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 problems. 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 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 \( W_N \), which is the square root inverse of the Toeplitz covariance matrix

\[ R_N = W_N^{-1} \]

A recursive inversion of \( W_N \) in 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

\[ y_n = \sum_{i=1}^{p} a_i y_{n-i} + \sum_{i=1}^{q} b_i e_{n-i} \] (1)

where it is assumed that i) \( \{e_n\} \) is a white, Gaussian, zero mean noise sequence with unit variance, ii) \( a_p, a_0, b_q, b_0 \), iii) the polynomial matrices

TH0167-7/86/0000-0037 $1.00 © 1986 IEEE
\[ A(z) = 1 + \sum_{i=1}^{p} a_i z^{-i}, \quad B(z) = \sum_{i=0}^{q} b_i z^{-i} \] (2)

are asymptotically stable and have no common roots.

Define the process autocovariance at lag \( k \), as
\[ r(k) = \mathbb{E}[y_n + ky_{n+k}] \]. 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

\[ y_{n+1} = F x_n + g e_n \] (3)
\[ y_n = h x_n + b e_n \] (4)

with minimal dimension, i.e. \( X \in \mathbb{R}^p \), completely controllable (c.c.) and completely observable (c.o.), and where \( F \) is an asymptotically stable, nonsingular matrix. The initial state \( x_0 \) is a Gaussian random vector, independent of \( \{e_n\} \) with zero mean and covariance matrix \( P_0 \), satisfying

\[ P_0 = F P_0 F^T + g g^T. \] (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)

\[ \hat{x}_{n+1} = F \hat{x}_n + k_n v_n, \quad \hat{x}_0 = 0, \] (6)
\[ \hat{y}_n = y_n - h \hat{x}_n \] (7)

where the variance of the innovation process \( v_n \) is given by

\[ d_n = h P_n h^T + b b^T, \] (8)

being related to the discrete Riccati equation

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

In (9), \( k_n \) stands for the filter gain and \( P_0 \) is given by (5).

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

\[ \begin{bmatrix} \tilde{v}_0 & \tilde{v}_1 & \ldots & \tilde{v}_n \end{bmatrix} = \tilde{W}_n \begin{bmatrix} y_0 & y_1 & \ldots & y_n \end{bmatrix}^T, \] (10)

where \( \tilde{W}_n \) is a lower triangular matrix of order \( N+1 \). The elements on line \( i \) of \( \tilde{W}^{-1}_n \) (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

\[ \begin{bmatrix} y_0 & y_1 & \ldots & y_n \end{bmatrix}^T = \tilde{W}_n \begin{bmatrix} \tilde{v}_0 & \tilde{v}_1 & \ldots & \tilde{v}_n \end{bmatrix}^T. \] (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

\[ (A_n)_{ij} = \begin{cases} b_{i-j} & \text{if } 0 \leq j < i \leq N \setminus 0 \text{ otherwise} \end{cases} \] (12)

and define the matrix

\[ B_n = A_n \tilde{W}_n, \] (13)

which has the following structure (Ref. 5)

\[ A_n = \begin{bmatrix} a_1 & \ldots & a_{N-1} & 0 \end{bmatrix}, \] (14)

the elements \( a_n(i) \) being expressed as

\[ a_n(i) = \begin{bmatrix} b_0 & b_1 & \ldots & b_{N-1} & 0 \end{bmatrix}. \] (15)

For instance, the elements on the first \( i-q \) columns of the matrix \( \tilde{W}_1 \) \((s < i < N)\) satisfy the linear recursion (Ref. 5)

\[ \tilde{W}_1^i + a_{i+1} \tilde{W}_1^{i+1} + \ldots + a_{i-s} \tilde{W}_1^{i-s} = 0, \quad 0 < s < i < N \] (16)

defined by the AR component of the process and where \( \tilde{W}_{1} = (\tilde{W}_1)_{ij} \) for \( 0 < i, j < N \). Consequently, for each \( i \), the column vector

\[ \mathbf{a} = [a_1, a_2, \ldots, a_p]^T \] (17)

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

\[ C_i \mathbf{a} = f_i, \quad p < i < N \] (18)

where
The equation (21) is the asymptotic MA counterpart of the AR-defined linear relation (16). In fact, 
\[ \lim a(i) = b_m^T, \quad 0 \leq i \leq q \] (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 \( \tilde{W}_1 \); this result is generalized here for the first \( i-p \) columns of that matrix. Let
\[ a(i) = [a_0(i) \ a_1(i) \ ... \ a_{q-1}(i)]^T \] (23)
which together with \( a_0(i) \) converges to the MA component of the process. For each value of \( i \), (21) may be written in matrix format as
\[ \tilde{C}_i \quad a(i) = \tilde{f}_i \quad p \leq i \leq N \] (24)
where
\[ \tilde{C}_i = \begin{bmatrix} a_{i-1} & a_{i-2} & ... & a_{i-q} \\ a_0 & a_{i-1} & ... & a_{i-q} \\ ... & ... & ... & ... \\ a_{i-1} & a_{i-2} & ... & a_{i-q} \\ a_{i-p-1} & a_{i-p-2} & ... & a_{i-p-1} \end{bmatrix} \] (25)
\[ \tilde{f}_i = -\frac{1}{2} \begin{bmatrix} a_{i-1} & a_{i-2} & ... & a_{i-p-1} \end{bmatrix}^T \] (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 \( i \geq p+q \), 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 \( \tilde{a}^i \) could be determined by solving jointly any \( p \) of the preceding equations (16) while the vector \( \tilde{a}^i \), 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 \( \tilde{a}^i \) and \( \tilde{a}^j \) 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 \( \tilde{W}_1 \) and \( \tilde{a}^i \) 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 \( \tilde{a}(k) \) be the least-squares solution of the system of all the linear equations (16), established for \( p \leq k \leq i \). A recursive estimation of the AR component is given by
\[ \tilde{a}(k) = \tilde{a}(k-1) + H_k \tilde{f}_k - C_k \tilde{z}(k-1) \] (27)
where \( \tilde{a}(k) \) and \( \tilde{a}(k-1) \) were defined above, \( C_k \) and \( f_k \) are as in (19) and (20) and \( H(k) \) is defined by
\[ H(k) = C_k^T(k) C(k) \] (28)
\[ C_k = [C_p \ T \ C_{p+1} \ T \ ... \ C_N \ T]^T \] (29)

For \( i \geq p+q \), 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 \( a(i) \) to the \( b_m \) coefficients has been attained, a recursive scheme on the MA estimation may be implemented as
\[ \hat{a}(k) = \hat{a}(k-1) + H(k) \tilde{f}_k - C_k \hat{z}(k-1) \] (30)
where \( \hat{a}(k) \) is the least-squares solution of the system of all the linear equations (21) for \( N \leq k \), \( C_k \) and \( \tilde{f}_k \) are as in (25) and (26) and
\[ H(k) = C_k^T(k) C_k \] (31)
\[ \hat{a}(k) = [C_N \ T \ C_{N+1} \ T \ ... \ C_N \ T]^T \] (32)

The recursion (30) is started with \( \hat{a}(N) = \tilde{a}(N) \) being assumed that the transient associated with the \( a_m(i) \) 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 \( b = 1 \) and pole-zero pattern displayed in table 1.
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 preliminary results on order estimation, where the dual roles of both the AR and MA components 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 higher order $N$ is displayed in Fig.3 for an ARMA(2,1) process with poles $-0.3; -0.6$ and zero $-0.8$. 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.

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**


