LD^2–ARMA Identification Algorithm

M. Isabel Ribeiro, *Member, IEEE*, and José M. F. Moura, *Senior Member, IEEE*

***Abstract*—This paper presents the LD^2-ARMA identifier, a novel algorithm that solves the essentially nonlinear autoregressive moving-average identification problem with a linear procedure, in two steps: an order selection algorithm followed by an ARMA parameter estimator. The determination of the AR and MA coefficients involves the solution of two dual systems of linear equations. These systems decouple the estimation of the autoregressive (AR) component from the estimation of the moving average (MA) component. The selection of the number of poles p_0 and of the number of zeros q_0 is accomplished by a scheme that minimizes the mismatch of the data to each proposed ARMA (p, q) model. Simulated experiments on the proposed order selection procedure are presented. The statistical analysis and extensive simulation results are discussed in a companion paper.**

I. Introduction

A. ARMA Estimation

An important question in many problems is that of fitting models to a series of measurement points. The available *a priori* information and the ultimate purpose of the model may change with the specifics of each application, e.g., econometric time series analysis, speech processing, bioengineering. Accordingly, there are alternative standard classes of models to fit the time series. Among them, the autoregressive (AR) and the autoregressive moving-average (ARMA) models play a relevant role.

In the absence of the moving-average (MA) part, the estimation of the process parameters has relatively well-established methods [16], [19]. These explore the linear relations between the autocorrelation function and the AR coefficients, by solving the Yule–Walker equations. In contrast with the linearity displayed by the estimation of AR processes, ARMA identification is a nonlinear problem. Optimization techniques based on the maximum-likelihood (ML) estimator [2], [15], and on nonlinear least squares techniques [14], have been applied to this problem. A common approach is a three step sequential procedure based on the modified Yule–Walker (MYW) equations (see e.g., [10], [11], [14], [31]). First, a sample autocovariance function is estimated from the data. Then, the AR component is obtained, and finally, the MA coefficients follow as a function of the previously computed AR coefficients. Similar sequential estimation procedures but based on the reflection coefficient sequence are presented in [5], [6], [18], [21].

A trend in spectral estimation studies, e.g., [30], is to solve a basically nonlinear problem by a linear procedure. In this paper we take the same point of view and solve the nonlinear problem of ARMA identification by a *linear, dual, decoupled* procedure—the LD^2-ARMA algorithm, which estimates independently the AR and the MA coefficients. Characteristics of our proposed technique include the following:

- The *linearity* of the algorithm. The determination of the AR and of the MA coefficients, as well as the order selection, involves only the solution of systems of linear equations.
- The *duality* of the procedure. The AR and the MA estimation are obtained by the same type of operations—the solution of a system of linear equations.
- The *decoupling* of the AR and of the MA component procedures, i.e., neither of them depends on previously estimated values of the other, thus differing from reported techniques.
- The use of a sample *reflection coefficient (RC) sequence* instead of a sample covariance sequence like MYW-based ARMA algorithms. The present ARMA estimator may be thought of as a square-root MYW-type algorithm exhibiting the improved stability properties known to square root algorithms.

B. Order Selection

In any identification sheme, the structure (e.g., the number of poles p_0 and of zeros q_0), if unknown, has to be determined prior to or within the estimation procedure. Two standard techniques for model order selection, the AIC [1] and the MDL [29], obtain the number of poles and zeros by minimizing a functional that accounts for the residual power and for the overparameterization.

The ld^2-ARMA identification algorithm detailed in this paper accomplishes the order selection by a scheme that minimizes the mismatch of the data to each proposed ARMA(p, q) model, no residual power minimization being performed. Thus, unlike AIC and MDL procedures, the overall identification algorithm is based on a single pass over the data, since there is no need to estimate the cross-correlation sequence between the data and the error process. Once the number of poles and zeros is selected, the algorithm simultaneously provides the estimated values of the parameters of the AR and of the MA components of the corresponding model.

Manuscript received August 10, 1989; revised August 4, 1990. This work was supported in part by NATO Research Grant 585/83.

M. I. Ribeiro is with CAPS/INIC, Instituto Superior Técnico, P-1096 Lisboa Codex, Portugal.

J. M. F. Moura is with LASIP, Department of Electrical and Computer Engineering, Carnegie-Mellon University, Pittsburgh, PA 15213.

IEEE Log Number 9100761.

In Section II, we formulate the identification problem for ARMA processes, state the hypotheses underlying the paper, and discuss a particular solution structure that solves it. Useful notation is introduced in Section II. In Section IV, the required theoretical relations are described. The order estimation algorithm is discussed in Section V, stressing the *linearity*, the *duality*, and the *decoupling* of the LD²-ARMA procedures. We introduce the functional to be minimized, present required properties, and propose the selection scheme for the number of poles and zeros. In Section VI, the LD²-ARMA implementation is presented. Some simulated examples compare the described order selection algorithm with the AIC criterion introduced by Akaike [1], and with the MDL suggested by Akaike [4], and Rissanen [29]. A companion paper [28], carries out the statistical analysis of our algorithm and illustrates its behavior with a set of extensive simulation results.

II. Problem Formulation

Let $\{y_0, y_1, \cdots, y_{L-1}\}$ be a finite sample of length L drawn from a stationary scalar process $\{y_n\}$ satisfying the recursion

$$y_n + \sum_{i=1}^{p_0} a_i y_{n-i} = \sigma\left[e_n + \sum_{i=1}^{q_0} b_i e_{n-i}\right] \qquad (1)$$

where $\{e_n\}$ is white noise, with zero mean and unit variance. The process $\{y_n\}$ is an ARMA(p_0, q_0) process. Model (1) is assumed to be stable and minimum phase, the numerator and denominator polynomials of its transfer function

$$B(z) = \sum_{i=0}^{q_0} b_i z^{-i}, \qquad b_0 = 1,$$
$$A(z) = \sum_{i=0}^{p_0} a_i z^{-i}, \qquad a_0 = 1 \qquad (2)$$

having no common roots. The vectors

$$\boldsymbol{b} = [b_1 \;\; b_2 \;\cdots\; b_{q_0}]^T, \qquad \boldsymbol{a} = [a_1 \;\; a_2 \;\cdots\; a_{p_0}]^T \qquad (3)$$

collect the MA coefficients $\{b_i\}$ and the AR coefficients $\{a_i\}$ of the process. No assumption is made on the relation between p_0 and q_0. In particular, in AR studies, $q_0 = 0$, whereas $p_0 = 0$ leads to the identification of MA processes.

The ARMA identification problem is now stated. Given a set of observations, $\{y_0, y_1, \cdots, y_{L-1}\}$ from an ARMA(p_0, q_0) process, find estimates of the number of poles $\hat{p}_0$, of the number of zeros $\hat{q}_0$, of the AR and of the MA components, $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$, and of $\widehat{\sigma^2}$, i.e., determine the unknown parameter vector

$$\theta = [p_0, q_0, \boldsymbol{a}^T, \boldsymbol{b}^T, \sigma^2]^T. \qquad (4)$$

The solution adopted in this paper for the ARMA identification problem is represented in Fig. 1. The parallelism in this block diagram emphasizes the duality and the decoupling of the AR and of the MA estimation procedures.

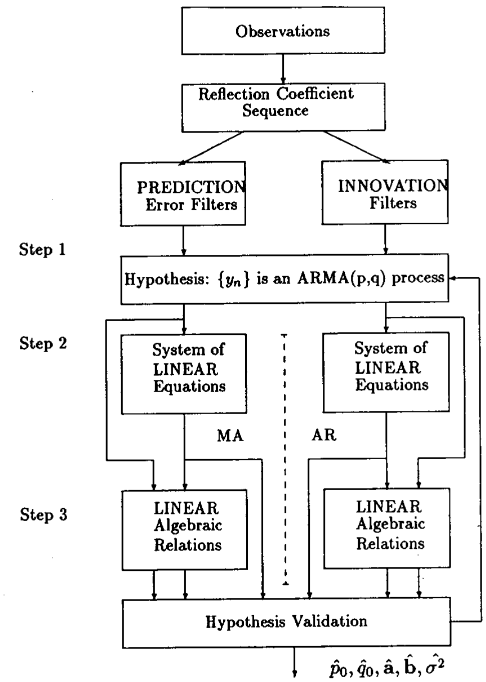


Fig. 1. Solution structure.

Starting from the set of observations of the ARMA(p_0, q_0) process, the identification algorithm obtains a second-order characterization in terms of a sample reflection coefficient (RC) sequence. From this, the algorithm computes the associated prediction and innovation filters' coefficients. The relationships between these filters, the RC sequence, and the ARMA(p_0, q_0) process will be detailed later in this section. Then, the algorithm iterates among three steps:

Step 1: Hypothesizes that the data corresponds to an ARMA(p, q) model.

Step 2: Obtains the estimated values of the AR and of the MA components of the corresponding model. These are the solution of two decoupled systems of linear equations.

Step 3: Tests the hypothesis.

The stopping condition of the algorithm is determined by the hypothesis validation (step 3). When the hypothesis is accepted, the corresponding pair (p, q) defines the estimated model structure, its AR and MA coefficients being given by step 2. If the hypothesis is rejected, the identification scheme iterates on (p, q).

For the sake of clarity, we describe briefly the kind of operations performed in steps 2 and 3. The two systems of linear equations considered in step 2 are constructed from the coefficients of increasing order prediction and innovation filters. The number of unknowns and the number of equations of each system is consistent with the pair (p, q) considered in step 1. The decoupling of both systems is clear from Fig. 1. The duality of the two estimation procedures will be proved in Section IV. The hypothesis validation (step 3) is carried out by minimizing a functional d. For each pair (p, q), d is evaluated by using the current estimates of the AR and of the MA coefficients and two sets of linear algebraic relations between these.

In the remaining of this section, we present the prediction and innovation filters associated with the ARMA process and its second-order characterization given by the reflection coefficient sequence.

A. Prediction/Innovation Filters

Let $\{v_n\}$

$$v_n = y_n - E[y_n | y_0, y_1, \cdots, y_{n-1}] \tag{5}$$

be the innovation sequence associated with the ARMA process (1) and represented by

$$A_n(z) = \sum_{i=0}^{n} a_i^n z^{-i}, \qquad a_0^n = 1 \tag{6}$$

the corresponding prediction error filter of order n, i.e.,

$$v_n = \sum_{i=0}^{n} a_i^n y_{n-i}, \qquad a_0^n = 1. \tag{7}$$

Notice that a_n^n, the coefficient of order n of the nth order filter, is the reflection coefficient of order n of the ARMA process. We will denote it alternatively by $c_n = a_n^n$. From the definition of the innovation sequence, it follows that [8]

$$v_n = y_n - E[y_n | v_0, v_1, \cdots, v_{n-1}] \tag{8}$$

leading to

$$y_n = \sum_{i=0}^{n} W_i^n v_{n-i}, \qquad W_0^n = 1 \tag{9}$$

where the set $\{W_i^n, 0 \le i \le n\}$ collects the coefficients of the innovation filter of order n, represented by

$$B_n(z) = \sum_{i=0}^{n} W_i^n z^{-i}. \tag{10}$$

As n goes to ∞, $A_n(z)$ and $B_n(z)$ represent long AR(n) and long MA(n) models which are equivalent to the stable, minimum-phase original ARMA model. This corresponds to the Wold and the Kolmogorov decompositions, respectively.

In [8], the coefficients of the prediction error filters a_i^n and of the innovation filters, W_i^n, are expressed as a function of the discrete Kalman–Bucy filter dynamics for a model with noise free observations. The modifications of these for the total correlated noise model of (A.5), (A.6) is carried out in [24]. See also [12]. In the next section we will use these coefficients to define two lower triangular matrices that play an important role on the estimation algorithm.

III. Notation

In this section we present useful notation. For the pairs (p, q) and (p_0, q_0), define

$$\bar{p}_0 = \max\{p_0, q_0\} \qquad \bar{p} = \max\{p, q\}.$$

In the sequel, the subindex N stands for the highest order prediction and innovation filter used by the estimation procedure. Unless otherwise stated, the lines and columns of the matrices are numbered starting from zero. Let $\mathbf{\Omega}(k)$ be the vector of order q_0

$$\mathbf{\Omega}(k) = [\Omega_1(k) \;\; \Omega_2(k) \;\; \cdots \;\; \Omega_{q_0}(k)]^T, \qquad \bar{p}_0 \le k \le N \tag{11}$$

with elements, $\Omega_i(k)$, specified later and let $\boldsymbol{J}_n$ be the permutation matrix of order n, i.e.,

$$\boldsymbol{J}_n = \begin{bmatrix} & & 1 \\ \mathbf{0} & \cdot^{\cdot^{\cdot}} & \\ & 1 & \mathbf{0} \\ 1 & & \end{bmatrix}, \qquad \forall n \in N. \tag{12}$$

Throughout the paper, we make the convention

$$a_i^n = W_i^n = 0, \qquad \text{for } i > n, \qquad \text{or } i < 0. \tag{13}$$

Define the lower triangular, unit diagonal matrices of order $N + 1$ as

$$\boldsymbol{W}_N^{-1} = \begin{bmatrix} 1 & & & & \\ a_1^1 & 1 & & \mathbf{0} & \\ a_2^2 & a_1^2 & 1 & & \\ & \cdots & & \ddots & \\ a_N^N & a_{N-1}^N & \cdots & a_1^N & 1 \end{bmatrix}$$

$$\boldsymbol{W}_N = \begin{bmatrix} 1 & & & & \\ W_1^1 & 1 & & \mathbf{0} & \\ W_2^2 & W_1^2 & 1 & & \\ & \cdots & & \ddots & \\ W_N^N & W_{N-1}^N & \cdots & W_1^N & 1 \end{bmatrix}. \tag{14}$$

These matrices play a key role on the dual ARMA estimation procedure. The following commments explain their meaning and clarify some options taken on the estimation algorithm.

Comment 1: The lines of matrices $\boldsymbol{W}_N^{-1}$ and $\boldsymbol{W}_N$ collect the coefficients of the successively increasing order prediction and innovation filters up to order N. This may be represented as

$$
W_N^{-1}\begin{bmatrix} 1 \\ z^{-1} \\ z^{-2} \\ \vdots \\ z^{-N} \end{bmatrix} = \begin{bmatrix} 1 \\ z^{-1}A_1(z^{-1}) \\ z^{-2}A_2(z^{-1}) \\ \vdots \\ z^{-N}A_N(z^{-1}) \end{bmatrix},
$$

$$
W_N\begin{bmatrix} 1 \\ z^{-1} \\ z^{-2} \\ \vdots \\ z^{-N} \end{bmatrix} = \begin{bmatrix} 1 \\ z^{-1}B_1(z^{-1}) \\ z^{-2}B_2(z^{-1}) \\ \vdots \\ z^{-N}B_N(z^{-1}) \end{bmatrix} \tag{15}
$$

with $A_n(z)$ and $B_n(z)$ defined in (6) and (10). An interpretation of the columns of W_N in terms of a time-varying impulse response is given in [13], [15].

Comment 2: The set of polynomials $\{A_1(z), A_2(z), \cdots, A_N(z)\}$ correspond to increasing orders AR models that can be approximately considered equivalent to the ARMA(p_0, q_0) process. Similar comments hold for $\{B_1(z), B_2(z), \cdots, B_N(z)\}$ and a set of increasing order MA models.

Comment 3: The reflection coefficient sequence $\{a_0^0, a_1^1, a_2^2, \cdots, a_N^N\}$ associated with the ARMA(p_0, q_0) is the first column of W_N^{-1}.

Comment 4: The matrix W_N is the Cholesky factor of the Toeplitz autocovariance matrix of the process R_N

$$
R_N = W_N D_N W_N^T \tag{16}
$$

where $[R_N]_{ij} = r(|i-j|)$, and $r(k) = E[y(n+k)\,y(n)]$. The diagonal matrix D_N in (16) is

$$
D_N = \text{diag}\,(d_0, d_1, \cdots, d_N) \tag{17}
$$

where d_n collects the power of the innovation process. In a similar way, W_N^{-1} is a Cholesky factor of R_N^{-1}.

Comment 5: In spite of the relationships between W_N and W_N^{-1} with the autocovariance matrix R_N, the identification algorithm of this paper does not start form an estimated value of R_N, no Cholesky factorization being performed. Instead, it estimates the prediction and the innovation filters' coefficients directly from the data.

Define A_N as a band diagonal, lower triangular, Toeplitz matrix of order $N+1$, constructed with the AR coefficients of the ARMA(p_0, q_0) process

$$
A_N = \begin{bmatrix}
1 & & & & & & \\
a_1 & 1 & & & & \mathbf{0} & \\
a_2 & a_1 & 1 & & & & \\
\vdots & & \ddots & \ddots & & & \\
a_{p0} & \cdots & \cdots & a_1 & 1 & & \\
 & a_{p0} & \cdots & \cdots & a_1 & 1 & \\
\mathbf{0} & & \ddots & & & \ddots & \ddots \\
 & & & a_{p0} & \cdots & \cdots & a_1 \;\; 1
\end{bmatrix}. \tag{18}
$$

Let β_N be a lower triangular, band diagonal matrix given by

$$
\beta_N = \begin{bmatrix}
\text{(shaded area)} & & & & \mathbf{0} \\
\mathbf{0} & \Omega_{q0}(\bar{p}_0) & \cdots & \Omega_1(\bar{p}_0) & 1 \\
\ddots & & \ddots & & \ddots \\
\mathbf{0} & \Omega_{q0}(k) & \cdots & \Omega_1(k) & 1 \\
\ddots & & \ddots & & \ddots \\
 & \Omega_{q0}(N) & \cdots & \Omega_1(N) & 1
\end{bmatrix} \quad \bar{p}_0 \le k \le N \tag{19}
$$

where the entries on the shaded area are not used in the ARMA estimation algorithm.

The following definitions consider a block partition of the matrices W_N^{-1} and W_N for each pair $(p, q) \in N \times N$. The notation is also particularized for $(p, q) = (p_0, q_0)$.

Consider the blocks of W_N^{-1} and W_N represented in Fig. 2, with the displayed dimensions. For the sake of clearity we explicitly represent the elements of $N_1(k, p, q)$ and $n_1^T(k, p, q)$, using (14)

$$
N_1(k, p, q) = \begin{bmatrix}
a_{k-q}^{k-q} & a_{k-q-1}^{k-q} & \cdots & a_{p-q+1}^{k-q} \\
\vdots & \vdots & & \vdots \\
a_{k-2}^{k-2} & a_{k-3}^{k-2} & \cdots & a_{p-1}^{k-2} \\
a_{k-1}^{k-1} & a_{k-2}^{k-1} & \cdots & a_{p}^{k-1}
\end{bmatrix}
$$

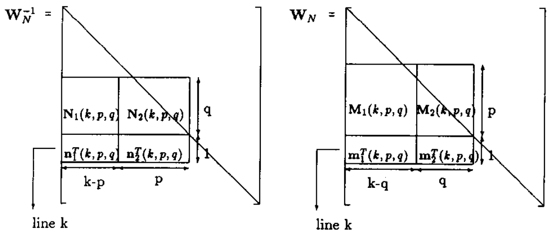
$$
n_1^T(k, p, q) = [a_k^k \;\; a_{k-1}^k \;\; \cdots \;\; a_{p+1}^k]. \tag{20}
$$

Vector $n_1^T(k, p, q)$ collects the $k-p$ highest order coefficients of the kth prediction error filter, while matrix $N_1(k, p, q)$ stacks these for the filters from order $k-1$ to $k-q$. Similarly to (20), the elements of $N_2(k, p, q)$ and $n_2^T(k, p, q)$ are obtained from (14) and Fig. 2(a). Correspondingly equivalent comments hold for the matrix W_N in Fig. 2(b), with the roles of p and q interchanged.

Define the matrix $\mathfrak{M}^T(N, p, q)$ and the vector $m(N, p, q)$ by

$$
\mathfrak{M}^T(N, p, q) = \begin{bmatrix}
M_1^T(\bar{p}, p, q) \\ \hline
M_1^T(\bar{p}+1, p, q) \\ \hline
\vdots \\ \hline
M_1^T(N, p, q)
\end{bmatrix}
$$

$$
m(N, p, q) = \begin{bmatrix}
m_1(\bar{p}, p, q) \\ \hline
m_1(\bar{p}+1, p, q) \\ \hline
\vdots \\ \hline
m_1(N, p, q)
\end{bmatrix}. \tag{21}
$$



(a) (b)

Fig. 2. Block partition of W_N^{-1} and W_N.

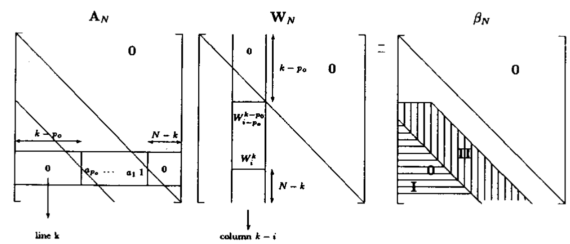


Fig. 3. Geometric interpretation of result 1A.

To simplify the notation, we omit in the above matrices, the explicit dependence on p and q when these equal the true model order parameters, i.e., $p = p_0$ and $q = q_0$.

IV. Parameter Estimation: Basic Relations

We derive a set of relations satisfied by the elements of W_N and W_N^{-1} which are the basis for the AR estimation, for the MA estimation, and for the order selection scheme.

A. Basic Relations for AR Estimation

Let A_N be the matrix (18). We will prove that β_N in (19) is

$$A_N \cdot W_N = \beta_N. \tag{22}$$

It is easy to see that β_N is lower triangular and unit diagonal as A_N is in (18) and W_N is in (14). Result 1A proves that β_N is the band diagonal, non-Toeplitz matrix in (19). Result 4A in section IV-C will define the band diagonal elements of β_N.

Result 1A: The elements of W_N satisfy

$$W_i^k + a_1 W_{i-1}^{k-1} + a_2 W_{i-2}^{k-2} + \cdots + a_{p0} W_{i-p0}^{k-p0} = 0,$$
$$q_0 + 1 \le i \le k, \bar{p}_0 \le k \le N. \tag{23}$$

Proof: In Fig. 3, we present a geometric interpretation of result 1A. The left-hand side of (23) is the product of the line k of A_N by the column $k - i$ of W_N. From (22), this product equals $[\beta_N]_{k,k-i}$, which, for the range of k and i defined in (23), corresponds to the dashed area in matrix β_N represented in Fig. 3.

To prove that the right-hand side (RHS) of (23) is zero involves two arguments, the Cayley–Hamilton theorem for area I and the theory of the invariant directions of the Riccati equation for area II. See Appendix 1 for the details. □

Result 1A says that the AR coefficients define a linear combination of the elements on the same column of the matrix W_N. Each combination defined by (23) corresponds to a null element of β_N below its main diagonal, as shown in Fig. 3.

With the block notation $M_1(k)$ and $m_1(k)$ introduced in Section III and grouping all $k - q_0$ equations in (23) for the same value of k, i.e., in correspondence with all the zeros in line k of β_N, we obtain a set of linear equations satisfied by the AR component coefficients.

Result 1B: The AR component satisfies the system of linear equations

$$M_1^T(k)\, J_{p0} a = -m_1(k), \qquad \bar{p}_0 \le k \le N. \tag{24}$$

□

Using (15), the result 1B may be interpreted as

$$\sum_{r=0}^{p0} a_r \left(\sum_{j=0}^{k-q0-1} W_{k-r-j}^{k-r} z^{-j} \right) = 0, \qquad \forall z \tag{25}$$

that represents a linear combination of the first $k - q_0$ coefficients of the polynomials $\{z^{-(k-p0)} B_{k-p0}(z^{-1}), \cdots, z^{-k} B_k(z^{-1})\}$. As (25) holds for every z, it corresponds to a set of $k - q_0$ linear equations.

Now, collect the linear equations corresponding to all the zeros of areas I and II in Fig. 3 using the definitions (21) for $p = p_0$ and $q = q_0$.

Result 1C: The AR component satisfies the system of linear equations

$$\mathfrak{M}^T(N)\, J_{p0} a = -m(N). \tag{26}$$

Result 1A states that each linear combination corresponds to a zero of β_N below its main diagonal. Result 1B collects all these combinations corresponding to the zeros of β_N in line k. Finally, result 1C represents the system of linear equations in correspondence with all the zeros of β_N below its main diagonal (see the shaded area in Fig. 3).

B. Basic Relations for MA Estimator

For the MA component and the entries of W_N^{-1}, we derive a set of linear equations using

$$\beta_N W_N^{-1} = A_N. \tag{27}$$

These equations are dual from those obtained for the AR component.

Result 2A: The elements of W_N^{-1} satisfy

$$a_i^k + \Omega_1(k) a_{i-1}^{k-1} + \Omega_2(k) a_{i-2}^{k-2} + \cdots + \Omega_{q0}(k) a_{i-q0}^{k-q0}$$
$$= 0, \qquad p_0 + 1 \le i \le k, \bar{p}_0 \le k \le N. \tag{28}$$

Proof: Using the structure of the matrices β_N and W_N^{-1}, in (27), the left-hand side (LHS) of (28) is the product of line k of β_N by column $k - i$ of W_N^{-1}. For the range of indices in (28), the element $[A_N]_{k,k-i}$ is zero (see (18)), thus concluding the proof. □

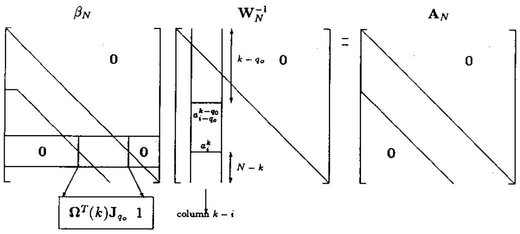


Fig. 4. Geometric interpretation of result 2A.

Fig. 4 is a geometric interpretation of this result. It is dual of Fig. 3.

Result 2A states that the coefficients of vector $\boldsymbol{\Omega}(k)$ in (11) define a linear combination of the elements of the same column of the matrix W_N^{-1}. The same combination, i.e., with the same coefficients, holds for the first $k - p_0$ columns of W_N^{-1}. Consequently, a system of $k - p_0$ linear equations may be established in correspondence with all the zeros on line k and below the main diagonal of A_N in Fig. 4. Result 2A generalizes to all the first $k - p_0$ columns of W_N^{-1} the corresponding relations for the first column (set of reflection coefficients), contained in [8]. As before, a matrix version results by using the notation of Section III.

Result 2B: The vector $\boldsymbol{\Omega}(k)$ satisfies the system of linear equations

$$N_1^T(k)\, J_{q_0}\boldsymbol{\Omega}(k) = -\boldsymbol{n}_1(k), \qquad \forall\, \bar{p}_0 \le k \le N. \tag{29}$$

□

This result may be interpreted in terms of increasing orders prediction error filters. Using (15), (29) may be rewritten as

$$\sum_{r=0}^{q_0} \Omega_r(k)\left(\sum_{j=0}^{k-p_0-1} a_{k-r-j}^{k-r} z^{-j}\right) = 0, \qquad \forall z \tag{30}$$

that represents a linear combination of the first $k - p_0$ coefficients of the polynomials $\{z^{-(k-q_0)}A_{k-p_0}(z^{-1}), \cdots, z^{-k}A_k(z^{-1})\}$. As (30) holds for every z, it corresponds to a set of $k - q_0$ linear equations.

Note that since $\boldsymbol{\Omega}(k)$ depends on k there is no analog to the matrix result 1C.

The MA estimation procedure is based on result 2B and on the following result which relates the MA component of the process with the asymptotic ($k \to \infty$) value of $\boldsymbol{\Omega}(k)$.

Result 3A:

$$\lim_{k \to \infty} \boldsymbol{\Omega}(k) = \boldsymbol{b}. \tag{31}$$

Proof: See Appendix 1. □

Equation (31) relates the MA component of the process with the asymptotic value of $\boldsymbol{\Omega}(k)$. In the sequel, we refer to the vector $\boldsymbol{\Omega}(k)$ as the MA component. This is an abuse of notation, since, by result 3A, only asymptotically does $\boldsymbol{\Omega}(k)$ converge to $\boldsymbol{b}$. In the following result we discuss the rate of this convergence. This clarifies when, for practical purposes, $\boldsymbol{\Omega}(k)$ can be taken as $\boldsymbol{b}$.

Result 3B: The rate of convergence of $\boldsymbol{\Omega}(k)$ to $\boldsymbol{b}$ is determined by the second power of the zeros of the original ARMA process.

Proof: See Appendix 1. □

Given the set of reflection coefficients $\{c_1, c_2, \cdots, c_k\}$, or equivalently the coefficients of the prediction error filters up to order k, the MA estimation problem is solved by obtaining the vector $\boldsymbol{\Omega}(k)$ as the solution of the system of linear equations (29) (see result 2B), and letting $k \to \infty$.

C. Basic Relations for Order Selection

When the correct number of poles p_0 and the correct number of zeros q_0 are not known *a priori* they have to be obtained prior to the estimation procedure. Here we present a set of coupled relations satisfied by the vectors $\boldsymbol{a}$ and $\boldsymbol{\Omega}(k)$ for known p_0 and q_0. These are the basis for step 3 of the order selection algorithm proposed in Section V.

Using (22), the nonnull elements on line k of β_N are obtained next.

Result 4A: The elements $\Omega_i(k)$, $1 \le i \le q_0$ are given by

$$W_i^k + a_1 W_{i-1}^{k-1} + a_2 W_{i-2}^{k-2} + \cdots + a_{p_0} W_{i-p_0}^{k-p_0} = \Omega_i(k)$$

$$1 \le i \le q_0,\ \bar{p}_0 \le k \le N. \tag{32}$$

Proof: The LHS of (32) is the product of the line k of A_N by the column $k - i$ of W_N. From (22), this product equals $[\beta_N]_{k,k-i}$, which, for the range of k and i defined in (32), corresponds to the band elements of β_N in line k. Result 2A holds by defining

$$[\beta_N]_{k,k-i} = \Omega_i(k). \tag{33}$$

We can use the notation in Section III to write (32) in compact form.

Result 4B:

$$M_2^T(k)\, J_{p_0}\boldsymbol{a} + \boldsymbol{m}_2(k) = J_{q_0}\boldsymbol{\Omega}(k). \tag{34}$$

The blocks in the previous algebraic relation are represented in Fig. 5.

By now using matrix equality (27), we establish a set of linear algebraic coupled relations between $\boldsymbol{a}$ and $\boldsymbol{\Omega}(k)$, involving the elements of W_N^{-1}. These are dual from those contained in results 4A and 4B.

Result 5A: The coefficients a_i, $1 \le i \le p_0$, are given by

$$a_i^k + \Omega_1(k)a_{i-1}^{k-1} + \Omega_2(k)a_{i-2}^{k-2} + \cdots + \Omega_{q_0}(k)a_{i-q_0}^{k-q_0} = a_i,$$

$$1 \le i \le p_0,\ \bar{p}_0 \le k \le N. \tag{35}$$

Proof: The LHS of (35) is the product of line k of β_N by the column $k - i$ of W_N^{-1}. From (27), this product equals $[A_N]_{k,k-i}$ which, for the range of k and i defined in (35) and the structure of A_N, equals a_i. □

The matrix version is given next.

Result 5B:

$$N_2^T(k)\, J_{q_0}\boldsymbol{\Omega}(k) + \boldsymbol{n}_2(k) = J_{p_0}\boldsymbol{a}, \qquad \bar{p}_0 \le k \le N. \tag{36}$$

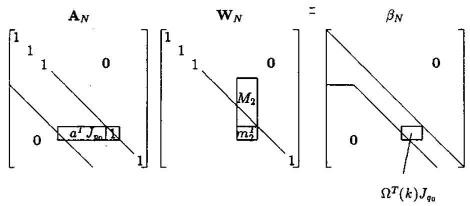


Fig. 5. Blocks presented in result 4B.

For the matrix equality $A_N W_N^{-1} = \beta_N$, the blocks in the previous algebraic relation are dual from these represented in Fig. 5.

D. Polynomial Interpretation

The results presented in Section IV-A through -C may be interpreted in terms of the polynomials $A_n(z^{-1})$ and $B_n(z^{-1})$. In fact, from (22), we have

$$\sum_{j=0}^{p_0} a_j z^{-(k-j)} B_{k-j}(z^{-1}) = z^{-(k-q_0)} \sum_{j=0}^{q_0} \Omega_{q_0-j}(k) z^{-j}, \quad \bar{p}_0 \le k \le N \quad (37)$$

while (27) leads to

$$\sum_{j=0}^{q_0} \Omega_j z^{-(k-j)} A_{k-j}(z^{-1}) = z^{-(k-p_0)} \sum_{j=0}^{p_0} a_{p_0-j}(k) z^{-j}, \quad \bar{p}_0 \le k \le N. \quad (38)$$

Notice that the LHS of (37) ((38)) is a linear combination of increasing order innovation filters (prediction error filters). Results 1A to 5B hold by equating the corresponding powers of z in both sides of (37) and (38). Also, it is clear from the two above equalities that the AR component is computed from a set of increasing orders MA models and that, asymptotically, the same happens for the MA component and increasing orders AR models.

V. ARMA Identification: Order Selection

This section presents the order selection scheme for an ARMA(p_0, q_0) process. When p_0 and q_0 are not known *a priori*, the number of poles and zeros as well as the AR and MA components are estimated through a scheme in Fig. 1 that iterates on the number of poles p and the number of zeros q. In this case, for each pair (p, q), we hypothesize that the data corresponds to an ARMA(p, q) process and evaluate the AR and the MA components of this model. This is a model fit procedure, accomplished by the solution of two systems of linear equations (see step 2 in Fig. 1). These systems are based on those defined in results 1C and 2B in Section IV for (p_0, q_0) but hypothesizing that p and q are the correct model orders.

For (p, q) $\neq$ (p_0, q_0), the evaluated ARMA(p, q) model does not coincide, in general, with the original process. The order selection algorithm is based on a model mismatch evaluation. A functional d quantifies this mismatch. For each pair (p, q), d is evaluated using the AR and the MA components of the fitted ARMA(p, q) model and the two sets of linear algebraic relations. These are established similarly as the coupled relations in results 4B and 5B in Section IV but admitting now that p and q are the correct model orders.

In this section, we assume that all required quantities, namely, the coefficients of the prediction error and of the innovation filters, are available. In practice, these quantities are estimated from the finite sample of data available.

A. Model Fit

We admit that p and q are the correct model orders and fit an ARMA(p, q) model by solving the two systems of linear equations defined similarly to (26) and (29) (see step 2 in Fig. 1).

Definition 1: Let ${}^1\boldsymbol{a}(N, p, q)$ be the minimum-norm vector $\boldsymbol{x}$ that minimizes the Cartesian norm

$$\|\mathfrak{M}^T(N, p, q)\boldsymbol{J}_p\boldsymbol{x} + \boldsymbol{m}(N, p, q)\|_2, \quad N \ge p + q. \quad (39)$$

□

Definition 2: Let ${}^1\boldsymbol{b}(k, p, q)$ be the minimum-norm vector $\boldsymbol{y}$ that minimizes the Cartesian norm

$$\|\boldsymbol{N}_1^T(k, p, q)\boldsymbol{J}_q\boldsymbol{y} + \boldsymbol{n}_1(k, p, q)\|_2, \quad p + q \le k \le N. \quad (40)$$

□

Depending on the rank of the system matrices in (39) and (40), the minimum value of the Euclidean norm in these expressions may not be attainable by a unique solution. Here the uniqueness is obtained by using the minimum-norm solution.

The vectors ${}^1\boldsymbol{a}(N, p, q)$ and ${}^1\boldsymbol{b}(k, p, q)$ are taken as the AR and the MA component estimates of the ARMA(p, q) process fit to the given data. In general, for an arbitrary pair (p, q), there is a nonzero error associated with the two systems of linear equations presented above. Let

$$\boldsymbol{e}_{\mathrm{MA}}(N, p, q) = [{}^{p+q}\boldsymbol{e}_{\mathrm{MA}}(N, p, q) | \cdots | {}^{N-1}\boldsymbol{e}_{\mathrm{MA}}(N, p, q) | {}^{N}\boldsymbol{e}_{\mathrm{MA}}(N, p, q)] \quad (41)$$

be the vector of errors associated with (39). Likewise, define ${}^k\boldsymbol{e}_{\mathrm{AR}}(k, p, q)$ to be the error associated with (40). Using the matrix notation defined in Section III, expressions for these errors follow by replacing the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in (39) and (40) by the corresponding solutions, ${}^1\boldsymbol{a}$ and ${}^1\boldsymbol{b}$.

Result 6:

$$\boldsymbol{e}_{\mathrm{MA}}(N, p, q) = \mathfrak{M}^T(N, p, q)\boldsymbol{J}_p\,{}^1\boldsymbol{a}(N, p, q) + \boldsymbol{m}(N, p, q) \quad (42)$$

$$^k e_{AR}(k, p, q) = N_1^T(k, p, q) J_q \, ^1 b(k, p, q) + n_1(k, p, q), \quad p + q \le k \le N \tag{43}$$

or, equivalently, using (42) and (41)

$$^k e_{MA}(N, p, q) = M_1^T(k, p, q) J_p \, ^1 a(N, p, q) + m_1(k, p, q). \tag{44}$$

□

Note that the independent vector $m(N, p, q)$ in (39) collects elements from line $p + q$ up to line N of W_N, (see (21)), while $n_1(k, p, q)$ in (40) only contains entries in line k of W_N^{-1}. The superindex used on the error notation in (43) and in the right-hand side of (41) stands for the line to which the error corresponds. Therefore, the error associated with (39), $e_{MA}(N, p, q)$ in (41), collects those obtained in correspondence with lines $p + q, \cdots, N$ in W_N.

In the upper part of Fig. 6, we represent graphically the operations leading to the two sets of vectors $\{^1 b, {}^k e_{AR}\}$ and $\{^1 a, {}^k e_{MA}\}$. The duality and independence of the two procedures is preserved when the correct model order is not known.

B. Model Mismatch

When $(p, q) \neq (p_0, q_0)$ there is, in general, a mismatch between the hypothesized model and the correct one. A quantification of this mismatch is proposed. The first step of this quantification uses two sets of coupled and linear algebraic relations defined, for each pair (p, q), as in results 4B and 5B. If we plug the vector $^1 a$ on the (p, q)-version of result 4B, we should obtain a replica of $^1 b$. Similar comments hold if we plug $^1 b$ on the (p, q)-version of result 5B. This suggests the following definitions:

Definition 3:

$$^2 b(k, p, q) = M_2^T(k, p, q) J_p \, ^1 a(N, p, q) + m_2(k, p, q), \quad p + q \le k \le N. \tag{45}$$

□

Definition 4:

$$^2 a(k, p, q) = N_2^T(k, p, q) J_q \, ^1 b(k, p, q) + n_2(k, p, q), \quad p + q \le k \le N. \tag{46}$$

□

These operations are represented in the bottom of Fig. 6.

C. Functional d

When $p = p_0$ and $q = q_0$, we naturally have, according to Section IV,

$$^1 a(N, p_0, q_0) = {}^2 a(N, p_0, q_0) = a \tag{47}$$

$$^1 b(k, p_0, q_0) = {}^2 b(k, p_0, q_0) = \Omega(k) \tag{48}$$

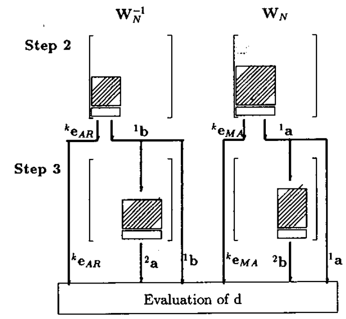


Fig. 6. Steps 2 and 3 of the identification algorithm.

$$^k e_{MA}(N, p_0, q_0) = 0 \tag{49}$$

$$^k e_{AR}(N, p_0, q_0) = 0. \tag{50}$$

These equalities do not generally hold for $(p, q) \neq (p_0, q_0)$.

Therefore, for an arbitrary pair (p, q), we evaluate the mismatch between the correct ARMA(p_0, q_0) and the assumed ARMA(p, q) model as the distance between four pairs of vectors. For each pair (p, q), these correspond to those shown in the first and second terms of the above equalities.

Definition 5: For each pair (p, q), the functional $d(N, p, q)$ is given by

$$d(N, p, q) = d_{AR}(N, p, q) + d_{MA}(N, p, q) \tag{51}$$

with

$$d_{AR}(N, p, q) = \sum_{k=p+q}^{N} \left\| \left[\frac{0_{k-p}}{^1 a(N, p, q)} \right] - \left[\frac{^k e_{AR}(k, p, q)}{^2 a(k, p, q)} \right] \right\|_2^2 \tag{52}$$

$$d_{MA}(N, p, q) = \sum_{k=p+q}^{N} \left\| \left[\frac{0_{k-q}}{^1 b(k, p, q)} \right] - \left[\frac{^k e_{MA}(N, p, q)}{^2 b(k, p, q)} \right] \right\|_2^2 \tag{53}$$

where $\|\cdot\|_2$ is the Euclidean norm. □

D. Properties of d

We present properties of $d(N, p, q)$ that will clarify how d is used in selecting the order. The analysis considers the pairs (p, q) belonging to the four distinct areas shown in Fig. 7.

P1. $\forall (p, q) \in$ Area I, $d(N, p, q) \neq 0$ for $N \ge p_0 + q_0$.

P2. $\forall (p, q) \in$ Area II, $d(N, p, q) \neq 0$ for $N \ge p + q_0$.

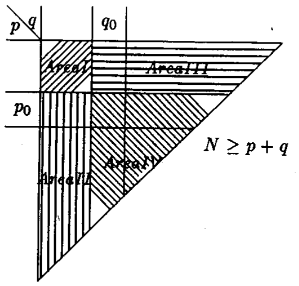


Fig. 7. Distinct areas for the analysis of d.

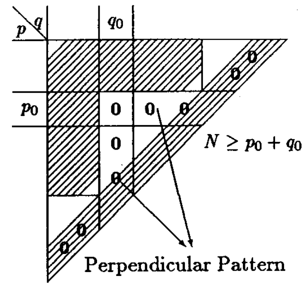


Fig. 8. Locus of the functional d considered in P1 to P5.

P3. $\forall (p, q) \in$ Area III, $d(N, p, q) \neq 0$ for $N \geq p_0 + q$.

P4. $d(N, p, q_0) = 0$ for $p \geq p_0$, $N \geq p + q_0$.

P5. $d(N, p_0, q) = 0$ for $q \geq q_0$, $N \geq p_0 + q$.

In Fig. 8, we sketch the locus corresponding to P1 to P5. The symbol **0** represents the set of pairs (p, q) where d is zero, while the shaded area represents a nonnull value of d. The white triangles in Fig. 8 are not covered by these properties.

The proof of these properties is based on the analysis of the solution of (39) and (40) as well as their associated errors. Due to their similarity we will briefly discuss the general case with z being the minimum norm vector that minimizes the Euclidean norm $\|Az + b\|_2$, with A an $m \times n$ matrix, with $m \leq n$ and rank $A = k \leq n$. The solution of this minimization problem is given by $z = -A^{\dagger}b$, where $A^{\dagger}$ is the Moore–Penrose inverse of A [22]. For a full rank matrix, $A^{\dagger} = (A^TA)^{-1}A^T$ [22]. If the independent vector b belongs to the range space of A, the error associated with the minimization is zero.

For the properties P1 to P3, notice from Fig. 7 that the ARMA(p, q) models computed for areas I–III have a number of poles p less than the true value p_0 and/or a number of zeros q less than q_0. For these models, the independent vectors m and $\boldsymbol{n}_1$ in the minimization problems (39) and (40) do not belong to the range space of the corresponding matrices, $\mathfrak{M}^T$ and N_1^T [25]. This leads to nonnull associated errors and, consequently, to a nonnull value of d (see (51)–(53)), justifying P1 to P3.

For ARMA(p, q) models corresponding to area IV, i.e., with both p and q greater than or equal to their correct values p_0 and q_0, the errors ${}^k\boldsymbol{e}_{\mathrm{AR}}$ and $\boldsymbol{e}_{\mathrm{MA}}$ associated with (39) and (40) are zero. In particular, this holds for the set of ARMA models considered in P4, i.e., with the correct number of zeros q_0 and a number of poles p greater than or equal to the correct one p_0. For these latter models in P4, the vectors, ${}^1\boldsymbol{b}$, ${}^2\boldsymbol{b}$, ${}^1\boldsymbol{a}$, and ${}^2\boldsymbol{a}$ given by definitions 1 to 4 are pairwise equal. Together with the null errors, this leads to a null d (see (51)–(53)), justifying P4. The location of the zeros corresponding to ${}^1\boldsymbol{b}$ coincide with those of the correct ARMA(p_0, q_0) model while the vector ${}^1\boldsymbol{a} \in R^p$ leads to p_0 poles at the correct locations, the remaining $p - p_0$ being at the origin. A dual explanation holds for P5. See [25] for a proof of these properties. For additional properties of the functional d and for a particularization to the class of Butterworth processes see [25].

E. Order Selection Algorithm

From the above discussion and list of properties, it results that for $N \geq p_0 + q_0$, a perpendicular pattern of the row with $p = p_0$ and the column with $q = q_0$ arises on the table of d, as shown in Fig. 8. This pattern is established for $N = p_0 + q_0$, and the length of its row and column increases as N increases. Every ARMA(p, q) model with (p, q) in the pattern has the same spectrum as the one corresponding to the intersection point [25]. In fact, the additional $p - p_0$ poles or $q - q_0$ zeros of these models are located at the origin. The order selection algorithm identifies this pattern and chooses the model with the smallest number of parameters. For the exact knowledge of the reflection coefficient sequence, this model has p_0 poles and q_0 zeros. Define the sets

$$\mathcal{I}(N, p, q) = \{(p, q): d(N, p, q) = 0 \vee p + q > N\} \tag{54}$$

$$\mathcal{L}(p, q) = \bigcap_{N \geq 1} \mathcal{I}(N, p, q) \tag{55}$$

where $\cap$ stands for set intersection. For each value of N, the set $\mathcal{I}(N, p, q)$: i) does not include the orders (p, q) of ARMA models for which $d(N, p, q) \neq 0$ for $N \geq p + q$, i.e., $\{(p, q): d(N, p, q) \neq 0,\ p + q \leq N\}$. These models are definitely not the correct one; ii) collects all the models for which there is not enough information for an accept or reject decision, i.e., $\{(p, q): d(N, p, q) = 0,\ p + q \leq N\}$ or $\{(p, q): p + q > N\}$.

The properties of d yield

$$\mathcal{P} = \{(p, q): (p = p_0, q \geq q_0) \vee (p \geq p_0, q = q_0)\} \subset \mathcal{L}(p, q)$$

i.e., models corresponding to the perpendicular pattern $\mathcal{P}$ with (p_0, q_0) as its intersection point are a subset of $\mathcal{L}(p, q)$. The model order (p_0, q_0), i.e., corresponding to the true orders, is obtained as the order of the model ARMA(p, q) belonging to $\mathcal{L}(p, q)$ with the smallest number of parameters

$$(p_0, q_0) = \arg\min_{(p,q) \in \mathcal{L}(p,q)} \{p + q\}. \tag{56}$$

When the order has been decided, the algorithm accepts the vectors ${}^1\boldsymbol{a}$ and ${}^1\boldsymbol{b}$ from the previous step 2 as the AR and MA parameter estimates. In the next section we modify this selection scheme for the case where the exact reflection coefficient sequence is replaced by an estimated one.

VI. LD^2–ARMA Algorithm Implementation

In practice, the identification has to be carried out based on a finite sample of length L. The proposed identification algorithm is implemented following the lines in Section V. The nonavailable quantities, namely, the successively increasing order prediction and innovation filter coefficients, are replaced by suitable estimates. Let $\hat{s}$ represent the estimated value of a quantity s. Then, starting from the set of observations, $\{y_0, y_1, \cdots, y_{L-1}\}$, the LD^2–ARMA iterates as:

1) It obtains an estimate of the reflection coefficient sequence $\{\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_N, \cdots\}$ using the Burg technique [9].

2) Computes the estimated matrix $\widehat{W_N^{-1}}$ for increasing values of N using point 1 above and the Levinson algorithm.

3) Computes $\widehat{W_N}$ by a recursive inversion of $\widehat{W_N^{-1}}$.

4) Hypothesizes that the data belongs to an ARMA(p, q) model, where $p + q \leq N$.

5) Fits the ARMA(p, q) model to the data, by solving (39) and (40) with the corresponding matrices replaced by its estimated values. The solutions are the AR and MA component estimates of the assumed model. If p_0 and q_0 are known *a priori*, the correct model structure is assumed in point 4, with no further processing being needed. If not, it proceeds to step 6.

6) Evaluates the model mismatch, $d(N, p, q)$, by computing the vectors in definitions 3 and 4 in Section V with the corresponding matrices replaced by its estimated values.

7) Tests the hypothesis. Due to the estimation errors on the reflection coefficient sequence, the functional properties P4 and P5 do not hold exactly. Therefore, the order selection procedure implements

$$(\hat{p}, \hat{q}) = \underset{(p,q) \in \mathcal{L}(p,q)}{\arg\min} \{p + q\} \tag{57}$$

modifying the definition presented in Section V of the set $\mathcal{L}(p, q)$ as

$$\mathcal{L}(p, q) = \bigcap_{N \geq 1}^{N_{\max}} \mathcal{I}(N, p, q) \tag{58}$$

$$\mathcal{I}(N, p, q) = \{(p, q): d(N, p, q) < \epsilon \vee p + q > N\} \tag{59}$$

where ϵ is a small positive constant and $N_{\max}$ is the maximum number of reflection coefficients estimated from the data.

If the hypothesis is accepted, the identification procedure is completed. Otherwise, it returns to point 4.

TABLE I

	Poles	Zeros	σ^2
Example 1 ARMA(2, 1)	0.9 −0.8	−0.5	1
Example 2 ARMA(6, 4)	0.95 exp [$\pm j50°$] 0.9 exp [$\pm j90°$] 0.95 exp [$\pm j130°$]	0.95 exp [$\pm j70°$] 0.95 exp [$\pm j110°$]	1
Example 3 ARMA(6, 4)	0.95 exp [$\pm j40°$] 0.9 exp [$\pm j90°$] 0.95 exp [$\pm j140°$]	0.95 exp [$\pm j65°$] 0.95 exp [$\pm j115°$]	1

TABLE II
EXAMPLE 1

	$L = 250$	$L = 500$	$L = 1000$	$L = 5000$
LD^2	46	62	85	100
AIC	54	43	45	45
MDL	91	96	98	100

TABLE III
EXAMPLES 2 AND 3

	Example 2			Example 3		
	LD^2 $\epsilon = 0.2$	AIC	MDL	LD^2 $\epsilon = 0.07$	AIC	MDL
$L = 500$	24	30	39	27	16	21
$L = 1000$	37	30	38	63	14	20
$L = 5000$	96	49	51	100	21	22

In the algorithm, points 1 and 3 are not essential, i.e., alternative estimates for W_N and W_N^{-1} can be used, e.g., [15].

Although a more thorough analysis of the algorithm is presented elsewhere [28], here we compare briefly the order determination characteristics of the LD^2–ARMA with the Akaike information criterion (AIC) introduced in [1] and the MDL suggested by Akaike [4], and developed by Rissanen [29]. The simulated results count the number of correct order selections ($\hat{p} = p_0$, $\hat{q} = q_0$), over 100 independent Monte Carlo experiments for the ARMA processes with pole-zero location displayed in Table I. The AIC and MDL procedures were implemented using the software package in [3].

The results for example 1 are presented in Table II where $\epsilon = 0.05$ is the threshold considered for the LD^2 order selection algorithm (see (59)). The results of examples 2 and 3 are displayed in Table III.

Tables II and III confirm the well-known statistical inconsistency of the AIC. On the contrary, the error in choosing the correct order goes to zero as $L \to \infty$ for both the MDL and the LD^2 algorithm. In example 1, which is a low-order ARMA model, for small sample sizes, MDL outperforms the LD^2–ARMA algorithm. When the order of the ARMA system increases as in examples 2 and 3, the LD^2–ARMA seems to have the upper hand over both AIC and MDL. The convergence of LD^2 is clearly faster as exhibited by the results for $L = 5000$ in example 2 and for $L \geq 1000$ for example 3.

VII. Conclusions

The LD^2–ARMA identifier described solves the nonlinear ARMA identification problem. It combines an order selection scheme with a linear, dual, decoupled algorithm for the estimation of the AR and of the MA components.

The essence of the algorithm explores linear relations between corresponding lag coefficients of successively higher order linear predictor and innovation filters that are fit to the data. The coefficients of these linear relations are exactly the AR parameters and (asymptotically) the MA parameters. These linear relations are of two natures: the uncoupled systems (26) and (29), which only involve either the AR or the MA coefficients; and the coupled relations (34) and (36), which express the AR (MA) parameters in terms of the MA(AR) parameters. By solving the uncoupled systems independently, we decouple the AR and MA estimation procedures. The order selection scheme uses the coupled system of equations. Once we have potential candidates for the AR and for the MA parameters, these are introduced in the coupled relations. The order estimate minimizes the resulting mismatch.

In practice, to use the algorithm described here, we first need to obtain suitable estimates for the coefficients of the higher order linear predictor and linear innovation fits to the data. We use estimates of these quantities obtained not from the covariance sequence but from the sequence of reflection coefficients. Section VI has presented simulation examples that compare the order selection capabilities of LD^2, AIC, and MDL. It is apparent that LD^2 and MDL are both consistent, and that as the order and complexity of the ARMA model increase, LD^2 exhibits a faster convergence rate. The companion paper [28] analyzes the statistical performance of the LD^2 identification algorithm and presents simulation results that further detail the behavior of the algorithm and compare it with other competing techniques.

Appendix I
Markovian/Innovation Representations

The ARMA(p_0, q_0) process admits a Markovian representation of order

$$\bar{p}_0 = \max\{p_0, q_0\} \tag{A.1}$$

given by

$$x_{n+1} = Fx_n + ge_n \tag{A.2}$$

$$y_n = hx_n + \sigma e_n \tag{A.3}$$

where $x_n \in R^{\bar{p}_0}$; the matrix F and the vectors g and h have consistent dimensions with those of x_n; and x_0 is a random vector with zero mean and covariance matrix P_0, satisfying

$$P_0 = FP_0F^T + gg^T. \tag{A.4}$$

In representation (A.2), (A.3), the state and output noises are totally correlated. For $p_0 < q_0$, the order of this state representation requires the introduction of additional poles at the origin. Alternative parameterizations are given in [6], [8], [15].

The discrete time Kalman–Bucy filter associated with (A.2), (A.3) leads to an innovation representation of $\{y_n\}$:

$$\hat{x}_{n+1} = F\hat{x}_n + k_n v_n \tag{A.5}$$

$$y_n = h\hat{x}_n + v_n \tag{A.6}$$

$$\hat{x}_0 = 0 \tag{A.7}$$

where $\{v_n\}$ is the innovation sequence associated with the ARMA process, defined by (5). This process is white, with zero mean and variance

$$d_n = E[v_n^2] \tag{A.8}$$

given by

$$d_n = hP_nh^T + \sigma^2 \tag{A.9}$$

where P_n satisfies the discrete Riccati equation

$$P_{n+1} = FP_nF^T - k_nd_nk_n^T + gg^T \tag{A.10}$$

$$k_n = (FP_nh^T + g\sigma)d_n^{-1} \tag{A.11}$$

with initial condition P_0 given by (A.4). In (A.5), k_n is the filter's gain given by (A.11).

Proof of Result 1A: The result is proved separately for the elements of β_N in areas I and II (see Fig. 3), and characterized by

$$[\beta_N]_{k,k-i} \in \text{Area I} \quad \text{iff } \bar{p}_0 \leq k \leq N, \quad 0 \leq k - i \leq k - \bar{p}_0 - 1 \tag{A.12}$$

$$[\beta_N]_{k,k-i} \in \text{Area II} \quad \text{iff } \bar{p}_0 \leq k \leq N, \quad k - \bar{p}_0 \leq k - i \leq k - q_0 - 1. \tag{A.13}$$

The product of line k of A_N by column $k - i$ of W_N leads to

$$W_i^k + \sum_{j=1}^{p_0} a_j W_{i-j}^{k-j} = [\beta_N]_{k,k-i} \tag{A.14}$$

where W_l^r satisfies (13). Solving iteratively the state equation (A.5) and comparing with (9), yields for $k \geq 0$

$$W_i^k = \begin{cases} 0, & \text{for } i < 0 \\ 1, & \text{for } i = 0 \\ hF^{i-1}k_{k-i}, & \text{for } i > 0. \end{cases} \tag{A.15}$$

For the set of indices k and $k - i$ belonging to area I, the right-hand side of (A.14) is zero due to the Cayley–Hamilton theorem when applied to matrix F. Therefore, all the elements in area I are zero. For area II

$$q_0 + 1 \le i \le \bar{p}_0, \quad \bar{p}_0 \le k \le N \quad (A.16)$$

which means, together with (A.1), that this area is nonempty iff $\bar{p}_0 = p_0 < q_0$ or, equivalently, iff F is a nonsingular matrix. Replacing (A.15) in (A.14) for the set of indices (A.16) and using the Cayley–Hamilton theorem on F

$$[\beta_N]_{k,k-i} = a_i - \sum_{j=i}^{p_0} a_j hF^{i-j-1}k_{k-i}, \qquad q_0 + 1 \le i \le p_0. \quad (A.17)$$

The Riccati equation associated with the ARMA(p_0, q_0) scalar process (1), has $p_0 - q_0$ independent invariant directions, the set of vectors $\{(F^{-T})^j h^T, 1 \le j \le p_0\}$ being a basis of the invariant direction space, i.e., $P_n(F^{-T})^j h^T = 0$ for $n \ge j, j \le p_0 - q_0$ [7]. This result, and the fact that the first $p_0 - q_0$ values of the anticausal impulse response of (A.2), (A.3) are zero, leads to

$$hF^{-j}k_n = \begin{cases} 1 & \text{for } j = 1 \\ 0 & \text{for } 2 \le j \le p_0 - q_0. \end{cases} \quad (A.18)$$

Rewriting (A.17) as

$$[\beta_N]_{k,k-i} = a_i - a_i hF^{-1}k_{k-i} - \sum_{j=2}^{p_0-i+1} a_{i+j-1} hF^{-j}k_{k-i}, \qquad q_0 + 1 \le i \le p_0 \quad (A.19)$$

and using (A.18), it follows that all the elements on area II are zero, concluding the proof. □

Proof of Result 3A: The proof is in two steps. First we establish a closed form for $\Omega_i(k)$ as a function of the solution of the Riccati equation (A.10). Second, the value of $\Omega_i(k)$ as $k \to \infty$ is obtained from the asymptotic behavior of the Riccati equation and from the Markov parameters of the model's process. In order to simplify the proof, we will consider that $p_0 \ge q_0$. See [25], for arbitrary p_0 and q_0.

From (32), $\Omega_i(k)$, $1 \le i \le q_0$ is the product of line i of A_N by column $k - i$ of W_N, i.e.,

$$\Omega_i(k) = W_i^k + \sum_{j=1}^{i} a_j W_{i-j}^{k-j}. \quad (A.20)$$

Replacing (A.11) in (A.15) and this last equation in (A.20) yields

$$\Omega_i(k) = a_i + \left[\sum_{j=1}^{i} a_{i-j} hF^j P_{k-i} h^T + \sum_{j=1}^{i} a_{i-j} hF^{j-1} g\sigma \right] d_{k-i}^{-1}. \quad (A.21)$$

The two representations of the ARMA process (1) and (A.2), (A.3) are equivalent. Consequently, the transfer function $\Phi(z)$ of the corresponding models are equal

$$\Phi(z) = \frac{\sigma B(z)}{A(z)} = \sigma + h(zI - F)^{-1} g. \quad (A.22)$$

Expanding both sides of (A.22) in a power series of z^{-1}, and equating the coefficients in the same powers, yields

$$\sigma b_i = a_i \sigma + \sum_{j=0}^{i-1} a_j hF^{i-1-j} g, \qquad i \ge 1. \quad (A.23)$$

Replacing (A.23) in (A.21) and using (7), the coefficients $\Omega_i(k)$ may finally be written as

$$\Omega_i(k) = \left[b_i \sigma^2 + \sum_{j=0}^{i} a_{i-j} hF^j P_{k-i} h^T \right] d_{k-i}^{-1}. \quad (A.24)$$

As $k \to \infty$, the solution of the discrete Riccati equation (A.10) goes to zero, [8], [25], and thus from (A.24), result 3A immediately holds. □

Proof of Result 3B: From (A.24), the evolution of $\Omega_i(k)$ is determined by that of the solution of the Riccati equation P_k (A.10). Therefore, the rate of convergence of $\Omega_i(k)$ to b_i is governed by that of P_k to its steady-state solution, the null matrix. We discuss the rate of this convergence by using the theory of the invariant directions of the Riccati equation [7]. As in result 5A, we will consider $p_0 \ge q_0$. See [25] for arbitrary p_0 and q_0 as well as for the generalization of this result for multivariable ARMA processes, where we use results of the invariant directions for multivariable systems from [23].

Let S be a linear transformation of order p_0, whose first $p_0 - q_0$ lines are a basis of the space of invariant directions of the Riccati equation (A.10) [7]:

$$S^T = [hF^{-1} | hF^{-1} | \cdots | hF^{-(p_0-q_0)} | S_1^T] \quad (A.25)$$

and S_1 is any $q_0 \times p_0$ matrix such that S is nonsingular. Under the transformation S, the solution of the Riccati equation (A.10) is given by [7]

$$\tilde{P}_k = SP_kS^T = \left[\begin{array}{c|c} \mathbf{0}_{(p_0-q_0)\times(p_0-q_0)} & \mathbf{0}_{(p_0-q_0)\times q_0} \\ \hline \mathbf{0}_{q_0\times(p_0-q_0)} & Q_k \end{array} \right], \qquad k \ge p_0 - q_0 \quad (A.26)$$

where Q_k satisfies a reduced q_0 order discrete Riccati equation

$$Q_{k+1} = \tilde{F}_{22} Q_k \tilde{F}_{22}^T - \tilde{Q}_k \tilde{k}_k^T + \tilde{g}_2 \tilde{g}_2^T, \qquad k \ge p_0 - q_0 \quad (A.27)$$

$$\tilde{k}_k = (\tilde{F}_{22} Q_k \tilde{h}_2^T + \tilde{g}_2 \sigma) \tilde{d}_k^{-1} \quad (A.28)$$

$$\tilde{d}_k = (\tilde{h}_2 Q_k \tilde{h}_2^T + \sigma^2) \quad (A.29)$$

and $\{\tilde{F}_{22}, \tilde{g}_2, \tilde{h}_2, \sigma\}$ is a reduced order system obtained from $\{F, g, h, \sigma\}$ under the transformation S and the block partition

$$\tilde{F} = SFS^{-1} = \left[\begin{array}{c|c} \tilde{F}_{11} & \tilde{F}_{12} \\ \hline \tilde{F}_{21} & \tilde{F}_{22} \end{array} \right], \qquad \tilde{g}_2 = Sg = \left[\begin{array}{c} \tilde{g}_1 \\ \hline \tilde{g}_2 \end{array} \right]$$

$$\tilde{h} = hS^{-1} = [\tilde{h}_1 | \tilde{h}_2]. \quad (A.30)$$

From (A.26), the rate of convergence of P_k is that of Q_k, this being determined by the rate of convergence to zero of

$$\mathcal{Q}_k = \Delta^{-[k-(p_0-q_0)]} U \Delta^{-[k-(p_0-q_0)]} \quad \text{(A.31)}$$

where U is a constant matrix, $\Delta = \text{diag}\{\lambda_i, i = 1, \cdots, q_0\}$ and $\{\lambda_i, i = 1, \cdots, q_0\}$ are the unstable eigenvalues of the Hamiltonian matrix associated with the reduced order Riccati equation (A.27)

$$\mathcal{H} = \begin{bmatrix} (R^T)^{-1} & (R^T)^{-1} \tilde{h}_2^T \tilde{h}_2/\sigma^2 \\ \mathbf{0} & R^T \end{bmatrix}, \quad R = \tilde{F}_{22} - \tilde{g}_2\tilde{h}_2/\sigma. \quad \text{(A.32)}$$

The characteristic equation associated with (A.32) is given by

$$\det(\lambda I - \mathcal{H}) = \det(\lambda I - R^T)\det(\lambda I - (R^T)^{-1}) \quad \text{(A.33)}$$

its zeros being pairwise inverse. Finally, we prove that the eigenvalues of R

$$\{\lambda : \det(\lambda I - R) = \det(\lambda I - \tilde{F}_{22} + \tilde{g}_2\tilde{h}_2/\sigma) = 0\} \quad \text{(A.34)}$$

coincide with the zeros of the reduced order system $\{\tilde{F}_{22}, \tilde{g}_2, \tilde{h}_2, \sigma\}$ and these with the zeros of the original system. The zeros of the system (A.2), (A.3) are unchanged under a coordinate transformation, being given by

$$\{\lambda : \det[\sigma + \tilde{h}(\lambda I - \tilde{F})^{-1}\tilde{g}] = 0\}. \quad \text{(A.35)}$$

The determinant in (A.35) is

$$\det[\sigma + \tilde{h}(\lambda I - \tilde{F})^{-1}\tilde{g}] = \frac{\sigma}{\det(\lambda I - \tilde{F})}\det[\lambda I - (\tilde{F} - \tilde{g}\tilde{h}/\sigma)]. \quad \text{(A.36)}$$

From the transformation S in (A.25) and using (A.30),

$$\tilde{F} - \tilde{g}\tilde{h}/\sigma = \begin{bmatrix} \mathbf{0}_{(p_0-q_0)\times(p_0-q_0)} & \mathbf{0}_{(p_0-q_0)\times q_0} \\ \tilde{F}_{12} - \tilde{g}_2\tilde{h}_1/\sigma & \tilde{F}_{22} - \tilde{g}_2\tilde{h}_2/\sigma \end{bmatrix} \quad \text{(A.37)}$$

and

$$\det[\lambda - (\tilde{F} - \tilde{g}\tilde{h}/\sigma)] = \lambda^{p_0-q_0}\det[\lambda I - (\tilde{F}_{22} - \tilde{g}_2\tilde{h}_2/\sigma]. \quad \text{(A.38)}$$

Also, from (A.36),

$$\det[\sigma + \tilde{h}_2(\lambda I - \tilde{F}_{22})^{-1}\tilde{g}_2] = \frac{\sigma}{\det(\lambda I - \tilde{F}_{22})}\det[\lambda I - (\tilde{F}_{22} - \tilde{g}_2\tilde{h}_2/\sigma)] \quad \text{(A.39)}$$

which together with (A.34), means that the eigenvalues of R coincide with the zeros of the reduced order system. From (A.38), these zeros coincide with the q_0 nonnull zeros of the original system. This concept is related to a result from optimal control connecting the closed loop system poles to the nonzero open loop transfer function zeros [17].

References

[1] H. Akaike, "Fitting autogressive models for prediction," *Ann. Inst. Stat. Math.*, vol. 21, pp. 243-247, 1969.

[2] H. Akaike, "Maximum likelihood identification of autoregressive moving average models," *Biometrika*, vol. 60, no. 2, pp. 255-265, 1973.

[3] H. Akaike, E. Arahata, and T. Ozaki, *A Time Series Analysis and Control Program Package*. Computer Science Monographs, Institute of Statistical Mathematics, Mar. 1975.

[4] H. Akaike, "On entropy maximization principle," in *Applications of Statistics*, P. R. Krishnaiah, Ed. Amsterdam: North Holland, pp. 27-41.

[5] G. Alengrin, R. S. Bucy, J. Moura, J. Pagés, and M. I. Ribeiro, "ARMA identification," *J. Optimiz. Theory Appl.*, vol. 55, no. 3, pp. 345-357, Dec. 1987.

[6] G. Alengrin and J. Zerubia, "A method to estimate the parameters of an ARMA model," *IEEE Trans. Automat. Contr.*, vol. 32, no. 12, pp. 1113-1115, Dec. 1987.

[7] R. S. Bucy, D. Rappaport, and L. M. Silverman, "Correlated noise filtering and invariant directions for the Riccati equation," *IEEE Trans. Automat. Contr.*, vol. 15, no. 5, pp. 535-540, Oct. 1970.

[8] R. S. Bucy, "Identification and filtering," *Math. Syst. Theory*, vol. 16, pp. 307-317, Dec. 1983.

[9] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford Univ., May 1975.

[10] J. A. Cadzow, "High performance spectral estimation—a new ARMA method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 5, pp. 524-529, Oct. 1980.

[11] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, no. 9, pp. 907-939, Sept. 1982.

[12] C. J. Demeure and L. L. Scharf, "Linear statistical models for stationary sequences and related algorithms for Cholesky factorization of Topelitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 1, pp. 29-42, Jan. 1987.

[13] J. P. Dugré, L. L. Scharf, and C. Gueguen, "Exact likelihood for stationary vector autoregressive moving average processes," *Signal Processing*, vol. 11, pp. 105-118, 1986.

[14] B. Friedlander and B. Porat, "The modified Yule-Walker method of ARMA spectral estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 20, no. 2, pp. 158-172, Mar. 1984.

[15] C. Gueguen and L. Scharf, "Exact maximum-likelihood identification of ARMA models: A signal processing perspective," in *Signal Processing: Theory and Applications*, M. Kunt and F. de Coulon, Eds. North-Holland, 1980, pp. 759-769.

[16] S. Kay, *Modern Spectral Analysis*. Englewood Clliffs, NJ: Prentice-Hall, 1988.

[17] R. Kalman,"When is a linear control system optimal?," *ASME J. Basic Eng.*, ser. D, pp. 51-66, Mar. 1964.

[18] S. Li and B. W. Dickinson, "Application of the lattice filter to robust estimation of AR and MA models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 502-512, Apr. 1988.

[19] S. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[20] J. M. F. Moura and M. I. Ribeiro, "Parametric spectral estimation for ARMA processes," in *Proc. 3rd ASSP Workshop Spectrum Estimation Modeling* (Boston, MA), Nov. 1986, pp. 37-40.

[21] J. M. F. Moura and M. I. Ribeiro, "Linear estimation of ARMA processes," in *Proc. 4th ASSP Workshop Spectrum Estimation Modeling* (Minneapolis, MN), Aug. 1988.

[22] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 2nd ed. Englewood Cliffs, NJ: 1977.

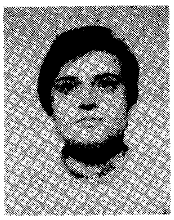
[23] M. Pavon, "Stochastic realization and invariant directions of the matrix Riccati equation," *SIAM J. Contr. Optimiz.*, vol. 18, no. 2, pp. 155-180, Mar. 1980.

[24] M. I. Ribeiro and J. M. F. Moura, "Dual estimation of the poles and zeros of an ARMA(p, q) process," Instituto Superior Técnico, CAPS Tech. Rep. 02/84, 1984 (revised as Tech. Rep. LIDS-P-1521, Massachusetts Institute of Technology, 1985).

[25] M. I. Ribeiro, "Identificação de processos autoregressivos de média móvel," Ph.D. dissertation, Instituto Superior Técnico, Tech. Univ. Lisbon, Mar. 1988; also Tech. Rep., LASIP, Elec. Comput. Eng. Dep., Carnegie-Mellon Univ., Pittsburgh, PA.

[26] M. I. Ribeiro and J. M. F. Moura, "ARMA processes: Order estimation," in *Proc. IEEE 1988 Int. Conf. Acoust., Speech, Signal Processing* (New York), Apr. 1988, pp. 2308-2311.

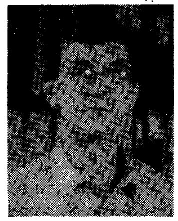
[27] M. I. Ribeiro, J. Zerubia, J. M. F. Moura, and G. Alengrin, "Comparison of two ARMA estimators," in *Proc. IEEE 1989 Int. Conf. Acoust., Speech, Signal Processing* (Glasgow, Scotland), May 1989, pp. 2186–2189.
[28] M. I. Ribeiro and J. M. F. Moura, "Statistical analysis of the LD^2 ARMA estimator," submitted for publication.
[29] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
[30] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, no. 9, pp. 975–989, Sept. 1982.
[31] G. C. Tiao and R. S. Tsay, "Consistency properties of least squares estimates of autoregressive parameters in ARMA models," *Ann. Stat.*, vol. 11, 1983.



M. Isabel Ribeiro (S'82–A'89–M'90) was born in Lisboa, Portugal in 1955. She received the E.E. degree in 1978, the M.Sc. and the Ph.D. degrees in electrical and computer engineering in 1983 and 1988, all from Instituto Superior Técnico (IST), Lisboa, Portugal.

In 1978, she joined the faculty of the Department of Electrical and Computer Engineering of IST where she has been an Assistant Professor (1988–1991) and an Associate Professor since 1991. She is also a Researcher at Centro de Análise e Processamento de Sinais (CAPS) in Lisboa, Portugal. She has taught courses in linear systems theory, control, and spectral methods and identification. Her research interests lie on control applications, robotics, and spectral methods in signal processing. At CAPS, she has been involved in the development of an autonomous guided vehicle for an industrial environment. She has presented her work at several international scientific meetings.

Dr. Ribeiro is a member of several societies of the IEEE and a member of Ordem dos Engenheiros.



José M. F. Moura (S'71–M'75–SM'90) received the engenheiro electrotécnico degree in 1969 from Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. in EECS degrees from the Massachusetts Institute of Technology (M.I.T.), Cambridge, in 1973 and 1975, respectively.

In 1975, he joined the Faculty of IST where he was an Assistant Professor, Professor Agregado (1978), and Professor Catedrático (1979). From 1984 to 1986 he was Genrad Associate Professor of Electrical Engineering and Computer Science (visiting) at M.I.T., also associated with LIDS. Since 1986, he has been a Professor of Electrical and Computer Engineering at Carnegie-Mellon University (CMU), Pittsburgh. He spent the summers of 1978–1981 as a Visiting Scientist with the Department of Aerospace Engineering, University of Southern California (USC), Los Angeles. His research interests lie in statistical signal processing (one and two dimensional), communications, and systems theory. He has over 80 technical contributions, including invited ones, published in international journals and conference proceedings, and is Coeditor of the book *Nonlinear Stochastic Problems* (Dordrect, The Netherlands: Reidel, 1983).

Dr. Moura is an Associate Editor for TRANSACTIONS ON SIGNAL PROCESSING (Sonar and Radar), is a member of several IEEE societies, of Sigma Xi, AMS, IMS, SIAM, and Ordem dos Engenheiros.