

STATISTIQUE ET ANALYSE DES DONNÉES

SERGE DEGERINE

Estimation autorégressive et traitement du signal

Statistique et analyse des données, tome 16, n° 2 (1991), p. 13-33

http://www.numdam.org/item?id=SAD_1991__16_2_13_0

© Association pour la statistique et ses utilisations, 1991, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**ESTIMATION AUTOREGRESSIVE
ET
TRAITEMENT DU SIGNAL**

Serge DEGERINE

Laboratoire LMC-IMAG
Université Joseph Fourier
BP 53X, F-38041 GRENOBLE Cedex

Résumé

Cet article présente des résultats récents obtenus en estimation autorégressive pour une situation fréquente en Traitement du Signal : séries courtes et modèle proche de la singularité. Il souligne le rôle essentiel des autocorrélations partielles dans ce domaine.

Mots clés : autocorrélation partielle, estimation autorégressive, maximum de vraisemblance, moindres carrés.

Classification AMS : 62 M 10, 62 M 15

Classification STMA : 12 070, 12 090

Abstract

This paper presents recent results achieved in autoregressive estimation for a frequent situation in Signal Processing : short series and nearly singular models. It emphasizes the main contribution of partial correlation coefficients in this field.

Keywords : partial autocorrelation, autoregressive estimation, maximum likelihood, least squares.

1. INTRODUCTION

Le modèle autorégressif joue un rôle important en traitement du signal. Il fournit une bonne approximation de la situation standard que constituent des sinusoides plongées dans un bruit blanc. D'où son utilisation dans les méthodes haute résolution d'analyse spectrale, qu'il s'agisse d'un phénomène temporel classique ou d'un phénomène spatial comme dans le traitement d'antennes. Un autre atout important est évidemment le fait qu'il conduit le plus souvent à des techniques linéaires. Cependant l'estimation du modèle se heurte à deux constantes inhérentes à ces situations : le modèle est proche de la singularité et les séries observées sont courtes. En effet c'est la proximité au cercle unité des pôles du filtre autorégressif qui permet d'obtenir des densités spectrales présentant

2. AUTOCORRELATIONS PARTIELLES ET MODELE AUTOREGRESSIF [Dégerine (1982)]

Les autocorrélations partielles sont peu utilisées dans la littérature statistique consacrée aux séries chronologiques par rapport à l'usage qui en est fait en traitement du signal (coefficients de réflexion). Elles ont été reconnues comme paramètres de la structure au second ordre d'une série d'abord pour le modèle autorégressif par Barndorff-Nielsen et Schou (1973) puis pour une situation quelconque par Ramsey (1974). Nous donnons ci-dessous un bref exposé sur ce sujet qui permet de préciser les notations qui seront utilisées par la suite.

Soit $x(\cdot) = \{x(t), t \in \mathbb{Z}\}$ une série chronologique scalaire réelle, stationnaire au second ordre et centrée. Il est pratique d'adopter une présentation géométrique en se plaçant dans l'espace de Hilbert réel $\overline{\mathcal{X}} \{x(t), t \in \mathbb{Z}\}$ engendré par cette série ; les variables étant centrées, le produit scalaire est défini par $\langle u, v \rangle = E(uv)$ et le carré de la norme représente une variance.

L'innovation partielle progressive d'ordre k est l'erreur de prédiction linéaire de $x(t)$ par son passé immédiat de longueur k :

$$\varepsilon(t; k) = \sum_{j=0}^k a(k, j)x(t-j), \quad a(k, 0) = 1, \quad \varepsilon(k, 0) = 1, \quad \varepsilon(t; 0) = x(t).$$

L'innovation partielle rétrograde, obtenue par inversion du temps et notée $\varepsilon^*(t; k)$, s'exprime à l'aide des mêmes coefficients $a(k, j)$ et a la même norme $\sigma(k)$ que $\varepsilon(t; k)$.

La fonction d'autocorrélation partielle $\beta(\cdot)$ définie par

$$\beta(k) = \frac{\langle \varepsilon(t; k-1), \varepsilon^*(t-k; k-1) \rangle}{\sigma^2(k-1)} \quad (0 \text{ si } \sigma(k-1) = 0), \quad k \in \mathbb{N},$$

caractérise, avec la variance $\sigma^2(0)$ de $x(t)$, la structure au second ordre de $x(\cdot)$ au même titre que la fonction d'autocovariance $\Lambda(k) = E\{x(t)x(t+k)\}$, $k \in \mathbb{N}$. C'est donc un paramètre équivalent à la fonction d'autocorrélation $\rho(\cdot) = \Lambda(\cdot) / \Lambda(0)$ mais dont les valeurs sont libres dans $]-1, +1[$ (pour $x(\cdot)$ non singulière) alors que $\rho(\cdot)$ est une fonction de type positif.

La récurrence fondamentale

$$\begin{aligned} \varepsilon(t; k) &= \varepsilon(t; k-1) - \beta(k)\varepsilon^*(t-k; k-1) \\ \varepsilon^*(t; k) &= \varepsilon^*(t; k-1) - \beta(k)\varepsilon(t+k; k-1) \end{aligned}$$

est à la base de l'algorithme de Levinson-Durbin

$$\begin{aligned} a(k, 0) &= 1, \quad a(k, k) = -\beta(k) \\ a(k, j) &= a(k-1, j) - \beta(k)a(k-1, k-j), \quad 1 \leq j \leq k-1 \\ \beta(k+1)\sigma^2(k) &= \Lambda(k+1) + \sum_{j=1}^k a(k, j)\Lambda(k+1-j) \\ \sigma^2(k+1) &= [1 - \beta^2(k+1)]\sigma^2(k) \end{aligned}$$

qui fournit les relations de passage entre $\Lambda(\cdot)$ (ou $\rho(\cdot)$) et $\beta(\cdot)$.

Le modèle autorégressif d'ordre p (AR(p))

$$\sum_{k=0}^p a(k)x(t-k) = \varepsilon(t), \quad a(0) = 1,$$

où $\varepsilon(\cdot)$ désigne un bruit blanc de variance σ_ε^2 , se caractérise par $\beta(k) = 0, k > p$ et $\beta(p) \neq 0$. Les paramètres du modèle sont soumis à la contrainte de stabilité du filtre autorégressif,

$$\phi(z) = \sum_{k=0}^p a(k)z^{p-k} \neq 0, \quad |z| \geq 1.$$

Les zéros de ce polynôme, appelés pôles du système, sont donc à l'intérieur du cercle unité et leur proximité au bord se traduit par des pics dans la densité spectrale

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} |\phi(e^{i\lambda})|^{-2}, \quad \lambda \in [-\pi, +\pi].$$

Les coefficients autorégressifs $a(k) = a(p, k)$, les autocorrélations $\rho(k)$ et les autocorrélations partielles $\beta(k), k = 1, \dots, p$ se correspondent par l'algorithme de Levinson-Durbin et $\sigma_\varepsilon^2 = \sigma^2(p)$.

3. MAXIMUM DE VRAISEMBLANCE

On suppose que l'on observe n séquences de longueur m , indépendantes, notées $x_j(1), \dots, x_j(m)$, $j = 1, \dots, n$, en omettant l'indice $j = 1$ dans le cas d'une seule séquence. Sous l'hypothèse d'un modèle AR(p) gaussien avec $p < m$, maximiser la vraisemblance équivaut à minimiser, par rapport à la matrice d'autocovariance Λ d'ordre m , la fonction

$$(3.1) \quad L(\Lambda; X) = \ln\{\det(\Lambda)\} + \text{tr}\{\Lambda^{-1}X\},$$

où X désigne la matrice de covariance empirique

$$X[i,k] = \frac{1}{n} \sum_{j=1}^n x_j(i)x_j(k), \quad i,k = 1, \dots, m.$$

La difficulté provient de ce que la fonction $L(\cdot; X)$ se présente comme la différence de deux fonctions convexes [Dégerine (1985)].

La situation générale modélisée ci-dessus recouvre deux situations standards : estimation d'une matrice de Toeplitz ($p = m-1$) en traitement d'antennes et estimation spectrale autorégressive ($n = 1$).

a) Exhaustivité.— D'après (3.1) la vraisemblance ne dépend des observations qu'à travers la covariance empirique X par le biais de la fonction $\text{tr}\{\Lambda^{-1}X\}$. Nous indiquons ci-dessous les diverses réductions possibles de la statistique X .

Soit \mathcal{S}_m l'espace vectoriel des matrices carrées symétriques d'ordre m muni du produit scalaire $\langle A, B \rangle = \text{tr}(AB)/m$. Le cône des matrices définies positives (resp. semi-définies positives) est noté \mathcal{S}_m^+ (resp. $\bar{\mathcal{S}}_m^+$); le même principe de notation sera utilisé

pour les sous-espaces de \mathcal{S}_m considérés ci-après. Une matrice carrée M est dite centrosymétrique lorsqu'elle vérifie $JMJ = M$ où J est la matrice de réflexion

$$J = \begin{bmatrix} (0) & \cdot & 1 \\ 1 & \cdot & (0) \end{bmatrix}.$$

L'opérateur J appliqué à un vecteur x inverse l'ordre de ses composantes. Les matrices centrosymétriques de \mathcal{S}_m sont persymétriques (symétriques par rapport à l'antidiagonale) et forment le sous-espace \mathcal{C}_m des matrices doublement symétriques. Les matrices de

Toeplitz symétriques d'ordre m , pour lesquelles $M[i,j]$ ne dépend que de $|i-j|$, constituent à leur tour un sous-espace \mathcal{C}_m de \mathcal{C}_m . Enfin $\mathcal{C}_m(p)$ désigne le sous-espace de \mathcal{C}_m formé des matrices vérifiant $M[i,j] = 0$ pour $|i-j| > p$ (matrices bande diagonale). Les espaces \mathcal{S}_m et \mathcal{C}_m sont stables par inversion mais ce n'est pas le cas de \mathcal{C}_m . Par contre l'hypothèse AR(p) correspond aux matrices Λ de \mathcal{C}_m^+ dont l'inverse appartient à $\mathcal{C}_m(p)$.

Une première réduction consiste à remplacer la matrice de covariance empirique X de $\bar{\mathcal{S}}_m^+$ par sa projection sur \mathcal{C}_m , ${}_sX = (X + JXJ)/2$. Cela équivaut à ajouter à l'ensemble des séquences observées celles obtenues par inversion du temps. La matrice ${}_sX$ est dans $\bar{\mathcal{S}}_m^+$. Elle peut être réduite à nouveau en la projetant sur $\mathcal{C}_m(p)$, cette nouvelle matrice n'est plus nécessairement dans $\bar{\mathcal{S}}_m^+$; on traduit simplement ici que le modèle AR(p)

dépend des produits entre les observations décalées dans le temps d'au plus p . Ces deux réductions résultent de l'invariance, par projection de X , du produit scalaire $\text{tr}\{\Lambda^{-1}X\}$. Une réduction plus conséquente [Pham and Degerine (1990)] peut être obtenue de la manière suivante. Posant $L = \sigma^2(p)\Omega$ la minimisation par rapport à Λ de $L(\Lambda;X)$ équivaut à celle de

$$(3.2) \quad L^*(\Omega;X) = \ln\{\det(\Omega)\} + m \ln\{\text{tr}(\Omega^{-1}X)\},$$

par rapport à Ω et $\sigma^2(p)$ est estimé en fonction de la solution $\hat{\Omega}$ par

$$\hat{\sigma}^2(p) = \frac{1}{m} \text{tr}(\hat{\Omega}^{-1}X).$$

La matrice Ω ne dépend que des paramètres autorégressifs $a(0) = 1, a(1), \dots, a(p)$ et posant $a(j) = 0$ pour $j > p$ la formule de Gohberg-Semencul

$$\Omega^{-1} = U^T U - V^T V = U U^T - V V^T,$$

où

$$U = \begin{bmatrix} a(0) & a(1) & \dots & a(m-1) \\ 0 & a(0) & \dots & a(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a(0) \end{bmatrix}, \quad V = \begin{bmatrix} a(m) & a(m-1) & \dots & a(1) \\ 0 & a(m) & \dots & a(2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a(m) \end{bmatrix}$$

montre que $\text{tr}(\Omega^{-1}X)$ s'exprime en fonction des innovations partielles d'ordre p . En effet pour une séquence $x = [x(1), \dots, x(m)]^T$, et en convenant de poser égales à zéro les valeurs $x(j)$ non observées, on peut écrire :

$$x^T \Omega^{-1} x = \sum_{t=1}^m \varepsilon^{*2}(t;p) - \sum_{t=m+1}^{m+p} \varepsilon^2(t;p) = \sum_{t=1}^m \varepsilon^2(t;p) - \sum_{t=p}^{-1} \varepsilon^{*2}(t;p)$$

Ainsi $\text{tr}(\Omega^{-1}X)$, qui apparaît comme une somme de formes quadratiques par rapport aux composantes de chaque séquence, est aussi une forme quadratique (pas nécessairement positive) par rapport aux paramètres autorégressifs. Après développement on obtient :

$$\text{tr}(\Omega^{-1}X) = \sum_{i,j=0}^p a(i) a(j) Q[i,j],$$

où la matrice Q satisfait $Q[i,j] = Q[j,i] = -Q[m-j,m-i]$ et est alors définie par

$$Q[i,j] = \sum_{k=1}^{m-i-j} X[k+i,k+j], \quad i+j < m.$$

b) Existence.— Sur le plan pratique les conditions $n \geq m/2$, pour l'estimation d'une matrice de Toeplitz, et $m > 3p/2$, dans l'estimation autorégressive au vu d'une seule séquence, sont des conditions d'existence presque sûre suffisantes à la fois simples, et réalistes [Dégerine (1987)]. En toute généralité, une condition nécessaire et suffisante d'existence presque sûre est que les n séquences observées ne soient pas simultanément interpolables par des sommes de sinusoides construites à partir d'au plus p fréquences distinctes [Dégerine (1991)]. Nous précisons ci-dessous ce résultat ainsi que les étapes de son obtention.

Les conditions d'existence sont liées au comportement de la fonction de vraisemblance au bord de son domaine de définition. Celui-ci n'est pas nécessairement composé de matrices singulières, l'hypothèse moyenne-mobile fournit un bel exemple de telles situations, il l'est par contre dans le cas autorégressif.

La limite non tangentielle de la fonction $L(\cdot; X)$ de (3.1) en une matrice singulière Λ_0 de $\bar{\mathcal{S}}_m^+$ satisfait [Dégerine, (1985)]

$$(3.3) \quad \lim_{\varepsilon \rightarrow 0^+} L\{\Lambda_0 + \varepsilon(\Lambda - \Lambda_0); X\} = +\infty \text{ si } \ker(\Lambda_0) \not\subset \ker(X), \quad -\infty \text{ sinon.}$$

Ainsi la condition $\ker(\Lambda_0) \subset \ker(X)$ est nécessaire mais l'exemple qui suit montre qu'elle n'est pas suffisante en toute généralité.

$$\Lambda_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Lambda_\varepsilon = \begin{bmatrix} e^{-2/\varepsilon} & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

lorsque ε tend vers zéro, Λ_ε tend vers Λ_0 et $L(\Lambda_\varepsilon; X) = \ln \varepsilon - 1/\varepsilon$ tend vers $-\infty$.

La propriété [Dégérine (1985)]

$$\ker(X) \cap \ker(\Lambda_0) = \{0\} \Rightarrow \lim_{\Lambda \rightarrow \Lambda_0} L(\Lambda; X) = +\infty$$

conduit à la condition suffisante $\ker(X) \cap \ker(\Lambda_0) = \{0\}$ beaucoup plus restrictive.

Dans le cas du modèle AR(p) la condition nécessaire $\ker(\Lambda_0) \subset \ker(X)$ s'avère suffisante pour presque tout X. La notion de décomposition spectrale des matrices de Toeplitz [Dégérine (1983)] permet d'expliciter cette condition et conduit surtout à une méthode de calcul pour l'estimation d'une telle matrice (cf. alinéa d) ci-après).

On appelle décomposition spectrale d'une matrice de \mathcal{C}_m^+ toute factorisation

$\Lambda = MDM^*$ dans laquelle les matrices

$$M = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 \\ e^{i\lambda_1} & \dots & e^{i\lambda_k} & \dots & e^{i\lambda_m} \\ \vdots & \dots & \vdots & \dots & \vdots \\ e^{i(m-1)\lambda_1} & \dots & e^{i(m-1)\lambda_k} & \dots & e^{i(m-1)\lambda_m} \end{bmatrix}, \quad D = \begin{bmatrix} d_1 & \dots & & & \\ & \ddots & & & \\ & & d_k & & \\ & & & \ddots & \\ & & & & d_m \end{bmatrix},$$

sont définies par des fréquences distinctes, $-\pi < \lambda_1 < \dots < \lambda_m \leq \pi$ et des masses d_k positives ou nulles, M^* désignant la matrice transposée conjuguée de M . Chaque fréquence λ_k , autre que 0 ou π , figure avec son opposé $-\lambda_k$ et les masses d_k correspondantes sont égales puisque les matrices Λ sont réelles. Soit F la mesure spectrale associée au couple (M, D) par $F(\{\lambda_k\}) = d_k$, $k = 1, \dots, m$. Les termes de la matrice d'autocovariance Λ sont donnés par la fonction d'autocovariance $\Lambda(\cdot)$, $\Lambda[i, j] = \Lambda(i-j)$, et la relation $\Lambda = MDM^*$ correspond à

$$\Lambda(k) = \int_{-\pi}^{+\pi} e^{ik\lambda} F(d\lambda), \quad k = 0, \dots, m-1.$$

Une matrice Λ de \mathcal{C}_m^+ admet exactement deux décompositions spectrales distinctes ${}_+M_+D_+M^*$ et ${}_M D_ M^*$. Ce sont celles des séries singulières d'ordre m dont les mesures spectrales ${}_+F$ et ${}_F$ sont définies par $\{\sigma^2(0); \beta(1), \dots, \beta(m-1)\}$, qui caractérise Λ , et le choix de ± 1 pour $\beta(m)$. Une matrice Λ_0 de rang $r < m$ constitue la matrice d'autocovariance d'ordre m d'une série singulière d'ordre r . L'unicité de la décomposition de Λ_0 peut être obtenue en retenant la matrice diagonale D carrée d'ordre r et la matrice M rectangulaire d'ordre $m \times r$ définies par la mesure spectrale de cette série.

Soit Λ_0 une matrice de Toeplitz symétrique singulière d'ordre m . La condition $\ker(\Lambda_0) \subset \ker(X)$ équivaut à $\text{Im}(X) \subset \text{Im}(\Lambda_0)$. Si r est le rang de Λ_0 et $-\pi < \lambda_1 < \dots < \lambda_r \leq \pi$ les fréquences de la décomposition spectrale MDM* évoquée plus haut, l'appartenance de x à l'image de Λ_0 s'exprime par sa décomposition en fonction des colonnes $M[.,k]$ de M :

$$x = \sum_{k=1}^r \gamma_k M[.,k] \Leftrightarrow x_t = \sum_{k=1}^r \gamma_k e^{i\lambda_k t}, \quad t = 1, \dots, m.$$

Cette condition équivaut aussi à ce que x satisfasse l'équation récurrente

$$(3.4) \quad \sum_{k=1}^r a(k)x_{t-k} = 0, \quad t = r+1, \dots, m; \quad \phi_r(z) = \sum_{k=1}^r a(k)z^{r-k} = \prod_{k=1}^r [z - e^{i\lambda_k}]$$

où les coefficients du polynôme ϕ_r sont obtenus par la récurrence de l'algorithme de Levinson-Durbin associée à une suite, d'ailleurs non unique, $\beta(1), \dots, \beta(r)$ satisfaisant $|\beta(k)| < 1, k = 1, \dots, r-1$ et $|\beta(r)| = 1$. Ces coefficients $a(0), \dots, a(r)$ sont symétriques ($\beta(r) = -1$) ou antisymétriques ($\beta(r) = +1$), ce qui implique que lorsque x satisfait (3.4), il en est de même pour la séquence inverse x_m, \dots, x_1 . Ainsi la condition nécessaire $\ker(\Lambda_0) \subset \ker(X)$ signifie que les séquences observées ne sont pas simultanément interpolables par une somme de sinusoides basées sur les r fréquences de Λ_0 puisque $\text{Im}(X)$ est le sous espace engendré par ces séquences.

Soit $d(X)$ le plus petit entier $r < m$ pour lequel les séquences observées satisfont simultanément une équation du type (3.4) dans laquelle le polynôme ϕ_r peut présenter des racines multiples. Cela équivaut à autoriser $|\beta(k)| \leq 1$ pour $k < r$ dans la récurrence de construction de ϕ_r . On pose $d(X) = m$ lorsqu'un tel r n'existe pas. Notons $\Lambda = \sigma^2(0)\Gamma$ la matrice d'autocovariance d'ordre m du modèle AR(p), $p < m$. La matrice d'autocorrélation Γ est caractérisée par les autocorrélations partielles $\beta(1), \dots, \beta(p)$ et la minimisation de $L(.,;X)$ équivaut à celle de la fonction

$$g(\Gamma; X) = \{\det(\Gamma)\}^{1/m} \text{tr}(\Gamma^{-1}X),$$

l'estimation de $\sigma^2(0)$ étant alors donnée en fonction de la matrice $\hat{\Gamma}$ minimisant cette fonction par

$$\hat{\sigma}^2(0) = \frac{1}{m} \text{tr}(\hat{\Gamma}^{-1}X).$$

Pour une matrice d'autocorrélation singulière Γ_0 située sur la frontière de l'hypothèse AR(p) on a [Dégerine (1991)]

$$d(X) > p \Rightarrow \lim_{\Gamma \rightarrow \Gamma_0} g(\Gamma; X) = +\infty.$$

Une séquence de longueur finie m satisfaisant une équation du type (3.4) avec racines multiples peut être approchée par des séquences satisfaisant une équation du même ordre avec racines simples en éclatant tout couple de racines égales en deux racines simples voisines. Les matrices X étant associées à un nombre fini de séquences, celles vérifiant $d(X) = r$ sans satisfaire, pour une matrice Λ_0 de rang r , $\text{Im}(X) \subseteq \text{Im}(\Lambda_0)$ constituent un ensemble d'intérieur vide et donc de probabilité nulle. Ainsi la condition nécessaire $\ker(\Lambda_0) \not\subseteq \ker(X)$ est suffisante pour presque tout X .

c) Unicité.— Alors que le problème d'existence est entièrement résolu, celui de l'unicité, pourtant primordial, ne l'est pas. Cette unicité est évidente pour un modèle AR(1) ou une matrice de Toeplitz d'ordre 2 puisque l'on dispose de la solution explicite dans ces deux situations (cf. alinéa d) ci-dessous). Dans le cas général elle paraît vraisemblable, par contre la solution de l'équation de vraisemblance peut ne pas être unique. C'est en particulier le cas dans l'estimation d'une matrice de Toeplitz [Dégerine (1992)]. Il est donc prudent, dans la mise en œuvre des méthodes de calcul, nécessairement itératives, de choisir une bonne initialisation.

d) Calcul.— La méthode du maximum de vraisemblance pour les modèles standards de séries chronologiques passe en général par des solutions sous-optimales y compris dans la mise en œuvre de techniques de gradient car les problèmes de dérivation par rapport aux paramètres habituels ne sont pas résolus ou peu contrôlables. Pour des séries longues et des modèles suffisamment réguliers ces approximations, ou d'autres méthodes

standards, sont satisfaisantes. Ce n'est plus le cas pour des séries courtes, surtout lorsque s'ajoute le problème de la "singularité".

Le premier calcul exact est obtenu dans le cadre de l'estimation d'une matrice de Toeplitz [Dégerine (1987)]. Il s'agit d'une méthode de relaxation basée sur le fait que le couple $({}_+M, {}_+M)$ constitue une paramétrisation des matrices d'autocorrélations Γ et que pour une matrice $\Lambda = MDM^*$, avec M connue, l'estimation de maximum de vraisemblance de D est unique et donnée par

$$(3.5) \quad \hat{D} = \text{Diag} \{M^{-1}XM^{*-1}\}$$

où $\text{Diag}(A)$ désigne la matrice diagonale définie par les éléments diagonaux de A . Pour $\Lambda = {}_+M{}_+D{}_+M^* = {}_+M{}_+D{}_+M^*$ on note ${}_+\hat{\Lambda}$ et ${}_- \hat{\Lambda}$ les matrices définies par (3.5). Partant d'une matrice Λ_0 de \mathcal{C}_m^+ le procédé itératif se poursuit comme suit :

$$\begin{aligned} \Lambda_1 &= {}_+\hat{\Lambda}_0 \text{ ou } \Lambda_1 = {}_- \hat{\Lambda}_0, \\ \Lambda_{k+1} &= {}_+\hat{\Lambda}_k \text{ si } \Lambda_k = {}_- \hat{\Lambda}_k, \text{ sinon } \Lambda_{k+1} = {}_- \hat{\Lambda}_k, \quad k \geq 1. \end{aligned}$$

Notons que les matrices Λ_k sont calculées de façon explicite, sans avoir à déterminer les fréquences λ_k des matrices M , et que les propriétés de convergence de la méthode sont démontrées.

La méthode ci-dessus ne résout pas le cas des modèles $AR(p)$, $p < m-1$. Dans un premier temps la méthode de Newton-Rhapon est appliquée sur la fonction $L^*(\Omega; X)$ de (3.2) qui ne dépend que des autocorrélations partielles. Le calcul numérique exact des deux premières dérivées par rapport à $\beta(k)$, $k = 1, \dots, p$, utilise la dérivation des relations de l'algorithme de Levinson-Durbin [Dégerine and Pham (1987)]. La méthode permet de bien contrôler le domaine de variation des paramètres par rapport à celles utilisant les coefficients autorégressifs $a(k)$. Cependant, dans les situations proches de la singularité, il est difficile d'accrocher un voisinage de la solution dans lequel la fonction $L^*(.; X)$ reste convexe.

La solution est finalement apportée par une méthode de relaxation portant directement sur les autocorrélations partielles [Pham (1988)]. La dérivée de la fonction $L^*(\Omega; X)$ par rapport à chaque $\beta(k)$, les autres étant fixés, s'annule en une seule valeur de $]-1, +1[$ correspondant à un minimum ; de plus cette valeur est une solution identifiable d'une équation du troisième degré. Cette méthode sera notée MV dans le paragraphe consacré aux simulations.

4. AUTOCORRELATION PARTIELLE EMPIRIQUE ET ESTIMATION AUTOREGRESSIVE [Dégérine (1993)]

On introduit ici une estimation naturelle des autocorrélations partielles. Cela conduit à une nouvelle méthode d'estimation autorégressive comparable à celle de Yule-Walker et à d'autres méthodes de type moindres carrés.

a) **Autocorrélation partielle empirique.**— On considère une seule séquence de longueur m , $x(1), \dots, x(m)$. Les $(m-k)$ séquences progressives de longueur $(k+1)$ constituent les lignes d'une matrice d'ordre $(m-k) \times (k+1)$ notée $S_{k+1} : S_{k+1}[i,j] = x(i+j-1)$. Les colonnes de cette matrice, que nous noterons $s_k(t)$, $t = 1, \dots, k+1$, constituent, dans \mathbb{R}^{m-k} , une "représentation empirique" de la structure au second ordre d'une séquence de longueur $(k+1)$ issue du modèle régissant la séquence $x(1), \dots, x(m)$ observée. La matrice de covariance empirique,

$$X_{k+1} = \frac{1}{m-k} [S_{k+1}]' [S_{k+1}] ,$$

est, selon le produit scalaire

$$\langle u, v \rangle = \frac{1}{m-k} \sum_{t=1}^{m-k} u(t)v(t), \quad u, v \in \mathbb{R}^{m-k},$$

la matrice de Gram des vecteurs $s_k(t)$, $t = 1, \dots, k+1$, et constitue un estimateur sans biais de la matrice d'autocovariance Λ_{k+1} d'ordre $k+1$. En particulier le produit scalaire

$$\langle s_k(1), s_k(k+1) \rangle = \frac{1}{m-k} \sum_{t=1}^{m-k} x(t)x(t+k) ,$$

est un estimateur sans biais de $\Lambda(k)$, mais la matrice de Toeplitz associée à ces estimations, lorsque k varie, n'est pas nécessairement positive ; c'est pourquoi on considère les autocovariances empiriques biaisées,

$$(4.1) \quad \hat{\Lambda}(k) = \frac{1}{m} \sum_{t=1}^{m-k} x(t)x(t+k) .$$

Pour éviter cet artifice, on peut estimer directement l'autocorrélation partielle d'ordre k par la "corrélation partielle" (au sens du produit scalaire \langle , \rangle) entre $s_k(1)$ et $s_k(k+1)$ dans l'ensemble $\{s_k(1), s_k(2), \dots, s_k(k), s_k(k+1)\}$:

$$(4.2) \quad \frac{\langle \varepsilon_k(k+1;k-1), \varepsilon_k^*(1;k-1) \rangle}{\|\varepsilon_k(k+1;k-1)\| \|\varepsilon_k^*(1;k-1)\|}$$

où, par analogie avec les innovations, $\varepsilon_k(k+1;k-1)$ est la différence entre $s_k(k+1)$ et sa projection sur le sous-espace engendré par $\{s_k(2), \dots, s_k(k)\}$, $\varepsilon_k^*(1;k-1)$ étant défini de façon analogue.

Il est clair que cette méthode est réflexive car le changement de $x(1), \dots, x(m)$ en $x(m), \dots, x(1)$ ne change pas la valeur de (4.2). En fait nous ne retiendrons pas cette expression mais une version encore plus réflexive. En effet la structure de la séquence $s_k(1), \dots, s_k(k+1)$ dans \mathbb{R}^{m-k} n'est pas la même que celle de $s_k(k+1), \dots, s_k(1)$. En particulier dans (4.2) les normes de $\varepsilon_k(k+1;k-1)$ et de $\varepsilon_k^*(1;k-1)$ sont en général distinctes et la valeur de (4.2) ne coïncide donc pas avec le coefficient de $s_k(1)$ dans la "régession" de $s_k(k+1)$ sur $s_k(k), \dots, s_k(1)$. Cela correspond au fait que la matrice de covariance empirique X_{k+1} n'est pas doublement symétrique comme Λ_{k+1} . Or (4.2) n'est fonction que de X_{k+1} et il est alors naturel de remplacer cette matrice par la matrice doublement symétrique,

$${}_s X_{k+1} = \frac{1}{2} [X_{k+1} + JX_{k+1}J].$$

La matrice ${}_s X_{k+1}$ est la matrice de Gram, dans $\mathbb{R}^{2(m-k)}$, associée aux vecteurs colonnes de la matrice ${}_s S_{k+1}$ dont les lignes représentent l'ensemble des séquences de longueur $(k+1)$ observées dans le sens progressif ou rétrograde. On notera encore $s_k(t)$, $t = 1, \dots, k+1$, les colonnes de ${}_s S_{k+1}$. Les filtres donnant les erreurs de prédiction progressive ou rétrograde d'ordre k sont alors identiques,

$$\varepsilon_k(k+1;k) = \sum_{j=0}^k a_k(k,j) s_k(k+1-j), \quad \varepsilon_k^*(1;k) = \sum_{j=0}^k a_k(k,j) s_k(1+j),$$

et ces erreurs ont la même norme,

$$\sigma_k(k+1;k) = \|\varepsilon_k(k+1;k)\| = \sigma_k^*(1;k) = \|\varepsilon_k^*(1;k)\|.$$

C'est également le cas pour les erreurs d'ordre $(k-1)$, $\varepsilon_k(k+1;k-1)$ et $\varepsilon_k^*(1;k-1)$ et, notant $\eta_k(k+1;k-1)$ et $\eta_k^*(1;k-1)$ les erreurs normalisées correspondantes, l'autocorrélation partielle empirique d'ordre k est définie par

$$(4.3) \quad \hat{\beta}(k) = \langle \eta_k(k+1;k-1), \eta_k^*(1;k-1) \rangle, \quad k < 2m/3.$$

La condition $k < 2m/3$ équivaut à $(k+1) \leq 2(m-k)$ et garantit que la matrice ${}_sX_{k+1}$ est presque sûrement régulière, elle assure ainsi $\hat{\beta}(k) \neq \pm 1$.

b) Estimation autorégressive.— La méthode de Yule-Walker consiste à retenir le modèle AR(p) défini par les autocovariances biaisées $\hat{\Lambda}(k)$, $k = 1, \dots, p$, données par (4.1). Les paramètres du modèle sont alors obtenus par l'algorithme de Levinson-Durbin.

La méthode ACPE, pour AutoCorrélation Partielle Empirique, consiste à retenir le modèle AR(p) défini par $\hat{\beta}(k)$, $k = 1, \dots, p$, donnés par (4.3) et l'autocovariance empirique $\hat{\Lambda}(0)$ de (4.1) pour l'estimation de $\sigma^2(0)$. Cette méthode, comme celle de Yule-Walker, ne résulte pas globalement d'un critère de minimisation. Elle s'apparente cependant à certaines méthodes de type moindres carrés classiques en traitement du signal.

Considérons en effet les problèmes de la minimisation simultanée des erreurs progressives et rétrogrades d'ordre k pour $k = 1, \dots, p$. On cherche le filtre $\phi_k = \sum a(j) z^{p-j}$ solution de

$$\mathcal{P}_k : \text{Min}_{\phi_k} \frac{1}{2(m-k)} \sum_{t=1}^{m-k} [\varepsilon(t+k;k)^2 + \varepsilon^*(t;k)^2],$$

$$\varepsilon(t;k) = \sum_{j=0}^k a(j)x(t-j), \quad \varepsilon^*(t;k) = \sum_{j=0}^k a(j)x(t+j).$$

La méthode des Moindres Carrés Progressifs et Rétrogrades (MCPR) [Nuttal (1976), Ulrych and Clayton (1976)] retient le modèle AR(p) défini par la densité spectrale associée au filtre solution $\hat{\phi}_p$ à l'ordre p et à la valeur du minimum pour l'estimation de $\sigma^2(p)$. Notons qu'en l'absence de contraintes supplémentaires le filtre $\hat{\phi}_p$ n'est pas nécessairement stable.

La méthode de Burg(1968) retient le filtre obtenu à l'ordre p par résolution récursive des problèmes \mathcal{P}_k : le filtre d'ordre k est défini en fonction du filtre d'ordre $k-1$ solution de l'étape précédente ; ceci en assure la stabilité. La variance $\sigma^2(0)$ est estimée par $\hat{\Lambda}(0)$ de (4.1).

La méthode ACPE apparaît alors comme une solution intermédiaire : le filtre d'ordre p est défini par les autocorrélations partielles successives $\hat{\beta}(k)$, $k = 1, \dots, p$, comme pour la méthode de Burg, d'où sa stabilité ; mais à chaque étape, $\hat{\beta}(k) = \hat{a}(k)$, est l'opposé du dernier coefficient du filtre solution de \mathcal{P}_k sans contrainte.

c) **Remarques.**— La méthode ACPE semble coûteuse puisqu'elle nécessite p décompositions de Cholesky. En fait l'algorithme rapide de Marple (1980), pour le calcul du filtre à l'ordre p de la méthode MCPR, passe par le calcul des filtres des ordres inférieurs et contient ainsi en lui la résolution de la méthode ACPE.

La méthode proposée par Atal (1977) ou Dickinson (1978, 1979) constitue une approximation de la méthode ACPE qui consiste à estimer les autocorrélations partielles d'ordre k à partir des colonnes de la matrice des séquences de longueur $p+1$: $\beta(k)$ est estimé par la "corrélation partielle" entre $s_p(k+1)$ et $s_p(1)$ dans l'ensemble $\{s_p(1), \dots, s_p(k+1)\}$. Le calcul nécessite alors qu'une seule décomposition de Cholesky.

L'extension à un échantillon $x_j(1), \dots, x_j(m)$, $j = 1, \dots, n$, est immédiate. La matrice ${}_{ns}S_{k+1}$ est formée de l'ensemble des séquences de longueur $(k+1)$ observées dans le sens progressif ou rétrograde, elle est d'ordre $2n(m-k) \times (k+1)$ et la solution est obtenue à partir de la décomposition de Cholesky de la matrice de covariance empirique ${}_{ns}X_{k+1}$ associée. Dans cette situation, l'algorithme rapide ne peut pas être utilisé. L'extension des autres méthodes s'obtient également sans difficulté [Pham and Dégerine (1990)].

5. COMPARAISON DES METHODES [Dégerine (1993)]

Il n'existe pratiquement pas de résultats théoriques sur la qualité des estimateurs définis par les différentes méthodes, celle du maximum de vraisemblance (MV) comprise, hormis quelques résultats asymptotiques dans le cas du modèle AR(1). De plus, ces méthodes étant le plus souvent destinées à opérer sur des séries courtes, les propriétés asymptotiques perdent de leur intérêt. Avant de procéder à des comparaisons par simulation il est utile de faire quelques remarques sur le principe même des méthodes.

a) **Remarques.**— Nous considérons l'estimation des modèles AR(p) au vu d'une séquence de longueur m ; la transposition des commentaires au cas des échantillons et en particulier à l'estimation d'une matrice de Toeplitz ($p = m-1$) est évidente lorsqu'ils conservent un sens.

Notons tout d'abord que les cinq méthodes retenues sont réflexives au sens où le modèle AR(p) estimé est le même au vu de la séquence $x(1), \dots, x(m)$ ou de la séquence réfléchie $x(m), \dots, x(1)$. Seule la méthode MCPR ne garantit pas la stabilité du filtre initialement estimé. Les méthodes de Yule-Walker, Burg et ACPE sont récursives selon l'ordre p du modèle, c'est-à-dire que le modèle AR(k), $k < p$, sous-jacent au modèle AR(p) estimé est celui de la solution à l'ordre k ; ce n'est pas le cas pour les deux autres méthodes. Chaque méthode s'interprète comme un opérateur de l'ensemble des matrices doublement symétriques d'ordre m dans le sous-ensemble des matrices de Tœplitz d'ordre m associées aux modèles AR(p); alors seule la méthode de Yule-Walker ne constitue pas en ce sens un projecteur.

b) Simulations.— Elles sont réalisées à partir d'un fichier de 60 000 nombres pseudo-aléatoires de loi normale centrée réduite obtenu par la méthode de Box-Muller à l'aide du générateur,

$$y(i+1) = \{69\,069y(i) + 1\} \bmod(2^{32}), \quad y(0) = 3\,579.$$

Les méthodes sont comparées dans le domaine temporel. Pour chaque simulation, et pour chaque méthode, les estimations satisfont les relations qui existent entre leurs homologues théoriques; on a par exemple,

$$\hat{\sigma}^2(p) = \hat{\sigma}^2(0) \prod_{k=1}^p [1 - \hat{\beta}^2(p)],$$

et les données concernant MCPR sont celles qui résultent de la stabilisation du filtre solution lorsque cela est nécessaire.

Pour les paramètres scalaires $\sigma^2(0)$ et $\sigma^2(p)$, on indique le biais moyen observé sur r répétitions ainsi que la racine carrée de l'écart quadratique moyen. Pour les paramètres vectoriels $\beta(\cdot)$, $\rho(\cdot)$ et $a(\cdot)$, nous indiquons une synthèse des résultats obtenus sur les composantes scalaires,

$$\sqrt{\frac{1}{p} \sum_{k=1}^p \left[\frac{1}{r} \sum_{i=1}^r \hat{\theta}_i(k) - \theta(k) \right]^2}, \quad \sqrt{\frac{1}{p} \sum_{k=1}^p \frac{1}{r} \sum_{i=1}^r [\hat{\theta}_i(k) - \theta(k)]^2}$$

Les paramètres des modèles simulés, ainsi que les valeurs de m , n , p et r , sont indiqués dans les tables. Il s'agit de deux modèles autorégressifs proches de la singularité. La densité spectrale du modèle AR(4) comporte deux pics voisins et celle du modèle AR(8) deux fois deux pics voisins.

Nous comparons tout d'abord l'autocorrélation partielle empirique (ACPE) avec l'autocorrélation empirique (Yule-Walker) pour de longues séries. Les deux méthodes sont très semblables quant à l'estimation des autocorrélations $r(\cdot)$ et utilisent la même estimation de $\sigma^2(0)$; par contre la méthode de Yule-Walker donne des résultats désastreux pour les autres paramètres. Les autres méthodes ne se différencient pas d'ACPE sur des séries aussi longues.

TABLE I
BIAIS ET RACINE CARREE DE L'ECART QUADRATIQUE MOYEN
DES ESTIMATIONS

$$\beta(\cdot) : 0.716 ; -0.982 ; 0.704 ; -0.924 ; \sigma^2(0) = 7.617$$

$$m = 1000 \quad n = 1 \quad p = 4 \quad r = 25$$

Paramètres	YULE-WALKER		ACPE	
	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}
$\beta(\cdot)$	0.318	0.348	0.003	0.010
$\rho(\cdot)$	0.002	0.021	0.001	0.021
$a(\cdot)$	1.185	1.290	0.007	0.023

Nous constatons déjà que les méthodes doivent être comparées selon plusieurs critères car la non linéarité des relations entre les paramètres ne permet pas de trancher au vu des résultats obtenus sur un seul d'entre eux. On peut envisager des méthodes hybrides : en traitement du signal, il est fréquent de choisir une méthode pour la localisation des pics et une autre pour l'estimation de l'énergie associée.

Ensuite les méthodes, excepté Yule-Walker, sont comparées sur des séries courtes. La méthode de Burg surestime $\sigma^2(p)$ et ce biais important s'accompagne d'un fort écart quadratique moyen ; c'est la traduction du défaut de structure de la méthode que représentent les contraintes dans l'estimation récursive des $\beta(k)$; on retrouve ce défaut sur les paramètres $\beta(\cdot)$ et $a(\cdot)$ dans le cas du modèle AR(8) (Table III) ou de l'estimation de la matrice de Toeplitz (Table IV) mais pas pour le modèle AR(4) (Table II). Les pourcentages de polynômes instables $\hat{\phi}_p(\cdot)$ dans MCPR sont respectivement égaux à 35%, 14%, 80%, 60%, 41% et 12% dans les Tables II à IV. Les mauvais résultats concernant l'estimation de la variance $\sigma^2(0)$ de la série sont clairement liés à ce problème : sans contraintes, les racines du polynôme $\hat{\phi}_p(\cdot)$ peuvent être très proches du cercle unité. Cependant la méthode donne de bons résultats pour l'ensemble des autres paramètres.

TABLE II
BIAIS ET RACINE CARREE DE L'ECART QUADRATIQUE MOYEN
DES ESTIMATIONS

$\beta(\cdot)$: 0.716 ; -0.982 ; 0.704 ; -0.924 ; $\sigma^2(0) = 7.617$

Paramètres	BURG		ACPE		MV		MCPR	
	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}
m = 10 n = 1 p = 4 r = 2000								
$\beta(\cdot)$	0.077	0.157	0.081	0.185	0.073	0.182	0.096	0.184
$\rho(\cdot)$	0.073	0.204	0.060	0.200	0.020	0.118	0.028	0.126
$a(\cdot)$	0.358	0.590	0.364	0.622	0.289	0.580	0.368	0.611
$\sigma^2(0)$	0.081	7.358	0.081	7.358	0.434	6.740	7.273	131.17
$\sigma^2(4)^*$	11.02	40.56	-3.48	5.55	-3.67	5.55	-4.38	5.79
m = 20 n = 1 p = 4 r = 1000								
$\beta(\cdot)$	0.046	0.093	0.047	0.101	0.040	0.092	0.050	0.100
$\rho(\cdot)$	0.028	0.119	0.023	0.118	0.007	0.081	0.007	0.089
$a(\cdot)$	0.188	0.319	0.184	0.321	0.144	0.285	0.177	0.314
$\sigma^2(0)$	-0.030	6.005	-0.030	6.005	0.063	5.249	7.513	48.74
$\sigma^2(4)^*$	2.06	10.71	-1.50	3.47	-1.78	3.47	-1.95	3.53

* L'unité de $\sigma^2(4)$ est égale à 10^{-3}

TABLE III
BIAIS ET RACINE CARREE DE L'ECART QUADRATIQUE MOYEN
DES ESTIMATIONS

$\beta(\cdot)$: 0.716 ; -0.982 ; 0.704 ; -0.924 ; 0.012 ; -0.980 ; 0.000 ; -0.950 ; $\sigma^2(0) = 7.617$

Paramètres	BURG		ACPE		MV		MCPR	
	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}
m = 15 n = 1 p = 8 r = 2000								
$\beta(\cdot)$	0.216	0.268	0.072	0.175	0.074	0.212	0.154	0.218
$\rho(\cdot)$	0.083	0.283	0.076	0.295	0.005	0.165	0.016	0.156
$a(\cdot)$	1.723	2.907	0.754	1.756	0.341	1.920	2.091	3.024
$\sigma^2(0)$	-0.168	6.523	-0.168	6.523	0.344	6.095	-6.385	6.216
$\sigma^2(8)^*$	211.6	583.6	-1.74	2.43	-1.30	4.69	-2.40	2.64
m = 30 n = 1 p = 8 r = 1000								
$\beta(\cdot)$	0.165	0.201	0.047	0.101	0.044	0.095	0.089	0.135
$\rho(\cdot)$	0.007	0.176	0.007	0.176	0.022	0.155	0.019	0.168
$a(\cdot)$	1.459	1.982	0.277	0.651	0.213	0.559	0.262	0.572
$\sigma^2(0)$	-0.127	5.465	-0.127	5.465	0.059	5.709	0.652	39.188
$\sigma^2(8)^*$	27.96	37.09	-0.66	1.24	-0.76	1.41	-0.82	1.29

* L'unité de $\sigma^2(8)$ est égale à 10^{-5}

TABLE IV
BIAIS ET RACINE CARREE DE L'ECART QUADRATIQUE MOYEN
DES ESTIMATIONS

$\beta(\cdot)$: 0.716 ; -0.982 ; 0.704 ; -0.924 ; $\sigma^2(0) = 7.617$

Paramètres	BURG		ACPE		MV		MCPR	
	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}	Moy	(EQM) ^{1/2}
m = 5 n = 3 p = 4 r = 1000								
$\beta(\cdot)$	0.103	0.212	0.037	0.165	0.031	0.181	0.055	0.166
$\rho(\cdot)$	0.019	0.114	0.019	0.113	0.014	0.090	0.058	0.178
$\alpha(\cdot)$	0.512	0.733	0.195	0.546	0.151	0.575	0.216	0.550
$\sigma^2(0)$	0.074	4.485	0.074	4.485	0.606	5.964	1.842	44.524
$\sigma^2(4)^*$	8.50	19.91	-4.79	8.62	-4.84	7.55	-7.00	7.94
m = 5 n = 10 p = 4 r = 1000								
$\beta(\cdot)$	0.034	0.084	0.012	0.071	0.003	0.041	0.009	0.048
$\rho(\cdot)$	0.005	0.054	0.005	0.054	0.002	0.038	0.004	0.057
$\alpha(\cdot)$	0.171	0.293	0.064	0.226	0.015	0.130	0.029	0.154
$\sigma^2(0)$	-0.012	2.436	-0.012	2.436	0.031	2.244	29.03	314.13
$\sigma^2(4)^*$	4.24	8.48	-1.31	4.54	-1.38	4.26	-2.10	4.15

* L'unité de $\sigma^2(4)$ est égale à 10^{-3}

Finalement MV et ACPE sont les seules méthodes qui ne présentent pas de défauts importants. La méthode du maximum de vraisemblance est certainement la plus performante mais au prix d'un coût de calcul largement supérieur aux autres méthodes.

Références

Atal, B.S. (1977) On Determining Partial Correlation Coefficients by the Covariance Method of Linear Prediction. Presented at the 94th Meeting, Acoust. Soc. Amer.

Barndorff-Nielsen, O. and Schou, G. (1973) On the Parametrization of Autoregressive Models by Partial Autocorrelations. J. Multivariate Analysis, 3, pp.408-419.

Burg, J.P. (1968) A New Analysis Technique for Time Series Data. Presented at NATO Advanced Study Institute on Signal Processing, Enschede, Netherlands.

Dégerine, S. (1982) Partial Autocorrelation Function for a Scalar Stationary Discrete-Time Process. Proc. 3rd. Franco-Belgian Meeting of Statisticians. Bruxelles : Publications des Facultés Universitaires Saint-Louis, pp.79-94.

Dégerine, S. (1983) Décompositions spectrales d'une matrice d'autocovariance et applications. C. R. Acad. Sci. Paris, t. 296, Serie I, pp.565-568.

Dégerine, S. (1985) Comportement au bord et caractérisation d'un maximum pour la vraisemblance d'un vecteur aléatoire gaussien centré avec contrainte sur sa structure de covariance. C. R. Acad. Sci. Paris, t. 301, Série I, n° 5, pp.233-236.

Dégerine, S. (1987) Maximum Likelihood Estimation of Autocovariance Matrices from Replicated Short Time Series. J. Time Series Anal. Vol. 8, n° 2, pp.135-146,.

Dégerine, S. (1991) A Necessary and Sufficient Condition for the Existence of the Maximum Likelihood Estimate in Autoregressive Models. Soumis à IEEE, Trans. Signal Processing.

Dégerine, S. (June 1992) On Local Maxima of the Likelihood Function for Toeplitz Matrix Estimation. A paraître dans IEEE, Trans. Signal Processing.

Dégerine, S. (January 1993) Sample Partial Autocorrelation Function. A paraître dans IEEE, Trans. Signal Processing.

Dégerine, S. and Pham D.T. (1987) Une nouvelle approche du calcul de l'estimateur de maximum de vraisemblance exact d'un modèle autorégressif. Présenté au 11-ième Colloque GRETSI sur le traitement du signal et des images. Nice.

Dickinson, B.W. (July 1978) Autoregressive Estimation Using Residual Energies Ratios. IEEE Trans. Information Theory, Vol. IT-24, pp. 503-506.

Dickinson, B.W. (April 1979) Estimation of Partial Correlation Matrices Using Cholesky Decomposition. IEEE Trans. Automatic Control, Vol. AC-24, n° 2, pp.302-305.

Marple, S.L. Jr. (August 1980) A New Autoregressive Spectrum Analysis Algorithm. *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-28, pp.441-454.

Nuttal, A.H. (March 1976) Spectral Analysis of a Univariate Process with Bad Data Points, via Maximum Entropy and Linear Predictive Techniques. Naval Underwater Systems Center, Technical Report. TR-5303, New London, Conn..

Pham, D.T. (August 1988) Maximum Likelihood Estimation of Autoregressive Model by Relaxation on the Reflection Coefficients. *IEEE, Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp.1363-1367.

Pham, D.T. and Dégerine, S. (January 1990) Efficient Computation of Autoregressive Estimates Through a Sufficient Statistic. *IEEE, Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, pp.175-177.

Ramsey, F. (1974) Characterization of the Partial Autocorrelation Function. *Ann. Statist.* n°2, pp.1296-1301.

Ulrych, T.J. and Clayton, R.W. (1976) Time Series Modeling and Maximum Entropy, *Phys. Earth and Plan. Int.*, Vol. 12, pp.188-200.