IEEE Third ASSP Workshop
on Spectrum Estimation and Modeling

Boston, MA
November 17-18, 1986

LIDS-P-1628

PARAMETRIC SPECTRAL ESTIMATION FOR ARMA PROCESSES *

José M. F. Moura

Dept. of Electrical and Computer Eng.
Carnegie-Mellon University
Pittsburgh, PA 15213, USA

M. Isabel Ribeiro

CAPS, Instituto Superior Técnico
Av. Rovisco Pais, 1
P-1096 Lisboa Codex, Portugal

Abstract

In this paper we present an algorithm that decouples the autoregressive (AR) and the moving average (MA) estimation procedures in the ARMA parameter identification problem. The technique dualizes the roles of the AR and MA components of the process, exploring the linear dependencies between successively higher order innovation and prediction error filter coefficients associated with the process. A least squares recursive implementation of this algorithm and preliminary results on order estimation are discussed. Some simulated examples show the estimator performance for several ARMA(p,q) processes.

1. INTRODUCTION

The problem of identification of the parameters of an autoregressive moving-average process with p poles and q zeros, ARMA (p,q) is of considerable importance. In the absence of the moving average component, well tested techniques exist that lead to good pole estimation performance,(Refs. 1-4).Most of these methods explore the linear dependence between successive autocorrelation lags of the process.

When the power spectrum exhibits zeros, the poles of the process are still estimated with an acceptable performance by the usual techniques (see e.g. Ref. 1). However, the litterature on the subject widely recognizes that the available methods yield significant degraded estimates for the zeros. One of the underlying reasons stems from the two pass type character of the algorithms where the first pass estimates the AR parameters and filters the autoregressive component of the process and the second part identifies the MA parameters.

The present work describes an ARMA estimation algorithm that decouples the AR and MA estimation procedures, dualizing, in a certain sense, the roles of the AR and MA components of the process, which are independently estimated without corrupting each other. The AR component is obtained from the linear dependence exhibited by successively higher order innovation filter coefficients. The MA parameters are asymptotically estimated from the linear combinations satisfied by successively higher order prediction error filter coefficients. The AR and MA estimation schemes are implemented in a recursive way, the recursiveness standing for the filter's order.

This ARMA parametric spectral estimation technique departs from the usual approach of using autocorrelation estimates, to use the Burg technique (Ref. 4) that provides the mean square estimation of the prediction error filter coefficients. Those are the elements of the lower triangular matrix $\tilde{W}_N^{-1}$, which is the square root inverse of the Toeplitz covariance matrix

$$R_N = \tilde{W}_N \tilde{W}_N^T .$$

A recursive inversion of $\tilde{W}_N^{-1}$ leads to the innovation filter coefficients.

The underlying theory concerning this ARMA estimation algorithm is discussed, for the multivariable case, in Ref. 5; a nonrecursive implementation of the algorithm has been compared in Ref. 6 with an alternative scheme that uses the estimates of d-step ahead predictor coefficients. The emphasis here is on how the proposed algorithm performs under several ARMA(p,q) processes. Simulated examples show the transient effect associated with the MA component estimation and the existing tradeoff between this asymptotic behavior and the estimation errors on the prediction and innovation filter coefficients.

Preliminary experiments using an order determination algorithm are also presented. They are based on the migration pattern of the pole and zero configurations as higher order (purely) MA processes and higher order (purely) AR processes are fitted to the ARMA process.

2. OBSERVATION PROCESS MODEL

The observation process is assumed to be a scalar, zero mean, stationary sequence, of the class of the ARMA(p,q) processes, i.e. satisfying the linear difference equation

$$y_n + \sum_{i=1}^{p} a_i y_{n-i} = b_o e_n + \sum_{i=1}^{q} b_i e_{n-i}, \qquad (1)$$

where it is assumed that i) $\{e_n\}$ is a white, Gaussian, zero mean noise sequence with unit variance, ii) $q \le p$, $a_p \neq 0$, $b_q \neq 0$, iii) the polynomial matrices

*The work of the first author while at LIDS and MIT, on leave from IST (Portugal), was supported by the Army Research Office under contract DAAG-29-84-K-005. The work of the second author was supported by INIC (Portugal).

$$A(z)=1+\sum_{i=1}^{p} a_i z^{-i} \quad , \quad B(z)=\sum_{i=0}^{q} b_i z^{-i} \qquad (2)$$

are asymptotically stable and have no common roots.

Define the process autocovariance at lag k, as $r(k)=E[y_{n+k}y_n]$. There exists a class of state-variable models whose output is equivalent, in the sense defined in Ref. 6, to the process $\{y_n\}$ given by (1). We will select the element of the class

$$x_{n+1}=F\,x_n+g\,e_n \qquad (3)$$

$$y_n=h\,x_n+b_o\,e_n \qquad (4)$$

with minimal dimension, i.e. $x \in R^p$, completely controllable (c.c.) and completely observable (c.o.), and where F is an asymptotically stable, nonsingular matrix. The initial state x_o is a Gaussian random vector, independent of $\{e_n\}$, with zero mean and covariance matrix P_o, satisfying

$$P_o=F\,P_oF^T+g\,g^T. \qquad (5)$$

The ARMA estimation algorithm is based on the successively higher order prediction and innovation filter coefficients. Those are related with the Kalman-Bucy filter associated with (3)-(4)

$$\hat{x}_{n+1}=F\,\hat{x}_n+k_nv_n, \quad n\geq 0 \ , \quad \hat{x}_o=0, \qquad (6)$$

$$v_n=y_n-h\,\hat{x}_n \qquad (7)$$

where the variance of the innovation process v_n is given by

$$d_n=h\,P_n\,h^T+b_o^2\,, \qquad (8)$$

being related to the discrete Riccati equation

$$P_{n+1}=F\,P_n\,F^T-k_n\,d_n\,k_n^T+g\,g^T. \qquad (9)$$

In (9), k_n stands for the filter gain and P_o is given by (5).

For the normalized variance process, $\tilde{v}_n=v_n(d_n)^{-1/2}$, equation (7) may be writen, for $n\geq 0$ in a matrix form as

$$[\tilde{v}_o\ \tilde{v}_1 \dots \tilde{v}_N]^T=\tilde{W}_N^{-1}\,[y_o\ y_1 \dots y_N]^T\ , \qquad (10)$$

where $\tilde{W}_N^{-1}$ is a lower triangular matrix of order N+1. The elements on line i of $\tilde{W}_N^{-1}$ (lines and columns are numbered starting from zero) are the coefficients of the one-step ahead linear prediction, i-th order, error filter.

Pre-multiplying both sides of (10) by $\tilde{W}_N$, yields the process normalized innovation representation

$$[y_o\ y_1 \dots y_N]^T=\tilde{W}_N\,[\tilde{v}_o\ \tilde{v}_1 \dots \tilde{v}_N]^T. \qquad (11)$$

The duality between (10) and (11) is clear, the coefficients of the normalized innovation filter of order i, being the entries on line i of the lower triangular matrix $\tilde{W}_N$. A closed expression of the elements of $\tilde{W}_N$ and $\tilde{W}_N^{-1}$ is provided in Ref. 7.

3. ARMA ESTIMATION ALGORITHM

The dual, decoupled, AR and MA estimation algorithms presented in this section are based on the linear dependence of the coefficients of successive order normalized innovation and linear predictor filters, presented at the same column of the matrices $\tilde{W}_N$ and $\tilde{W}_N^{-1}$.

Pre-multiply the normalized innovation representation of the process (11) by the lower triangular, band diagonal matrix A_N with entries

$$(A_N)_{ij\ \ i,j=0,1,\dots,N}=\begin{cases} a_{i-j} & 0\leq i-j\leq p \\ 0 & \text{other ise} \end{cases} \qquad (12)$$

and define the matrix

$$\beta_N=A_N\,\tilde{W}_N\,, \qquad (13)$$

which has the following structure (Ref. 5)

$$\beta_N=\begin{bmatrix} \alpha_0(0) & & & & & & & \\ \alpha_1(1) & \alpha_0(1) & & & & 0 & & \\ \vdots & \vdots & \ddots & & & & & \\ \alpha_q(q) & \alpha_{q-1}(q) & \cdots & \alpha_0(q) & & & & \\ \vdots & & \ddots & & \ddots & & & \\ \alpha_{p-1}(p-1) & \cdots & \alpha_q(p-1) & \alpha_{q-1}(p-1) & \cdots & \alpha_0(p-1) & & \\ 0 & 0 & \cdots & 0 & \alpha_q(p) & \cdots & \alpha_1(p) & \alpha_0(p) \\ & & & & & \ddots & & \ddots \\ & & 0 & & & & & \\ & & & & \alpha_q(N) & \cdots & \alpha_1(N) & \alpha_0(N) \end{bmatrix} \qquad (14)$$

the elements $\alpha_m(i)$ being expressed as

$$\alpha_m(i)=\{b_m b_o-\sum_{r=m+(p-q)+1}^{p} a_r hF^{m-r}P_{i-m}h^T\}d_{i-m}^{-1/2}. \qquad (15)$$

$$0\leq m\leq q,\ i\geq p$$

From (13) and the structure exhibited by (14), the elements on the first i-q columns of the matrix $\tilde{W}_i$ $(p\leq i\leq N)$ satisfy the linear recursion (Ref. 5)

$$\tilde{W}_j^i+a_1\tilde{W}_j^{i-1}+\ \dots\ +a_p\tilde{W}_j^{i-p}=0 \qquad 0\leq j\leq i-q-1 \qquad p\leq i\leq N \qquad (16)$$

defined by the AR component of the process and where $\tilde{W}_j^i=(\tilde{W}_N)_{ij}$, for $0\leq i,j\leq N$. Consequently, for each value of i, the column vector

$$\underline{a}=[a_1\ a_2\ \dots\ a_p]^T \qquad (17)$$

satisfy a system of linear equations, written in matrix format as

$$C_i\ \underline{a}=f_i \quad , \quad p\leq i\leq N \qquad (18)$$

where

$$C_i = \begin{bmatrix} \tilde{W}_0^{i-1} & \cdots & \tilde{W}_0^{i-p} \\ \vdots & & \vdots \\ \tilde{W}_{i-p}^{i-1} & \cdots & \tilde{W}_{i-p}^{i-p} \\ \vdots & \cdot^{\cdot^{\cdot}} & \\ \tilde{W}_{i-q-1}^{i-1} & \cdots \tilde{W}_{i-q-1}^{i-q-1} & \underline{0} \end{bmatrix} \tag{19}$$

$$f_i = -[\tilde{W}_0^i \; \tilde{W}_1^i \; \cdots \; \tilde{W}_{i-q-1}^i]^T. \tag{20}$$

Using again (13), the dual result of (16), relates the elements on the first i-p columns of the matrix $\tilde{W}_i^{-1}$ and the $\alpha_m(i)$ coefficients,

$$\alpha_0(i)\tilde{a}_j^i + \alpha_1(i)\tilde{a}_j^{i-1} + \ldots + \alpha_q(i)\tilde{a}_j^{i-q} = 0, \quad 0 \leq j \leq i-p-1 \tag{21}$$

where $\tilde{a}_j^i$ stands for the element $(\tilde{W}_N^{-1})_{ij}$.

The equation (21) is the asymptotic MA counterpart of the AR-defined linear relation (16). In fact,

$$\lim \alpha_m(i) = b_m, \quad 0 \leq m \leq q \tag{22}$$

the rate of this convergence being governed by the second power of the zeros of the observation process (Ref. 5).

In Ref. 7, the equation (21) was presented for j=0, i.e. for the first column of the matrix $\tilde{W}_i^{-1}$; this result is generalized here for the first i-p columns of that matrix.

Let

$$\underline{\alpha}(i) = [\alpha_1(i) \; \alpha_2(i) \ldots \alpha_q(i)]^T \tag{23}$$

which together with $\alpha_0(i)$ converges to the MA component of the process. For each value of i, (21) may be written in matrix format as

$$\tilde{C}_i \; \underline{\alpha}(i) = \tilde{f}_i \qquad p \leq i \leq N \tag{24}$$

where

$$\tilde{C}_i = \begin{bmatrix} \tilde{a}_0^{i-1} & \tilde{a}_0^{i-2} & \cdots & \tilde{a}_0^{i-q} \\ \vdots & \vdots & & \vdots \\ \tilde{a}_{i-p-1}^{i-1} & \tilde{a}_{i-p-1}^{i-2} & \cdots & \tilde{a}_{i-p-1}^{i-q} \end{bmatrix} \tag{25}$$

$$\tilde{f}_i = -d_i^{-1/2} \; [\tilde{a}_0^i \; \tilde{a}_1^i \; \cdots \; \tilde{a}_{i-p-1}^i]^T. \tag{26}$$

The systems of linear equations (18) and (24) exhibit the duality behavior of the AR and MA components of the process in their relation with the innovation and prediction error filter coefficients. We note that, for i>p+q, both (18) and (24) represent an oversized system of equations.

The above analysis is based on the exact knowledge of both the normalized innovation and prediction error filter coefficients. In that case, the vector $\underline{a}$ could be determined by solving jointly any p of the preceeding equations (16) while the vector $\underline{\alpha}(i)$, which converges to the MA component, is the solution of any q of the simultaneous relations (21), defined for that value of i. However, in the presence of a finite sample of the observation process, the $\tilde{W}_j^i$ and $\tilde{a}_j^i$ coefficients are replaced by suitable estimates, the latter provided by the Burg technique. The use of an oversized system of equations, has then statistical relevance compensating the errors on the $\tilde{W}_j^i$ and $\tilde{a}_j^i$ estimation. For the AR component, this is similar to (Ref. 1).

We will present a scheme that obtains the estimate of the AR component using all the linear relations (16) and that recursively updates the estimation when the coefficient estimation of an higher order innovation filter is available. Let $\hat{\underline{a}}(k)$ be the least-squares solution of the system of all the linear equations (16), established for $p \leq i \leq k$. A recursive estimation of the AR component is given by

$$\hat{\underline{a}}(k) = \hat{\underline{a}}(k-1) + M(k)^{-1}C_k^T[f_k - C_k \, \hat{\underline{a}}(k-1)] \tag{27}$$

where $\hat{\underline{a}}(k)$ and $\hat{\underline{a}}(k-1)$ were defined above, C_k and f_k are as in (19) and (20) and M(k) is defined by

$$M(k) = \mathcal{C}^T(k)\,\mathcal{C}(k) \tag{28}$$

$$\mathcal{C}(k) = [C_p^T \mid C_{p+1}^T \mid \cdots \mid C_k^T]^T. \tag{29}$$

For i>p+q, the least-squares solution of the oversized system of equations (24), has the above mentioned statistical relevance in the presence of a finite sample of the observation process. The solution is time-varying with i, preventing the simultaneous use of the linear relations (21) for different values of i. However, if the value of i is high enough so that the convergence of $\alpha_m(i)$ to the b_m coefficients has been attained, a recursive scheme on the MA estimation may be implemented as

$$\hat{\underline{b}}(k) = \hat{\underline{b}}(k-1) + \tilde{M}(k)^{-1}\tilde{C}_k^T[\tilde{f}_k - \tilde{C}_k \, \hat{\underline{b}}(k-1)], \tag{30}$$

where $\hat{\underline{b}}(k)$ is the least-squares solution of the system of all the linear equations (21) for $N^* \leq i \leq k$, $\tilde{C}_k$ and $\tilde{f}_k$ are as in (25) and (26) and

$$\tilde{M}(k) = \tilde{\mathcal{C}}^T(k)\,\tilde{\mathcal{C}}(k)$$

$$\tilde{\mathcal{C}}(k) = [\tilde{C}_N^T \mid \tilde{C}_{N+1}^T \mid \cdots \mid \tilde{C}_k^T]^T.$$

The recursion (30) is started with $\hat{\underline{b}}(N^*) = \underline{\alpha}(N^*)$ being assumed that the transient associated with the $\alpha_m(i)$ coefficients has died out for $i \geq N^*$.

4. SIMULATION RESULTS

In this section we present simulated examples that illustrate the estimation algorithm performance under several ARMA(p,q) processes. The figures 1 and 2 refer to two ARMA(4,4) processes, represented by (1) with $b_0=1$, and pole-zero pattern displayed in table 1.

Case 1		Case 2	
Poles	Zeros	Poles	Zeros
$0.97e^{\pm j45^\circ}$	$0.97e^{\pm j60^\circ}$	$0.97e^{\pm j45^\circ}$	$0.87e^{\pm j60^\circ}$
$0.9e^{\pm j90^\circ}$	$0.9e^{\pm j110^\circ}$	$0.9e^{\pm j90^\circ}$	$0.81e^{\pm j110^\circ}$

Table 1

In figure 1 we represent both the real and the mean estimated spectrum obtained with 100 Monte-Carlo runs. For T=5000 data points (Fig.1b), one can see that as N, the highest order filter considered in the estimation, increases, the zeros are better estimated with evident consequences on the spectral valeys estimation. However, for small values of T, there exists a tradeoff between the estimation errors in the prediction and innovation filter coefficients as N increases and the convergence of the $\alpha_m(N)$ coefficients to the corresponding b_m. This is shown in fig.1a) obtained with T=500 data points.

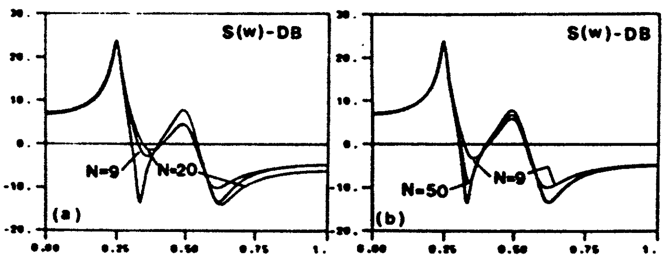

Fig.1 - Real and mean estimated spectrum, obtained with 100 Monte-Carlo runs, for Case 1.

The convergence rate associated with the MA estimation is shown in fig.2, where the real and estimated pole-zero pattern are compared for the two processes referred in table 1. The comparison is done for small values of N and T=5000 to prevent the effect of the errors on the filter coefficients. The following notation is used: x - real pole, o - real zero, ◆ - estimated pole and ◇ - estimated zero.

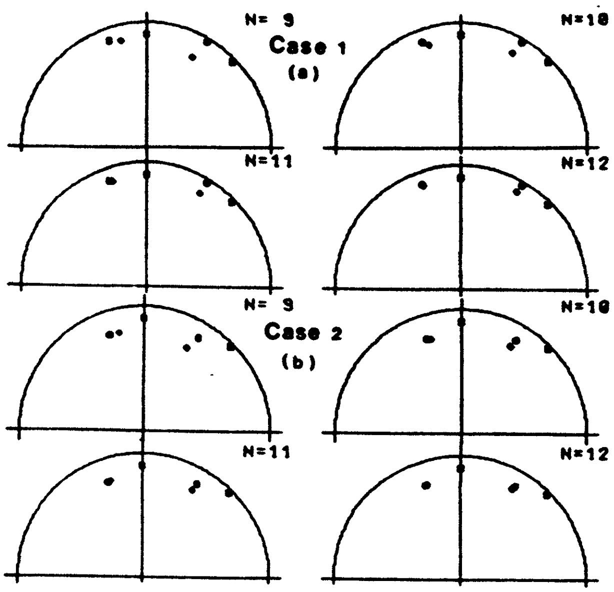

Fig.2-Real and estimated pole-zero pattern for T=5000 data points.

The zeros of the process referred to as case 2 in table 1 have smaller magnitude than those of case 1, the convergence of the estimated zeros to their real location being faster in fig.2-b) (case 2) than in fig.2-a) (case 1).

We present some preliminar results on order estimation, where the dual roles of both the AR and MA compoents is also evident. This is a joint work with R.S.Bucy, observed when the simulations for Ref. 6 were carried out. It is based on the migration pattern of the pole and zero configurations as higher order (purely) MA processes and higher order (purely) AR processes are fitted to the ARMA process.

The root pattern for the predictor and innovation filters of high order N is displayed in fig.3 for an ARMA(2,1) process with poles=.3; -.6 and zero=-.8. In fig.3-a) all but one the predictor roots lie in a Butterworth configuration with magnitude determined by the zero. The other root coincides with the pole=0.3. In a dual way, all but one innovation filter roots are displayed in a Butterworth pattern, defined by the highest pole, the outsider root estimating the zero of the process.

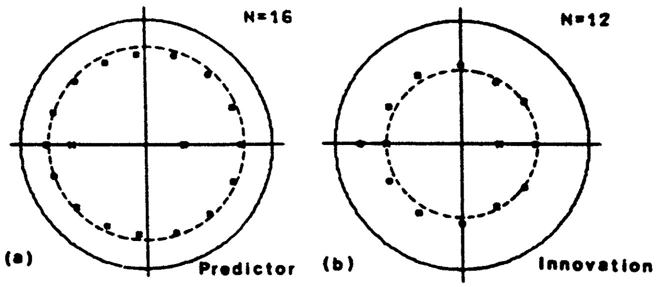

Fig.3 - Predictor and innovation roots pattern

The ideas herein presented on order estimation require further study. A more sophisticated approach will identify the order and pole/zero pattern iteratively, first estimating the outer most poles or zeros, filtering them out and progressing inwards.

REFERENCES

1. CADZOW, J. A., 'Spectral Estimation: An Overdetermined Rational Model Equation Approach', IEEE Proceedings, Vol.70, No.9, Sept. 1982.
2. KAY, S. M., MARPLE, S. L., 'Spectrum Analysis: A Modern Perspective', IEEE Proceedings, Vol.69, No.11, Nov. 1981.
3. FRIEDLANDER,B., 'Lattice Methods for Spectral Estimation', IEEE Proceedings, Vol.70, No.9, Sept. 1982.
4. BURG,J.P., 'Maximum Entropy Spectral Analysis', Ph.D. Thesis, Stanford University, 1975.
5. RIBEIRO,M.I., MOURA,J.M.F., 'Dual Estimation of the Poles and Zeros of an ARMA (p,q) process', CAPS, Technical Report 02/84, November 1984; revised as LIDS Technical Report No. LIDS-P-1521, MIT, September 1985, submitted for publication.
6. ALENGRIN,G., BUCY,R., MOURA,J., PAGES,J., RIBEIRO,I., 'Arma Identification', January 1985, revised as LIDS Technical Report No. LIDS-P-1588, MIT, August 1986.
7. BUCY,R.S., 'Identification and Filtering', Mathematical Systems Theory, Vol. 16, December 1983.