*}{C}_{n,i-1} - X_{n+1}^T C_{n,n-i} \right] \quad (37)$$

Posons  $J_n = H_n^{-1} F_n$  et  $G_n = \overset{*}{H}_n^{-T}$  il vient :

$$J_n = \sum_0^n C_{n,i} R_{i+1}^T \quad (38)$$

$$X_{n+1} = J_n G_n \quad (39)$$

Dans ce qui suit, la décomposition de Choleski, d'une matrice A symétrique définie positive, est définie par  $A = A^{1/2} A^{T/2}$ , où  $A^{1/2}$  est une matrice triangulaire inférieure ayant tous ses éléments diagonaux positifs. La décomposition "duale de Choleski" est définie par  $A = 1/2 A^{T/2} A$  ou  $1/2 A$  est une matrice triangulaire supérieure ayant ses éléments diagonaux positifs.

Les différents modèles normalisés se définissent alors, en explicitant la forme des matrices  $H_n$  et  $\overset{*}{H}_n$  pour  $n = 0, \dots, p$  :

(a) -  $H_n, \overset{*}{H}_n$  triangulaires inférieures, avec des éléments diagonaux positifs : cette normalisation a été introduite par Morf et all. [13], et est équivalente à la décomposition de Choleski des matrices  $\underline{R}$  et  $\overset{\vee}{\underline{R}}$ . Les paramètres du modèle  $C_{n,n}$  et  $\overset{*}{C}_{n,n}$  sont triangulaires inférieures, et les conditions initiales de l'algorithme (37) sont données par  $C_{0,0} = \overset{*}{C}_{0,0} = R_0^{1/2}$  et  $G_0 = R_0^{-T/2}$ .

(b) -  $H_n, \overset{*}{H}_n$  triangulaires supérieures (éléments diagonaux positifs) : considérons les factorisations (27); en utilisant (9) elles se mettent sous la forme :

$$\begin{aligned} (\underline{L} \underline{C} \underline{L}) \overset{\vee}{\underline{R}} (\underline{L} \underline{C} \underline{L})^T &= I \\ (\underline{L} \overset{*}{\underline{C}} \underline{L}) \underline{R} (\underline{L} \overset{*}{\underline{C}} \underline{L})^T &= I \end{aligned} \quad (40)$$

ou  $\underline{L} \underline{C} \underline{L}$  et  $\underline{L} \overset{*}{\underline{C}} \underline{L}$  sont 2 matrices triangulaires supérieures ; cette normalisation correspond donc à la décomposition duale de Choleski des matrices de corrélation  $\underline{R}$  et  $\overset{\vee}{\underline{R}}$ .

Les matrices  $C_{n,n}$  et  $\overset{*}{C}_{n,n}$  sont triangulaires supérieures et les conditions initiales données par  $C_{0,0} = \overset{*}{C}_{0,0} = -1/2 R_0$  et  $G_0 = -T/2 R_0$ .

(c) -  $H_n$  triangulaire inférieure,  $\overset{*}{H}_n$  triangulaire supérieure (éléments diagonaux positifs) : ce modèle normalisé introduit en [14], a permis d'établir la



relation entre le filtre AR vectoriel, et le filtre AR bidimensionnel [15] L'algorithme associé correspond aux décompositions de Choleski, et duale de Choleski de la matrice R ; les matrices  $C_{n,n}$  et  $\overset{*}{C}_{n,n}$  sont triangulaires (respectivement inférieure et supérieure) ; les conditions initiales sont :  $C_{0,0} = R_0^{-1/2}$ ,  $\overset{*}{C}_{0,0} = -1/2 R_0$ ,  $G_0 = -T/2 R_0$ .

Considérons les relations (32) ; si  $H_n$  et  $\overset{*}{H}_n$  sont triangulaires pour  $n = 0, \dots, p$ , les matrices  $Q_{n+1}$  et  $\overset{*}{Q}_{n+1}$  sont aussi triangulaires ; on en déduit que, pour les 3 normalisations précédentes, les matrices  $Q_{n+1}$  et  $\overset{*}{Q}_{n+1}$  s'obtiennent à partir des factorisations de Choleski (ou duale de Choleski) de  $\begin{bmatrix} I - X_{n+1} & X_{n+1}^T \\ X_{n+1} & X_{n+1} \end{bmatrix}$  et  $\begin{bmatrix} I - X_{n+1}^T & X_{n+1} \\ X_{n+1} & X_{n+1} \end{bmatrix}$ . On utilise ensuite (3) pour calculer  $G_{n+1}$ .

(d)  $-H_n = P_n \Delta_n P_n^T$ ,  $\overset{*}{H}_n = \overset{*}{P}_n \overset{*}{\Delta}_n \overset{*}{P}_n^T$  avec  $P_n P_n^T = \overset{*}{P}_n \overset{*}{P}_n^T = I$  et  $\Delta_n, \overset{*}{\Delta}_n$  sont 2 matrices diagonales ayant pour éléments diagonaux, les racines des valeurs propres des matrices  $D_n$  et  $\overset{*}{D}_n$ .  $H_n, \overset{*}{H}_n$  et leurs inverses sont des matrices symétriques.  $P_{n+1}, \overset{*}{P}_{n+1}, \Delta_{n+1}, \overset{*}{\Delta}_{n+1}$  se calculent à chaque itération, à partir des relations (32) et (39) et permettent d'obtenir  $X_{n+2}, Q_{n+2}, \overset{*}{Q}_{n+2}$ .

Les conditions initiales se déduisent de la décomposition  $R_0 = P_0 \Delta_0^2 P_0^T$ .

D'autres normalisations peuvent être utilisées ; en particulier des variantes des normalisations (a), (b), (c) sont obtenues en mettant les factorisations (27) sous la forme :

$$\underline{W} \underline{R} \underline{W}^T = \underline{\Delta} \quad \overset{*}{\underline{W}} \overset{*}{\underline{R}} \overset{*}{\underline{W}}^T = \overset{*}{\underline{\Delta}} \quad (41)$$

où  $\underline{W}$  et  $\overset{*}{\underline{W}}$  sont encore triangulaires, mais avec "1" sur la diagonale, et  $\underline{\Delta}, \overset{*}{\underline{\Delta}}$  sont 2 matrices diagonales.

Ces formes canoniques sont très importantes, pour les problèmes d'estimation et de reconnaissance, car les éléments diagonaux des matrices  $\underline{\Delta}$  et  $\overset{*}{\underline{\Delta}}$  sont les variances des erreurs de prédiction de chacune des composantes du signal aux différents ordres.

#### IV Propriétés de structure et de stabilité

Dans le cas scalaire ( $r = 1$ ) la stabilité d'un filtre récursif d'ordre  $n+1$  peut être vérifiée en

inversant les récursions (13) ; l'algorithme correspondant est équivalent à celui de Schur Cohn [16], et le filtre est stable si tous les coefficients de corrélation partielle ( $K_0$  à  $K_n$ ) sont en module inférieurs à 1. Dans le cas vectoriel, la relation (34) montre que les valeurs singulières des matrices  $X_{n+1}$  sont inférieures à 1 ; ce résultat a été obtenu par Morf et all. [13] pour le modèle (a). Le théorème suivant permet d'obtenir un résultat plus précis :

Théorème 1 [17] : A étant une matrice  $n \times n$  on a la relation :

$$\lambda_n^{1/2}(AA^+) \leq |\lambda_j(A)| \leq \lambda_1^{1/2}(AA^+) \quad (42)$$

$$j = 1, \dots, n$$

A partir de (34) et (42) on montre alors que les valeurs propres de  $X_{n+1}$  sont en valeur absolue inférieures à 1 :

$$0 < |\lambda_j(X_{n+1})| < 1 \quad n = 0, \dots, p-1 \quad (43)$$

D'autre part, les matrices  $K_n$  et  $\overset{*}{K}_n$  s'expriment en fonction de la matrice  $X_{n+1}$  :

$$K_n = -H_n X_{n+1} \overset{*}{H}_n^{-1} \quad \overset{*}{K}_n = -\overset{*}{H}_n X_{n+1}^T H_n^{-1} \quad (44)$$

il vient :

$$K_n \overset{*}{K}_n = H_n (X_{n+1} X_{n+1}^T) H_n^{-1}$$

$$\overset{*}{K}_n K_n = \overset{*}{H}_n (X_{n+1}^T X_{n+1}) \overset{*}{H}_n^{-1} \quad (45)$$

A et B étant 2 matrices  $n \times n$  les matrices produit AB et BA ont les mêmes valeurs propres [17] ; on en déduit alors l'égalité :

$$\lambda(K_n \overset{*}{K}_n) = \lambda(\overset{*}{K}_n K_n) = \lambda(X_{n+1} X_{n+1}^T) = \lambda(X_{n+1}^T X_{n+1}) \quad (46)$$

Les matrices  $K_n \overset{*}{K}_n$  et  $\overset{*}{K}_n K_n$  sont donc définies non négatives et il vient :

$$0 \leq \lambda_j(K_n \overset{*}{K}_n) < 1 \quad (47)$$

$$j = 1, \dots, r \quad n = 1, \dots, p.$$

(47) généralise le résultat bien connu dans le cas

scalaire : les coefficients de corrélation partielle sont inférieurs à 1 en valeur absolue, et cette propriété est une condition nécessaire et suffisante de stabilité (si  $R_0$  est positif).

Dans le cas vectoriel nous allons démontrer l'équivalence entre :

- a)  $\underline{R}$  définie positive,
- b)  $\underline{\overset{*}{R}}$  définie positive,
- c)  $\{D_i \quad i = 0, \dots, p\}$  définie positive,
- d)  $\{\overset{*}{D}_i \quad i = 0, \dots, p\}$  définie positive,
- e)  $\{H_i \quad i = 0, \dots, p\}$  vérifient  $\lambda(H_i) \neq 0$
- f)  $\{\overset{*}{H}_i \quad i = 0, \dots, p\}$  vérifient  $\lambda(\overset{*}{H}_i) \neq 0$
- g) (43) est vérifiée et  $R_0$  définie positive
- h) (47) est vérifiée et  $\overset{*}{R}_0$  définie positive

Démonstration : en utilisant (9) on montre facilement que  $\underline{R}$  et  $\underline{\overset{*}{R}}$  ont mêmes valeurs propres donc a) est équivalente à b).

Considérons les factorisations (27) il vient :

$$\underline{R} = \underline{C}^{-1} \underline{C}^{-T} \quad \underline{\overset{*}{R}} = \underline{\overset{*}{C}}^{-1} \underline{\overset{*}{C}}^{-T} \quad (48)$$

$\underline{C}^{-1}$  et  $\underline{\overset{*}{C}}^{-1}$  étant 2 matrices triangulaires, en appliquant le théorème 1 à (48) il vient :

$$\lambda_n^{1/2}(\underline{R}) \leq |\lambda_j(H_n)| \leq \lambda_1^{1/2}(\underline{R}) \quad (49)$$

$$\lambda_n(\underline{\overset{*}{R}}) \leq |\lambda_j(\overset{*}{H}_n)| \leq \lambda_1^{1/2}(\underline{\overset{*}{R}})$$

donc e) et f) sont équivalentes à a) et b). (31) montre alors que les matrices  $D_n$  et  $\overset{*}{D}_n$  sont définies positives.

Considérons l'équation (29) pour  $n = 0, \dots, p$  l'équivalence entre e) et g) est alors évidente d'après les résultats précédents.

Le filtre AR vectoriel étant stable si et seulement si d) est vérifiée (voir [11] page 84), g) et h) sont aussi des conditions nécessaires et suffisantes de stabilité. Comme dans le cas scalaire cette propriété permet de vérifier la stabilité des modèles successifs obtenus par un algorithme LWR, mais ne peut pas être utilisée pour tester celle d'un polynôme matriciel quelconque d'ordre  $n+1$ , car il faut connaître simultanément les polynômes direct et rétrogradé à l'ordre  $n+1$  pour calculer les coefficients de corrélation partielle.

Les résidus sont décroissants avec l'ordre du modèle : dans le cas vectoriel cette propriété s'explique en utilisant les valeurs propres, les déterminants, ou bien les éléments diagonaux des matrices résidus .

En appliquant l'inégalité de Minkowski ( $\sqrt[n]{\det(A+B)} \geq \sqrt[n]{\det A} + \sqrt[n]{\det B}$  pour  $n = 1$ ) aux relations (16) plusieurs auteurs [18] [19] ont obtenu le résultat suivant :

$$\det D_{n+1} < \det D_n \quad \det \overset{*}{D}_{n+1} < \det \overset{*}{D}_n \quad (50)$$

cette propriété de décroissance du déterminant est aussi vérifiée par des matrices  $H_n$  et  $\overset{*}{H}_n$  définies positives.

$H_{n+1}$  se met sous la forme :

$$H_{n+1} = H_0 Q_1 \dots Q_n Q_{n+1} \quad (51)$$

Si les matrices  $H_i$  sont triangulaires, il vient :

$$\lambda_j(H_{n+1}) = \lambda_j(H_0) \prod_{i=1}^{n+1} \lambda_j(Q_i) \quad (52)$$

Les valeurs propres des matrices  $H_0$  et  $Q_i$  étant inférieures à 1, les valeurs propres des matrices  $H_n$  forment r suites décroissantes :

$$\lambda_j(H_{n+1}) < \lambda_j(H_n) < \dots < \lambda_j(H_0) \quad (53)$$

le même résultat est obtenu pour des matrices  $\overset{*}{H}_{n+1}$  triangulaires.

Considérons les relations (16) sous la forme

$$D_n = D_{n+1} + F_n \overset{*}{D}_n^{-1} F_n^T \quad (54)$$

$$\overset{*}{D}_n = \overset{*}{D}_{n+1} + F_n^T D_n^{-1} F_n$$

Les matrices  $F_n \overset{*}{D}_n^{-1} F_n^T$  et  $F_n^T D_n^{-1} F_n$  ainsi que les matrices résidus sont symétriques et définies non négatives ; elles ont donc des éléments positifs sur la diagonale ; on en déduit le résultat :

$$(D_{n+1})_{ii} < (D_n)_{ii} \quad (\overset{*}{D}_{n+1})_{ii} < (\overset{*}{D}_n)_{ii} \quad (55)$$

$$i = 1, \dots, r \quad n = 0, \dots, p$$

Les éléments diagonaux des matrices résidus



forment donc  $r$  suites décroissantes. Cette propriété ne semble pas avoir été remarquée auparavant ; elle est importante, car elle montre que l'erreur d'estimation sur chacune des composantes du signal, diminue lorsque l'ordre du modèle augmente. Bien que les méthodes d'estimation linéaire vectorielles consistent seulement à minimiser l'erreur globale (Trace ou déterminant de (3)), les résultats (7) et (55) montrent que l'estimation a lieu de façon indépendante sur chacune des composantes du signal.

#### V Critères de reconnaissance

La méthode de reconnaissance consiste à construire une distance entre le signal inconnu (ou son modèle) et un signal de référence (modèle de référence). Nous proposons ici plusieurs critères de reconnaissance qui diffèrent suivant la méthode d'estimation et suivant la forme canonique utilisée.

Pour le modèle AR les critères de reconnaissance se définissent à partir de la variance de l'erreur de prédiction :

$$\Delta(k) = W(k) \underline{R} W^T(k) \quad (56)$$

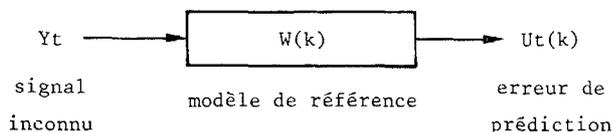
où  $\underline{R}$  est la matrice de corrélation calculée à partir du signal à tester, et  $W(k)$  est la matrice des paramètres du  $k$ -ième modèle de référence :

$$W(k) = \begin{bmatrix} W_{n,n} & \dots & W_{n,0} \end{bmatrix} (k) \quad (57)$$

Pour que  $\Delta(k)$  soit la variance de l'erreur de prédiction, il faut que les éléments diagonaux de  $W_{n,n}$  soient égaux à 1 ;  $W(k)$  est donc un des modèles normalisés défini par (41). Le modèle défini par  $W_{n,n} = I$  convient aussi, et au lieu de (56) on utilise dans ce cas la notation :

$$D(k) = A(k) \underline{R} A^T(k) \quad (58)$$

Au lieu de calculer la matrice de corrélation  $\underline{R}$  du signal, on peut aussi utiliser la méthode du filtrage inverse du signal par le modèle de référence :



De façon à avoir un critère indépendant de la variance du signal à tester, la matrice  $\underline{R}$  est normalisée par :

$$(\underline{R}_0)_{ii} = 1 \quad i = 1, \dots, r \quad (59)$$

Un premier critère lié à l'estimation par les MC est donné par :

$$\text{Trace } \Delta(k) \quad (60)$$

qui doit être comparé à la valeur minimale :

$$\text{Trace } \Delta = \text{Trace } W \underline{R} W^T \quad (61)$$

$\underline{W}$  étant le modèle optimal obtenu par identification du signal à tester. Un autre critère lié à l'estimation par le MV dans le cas gaussien a pour valeur :

$$\det \Delta(k) \quad (62)$$

Bien qu'elles correspondent à des méthodes d'estimation équivalentes (les minima sont obtenus pour le même modèle optimal  $k$ ), les deux fonctions (60) et (62) ne sont pas égales et donnent donc des critères de reconnaissance différents.

Nous remarquons aussi que les deux fonctions matricielles  $D(k)$  et  $\Delta(k)$  ne sont pas équivalentes ; en effet  $D(k)$  est la distance obtenue en projetant  $Y_t$  sur l'espace engendré par les vecteurs  $Y_{t-i}$  pour  $i = 1, \dots, p$  ; en revanche la distance  $\Delta(k)$  est obtenue en projetant chacune des composantes du vecteur  $Y_t$  sur l'espace engendré par les composantes de  $Y_t$  d'indices inférieurs et par les  $Y_{t-i}$   $i = 1, \dots, p$ .

Le critère Trace  $D(k)$  est donc différent des critères Traces normalisés (60). Par contre  $\det D(k)$  est équivalent aux critères déterminants normalisés (62) ; en effet on a la relation :

$$W(k) = W_{n,n} A(k) \quad (63)$$

en remarquant que  $W_{n,n}$  est triangulaire avec des 1 sur la diagonale, il vient :

$$\det \Delta(k) = \det D(k) \quad (64)$$

Nous avons donc montré que dans le cas vectoriel, la technique de reconnaissance utilisant le filtrage

inverse du signal par un modele de référence, permet l'utilisation de plusieurs critères différents de reconnaissance.

REFERENCES :

- [1] E.J. HANNAN, "The identification problem for multiple equation systems with moving average errors", *Econometrica*, Vol.39, n° 5 - Sep. 1971.
- [2] R.L. KASHYAP, R.E. NASBURG, "Parameter estimation in multivariate stochastic difference equations", *IEEE Trans. on Auto. Control*, VOL.AC-19, n° 6, Dec. 1974.
- [3] E.J. HANNAN, "The estimation of ARMA models", the *Annals of statistics*, Vol.3, n° 4, 975-981 - 1975.
- [4] J.P. GAMBOTTO, "Processus bidimensionnels et modèle autoregressif vectoriel : application à la modélisation des textures", 7ème colloque GRETSI, NICE - 1979.
- [5] N. LEVINSON, "The Wiener rms error criterion in filter design and prediction", *J. MATH Phys*, Vol 25, pp 261-278 - Jan. 1947.
- [6] R. WIGGINS, E. ROBINSON, "Recursive solution to the multichannel filtering problem", *J. Geophys. Res.* Vol.70, pp 1885-1891 - April 1965.
- [7] J. WHITTLE, "On the fitting of multivariate autoregressions and the approximate canonical factorization of the spectral density matrix", *Biometrika*, Vol.50, pp 129-134 - 1963.
- [8] T.W. ANDERSON, "Introduction to multivariate statistical analysis", NEW-YORK, J. Wiley sons, Inc - 1958.
- [9] C. GUEGUEN, L.L. SCHARF, "Exact maximum likelihood identification of ARMA models : a signal processing perspective", pp 759-769, *Signal Processing : Theories and Applications EURASIP* - 1980.
- [10] H. AKAIKE, "Maximum likelihood identification of gaussian autoregressive moving average models", *Biometrika*, Vol.60, n° 2, pp 255-265 - 1973.
- [11] J.P. GAMBOTTO, "Methodes d'estimation linéaire multidimensionnelles : application à la reconnaissance et à la segmentation des textures", These D.I., Dec. 1979, ENST, PARIS.
- [12] P.S. DWYER, "Some application of matrix derivatives in multivariate analysis", *Amer. Math. Assoc. Journal*, 606-625, June 1967.
- [13] M. MORF, A. VIEIRA, T.KAILATH, "Covariance characterization by partial autocorrelation matrices", *Ann. Stat.* Vol.6, pp 643-648 - 1978.
- [14] J.P. GAMBOTTO, C. GUEGUEN, "A multidimensional filtering approach to pattern recognition with application to texture classification and segmentation", *Int. Signal. Proc. Conf. Firenze* - 1978.
- [15] J.P. GAMBOTTO, "Two dimensional time series for textures", A paraitre dans *NATO ASI on Image Processing - REIDEL, HOLLAND* - 1980.
- [16] A. VIEIRA, T. KAILATH, "On another approach to the Schur-Cohn criterion", *IEEE Trans on Circuits and Systems*, April 1977.
- [17] M. MARCUS, "Basic theorems in Matrix theory", National Bureau of Standards, WASHINGTON DC ; *Applied Math. Series*, 37 - 1964.
- [18] D.C. YOULA, N.N. KAZANJIAN, "Bauer-type factorization of positive matrices, and the theory of matrix polynomials orthogonal on the unit circle", *IEEE on Circuits and Systems*, Vol CAS-25. n°2 Feb. 1978.
- [19] P. DELSARTE, Y.V. GENIN, Y.G. KAMP, "Orthogonal polynomial matrices on the unit circle", *IEEE Trans. on Circuits and Systems*. Vol CAS-25. n°3, March 1978.

