E3.2

Comparison of two ARMA Estimators

M.Isabel Ribeiro
CAPS/INIC, Instituto Superior Técnico, P-1096 Lisbon Codex, Portugal
Josiane Zerubia
University of Southern California and Signal and Image Processing Inst., USA
on leave from LASSY, University of Nice, France
José M. F. Moura
Dept. of ECE, Carnegie Mellon University, Pittsburgh PA 15213, USA
Gerard Alengrin
LASSY, University of Nice, Nice, France

Abstract

In this paper, two alternative ARMA estimators are compared both theoretically and through simulation analysis. The first is a dual algorithm which estimates the MA and the AR components as the solution of two linear and independent systems of equations. On the second estimator, the AR coefficients result from a system of linear equations, while the MA component is obtained from a fast filtering algorithm initialized with the previous AR estimated coefficients.

Introduction

The estimation of ARMA or AR+noise processes is an area of active research. Techniques based on Maximum Likelihood require a large computational effort and may fail to converge. Here, we compare by simulation two alternative methods recently proposed [1], [6]. These use the reflection coefficient sequence estimated directly from the data.

The first method, [6], is a dual algorithm that estimates the MA and the AR components as the solution of two linear, independent, and overdimensioned systems of equations. These are obtained from the coefficients of the innovation and prediction error filters of increasing orders associated with the ARMA process. With this method, the MA estimation does not depend on the determination of the AR component. Likewise, the AR estimation results independently of the MA estimation. The method effectively decouples the two tasks of AR and MA identification.

The second technique, [1], uses only the prediction error filter coefficients. The AR component is obtained by combining the first step predictor and a transient Kalman gain in a two stage procedure that solves a system of linear equations and implements a set of linear algebraic relations. The estimation of the MA component starts from the previous AR estimated coefficients and is based on a fast nonlinear iterative algorithm derived from the Chandrasekar equations.

In Section 2, we formulate the estimation problem. The two ARMA estimation algorithms are briefly presented and compared in Section 3. Simulation results are displayed in Section 4. Finally, Section 5 concludes the paper.

Problem Formulation

Let $\{y_n\}$ satisfy the ARMA(p,q) recursion

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

where $\{e_n\}$ is a white noise process with zero mean and variance σ^2. We assume that (1) is stable, minimum-phase, and that the associated transfer function has no common poles and zeros. The model structure, i.e., the number of poles, p, and zeros, q, is known a priori, see [7], [9] for an order determination algorithm. For the sake of simplicity, we assume that $p \geq q$, although both estimation algorithms apply for $p < q$. The problem is to estimate the AR and MA coefficients from a finite sample of the process. Define $\{v_n\}$

$$v_n = y_n - E[y_n \mid y_0, y_1, \cdots, y_{n-1}] \qquad (2)$$

as the innovation process associated with the ARMA process and let $\{a_i^n,\ 0 \leq i \leq n\}$ be the coefficients of the prediction error filter of order n,

$$v_n = \sum_{i=0}^{n} a_i^n y_{n-i}, \quad a_0^n = 1,\ n \geq 0. \qquad (3)$$

In (3), a_n^n is the n^{th} order reflection coefficient associated with the ARMA process.

From the definition of the innovation sequence, y_n may be represented as [3],

$$y_n = \sum_{i=0}^{n} W_i^n v_{n-i}, \quad W_0^n = 1, \quad n \geq 0, \qquad (4)$$

CH2673-2/89/0000-2186 $1.00

known as the innovation filter of order n. The set $\{W_i^n,\ 0 \le i \le n\}$ collects the coefficients of the innovation filter of order n. We use the notation of [6], [7] and assume that

$$W_j^n = a_j^n = 0 \text{ for } j < 0. \tag{5}$$

Let $\mathbf{W}_N^{-1}$ be the triangular matrix of order N+1,

$$\mathbf{W}_N^{-1} = \begin{bmatrix} 1 & & & & \\ a_1^1 & 1 & & \mathbf{0} & \\ a_2^2 & a_1^2 & 1 & & \\ \dots & \dots & & \ddots & \\ a_N^N & a_{N-1}^N & \cdots & a_1^N & 1 \end{bmatrix}. \tag{6}$$

where each line collects the increasing order prediction error filter coefficients (up to order N). The inverse of this matrix, $\mathbf{W}_N$, has a dual meaning. Line n corresponds to the coefficients of the innovation filter of order n, $0 \le n \le N$, [3],

$$[\mathbf{W}_N]_n = [W_n^n \dots W_1^n\ 1\ 0 \dots 0] \tag{7}$$

ARMA Estimation

We present briefly two alternative algorithms for ARMA estimation and comment upon its implementation from a finite lenght sample. Unless otherwise stated, we assume exact knowledge of the prediction and innovation filter coefficients. Let $\mathbf{A}_N$ be the $(N+1) \times (N+1)$ Toeplitz matrix,

$$\mathbf{A}_N = \begin{bmatrix} 1 & & & & & & & \\ a_1 & 1 & & & & \mathbf{0} & & \\ a_2 & a_1 & 1 & & & & & \\ \vdots & & & \ddots & & & & \\ a_p & \dots & \dots & a_1 & 1 & & & \\ & a_p & \dots & a_2 & a_1 & 1 & & \\ \mathbf{0} & & \ddots & & & & \ddots & \\ & & & a_p & \dots & \dots & a_1 & 1 \end{bmatrix} \tag{8}$$

and define the matrix β_N as

$$\beta_N \mathbf{W}_N^{-1} = \mathbf{A}_N. \tag{9}$$

Algorithm 1 [6]

Using the Cayley-Hamilton theorem and the theory of the invariant directions of the Riccati equation, it is proved in [8] that β_N has the following structure for $p \ge q$,

$$\beta_N = \begin{bmatrix} 1 & & & & & & \\ \Omega_1(1) & 1 & & & \mathbf{0} & & \\ \Omega_2(2) & \Omega_1(2) & 1 & & & & \\ \vdots & & & & & & \\ & \Omega_q(p) & \dots & \Omega_1(p) & 1 & & \\ \mathbf{0} & \ddots & & & & \ddots & \\ & & \Omega_q(N) & \dots & \Omega_1(N) & 1 \end{bmatrix}. \tag{10}$$

Further, [8],

$$\lim_{N \to \infty} \Omega_i(N) = b_i,\ 1 \le i \le q. \tag{11}$$

The algorithm obtains the MA and the AR components by dual, independent procedures that solve systems of linear equations.

MA Component

For $p + q \le n \le N$, by equating the elements of line n on both sides of (9), corresponding to the first $n - p$ zeros of line n of $\mathbf{A_N}$, yields the system of $n - p$ linear equations,

$$a_j^n + \Omega_1(n) a_{j-1}^{n-1} + \dots + \Omega_q(n) a_{j-q}^{n-q} = 0, \quad p+1 \le j \le n \tag{12}$$

The solution of (12) is the set of coefficients $\{\Omega_i(n), 1 \le i \le q\}$ which converge with n to the MA component (see (11)). Note that n is the highest order reflection coefficient present in (12).

AR Component

For $p \le n \le N$, the a_i coefficients satisfy the system of linear equations,

$$\begin{cases} W_j^p + a_1 W_{j-1}^{p-1} + \dots + a_p W_{j-p}^0 = 0 & q+1 \le j \le p \\ W_j^{p+1} + a_1 W_{j-1}^p + \dots + a_p W_{j-p}^1 = 0 & q+1 \le j \le p+1 \\ \vdots & \\ W_j^N + a_1 W_{j-1}^{N-1} + \dots + a_p W_{j-p}^{N-p} = 0 & q+1 \le j \le N \end{cases} \tag{13}$$

obtained by equating the lines $\{p, p+1, \dots, N\}$ on both sides of

$$\mathbf{A}_N \mathbf{W}_N = \beta_N. \tag{14}$$

At least for $n > p+q$, (12) and (13) are overdimensioned systems of equations. For the full derivation of the algorithm see [8], [9].

Algorithm 2 [1]

The algorithm obtains the AR component by a two stage procedure that solves a system of linear equations. The MA component is obtained through a fast iterative algorithm whose initial condition depends on the previously computed AR component.

AR Component

1. Particularizing (12) for $n = p + q$, obtain $\{\Omega_i(p+q),\ 1 \le i \le q\}$.
2. Compute the p AR coefficients by replacing the Ω_i's in the linear relations

$$a_i = a_i^{p+q} + \sum_{j=1}^{q} \Omega_j(p+q) a_{i-j}^{p+q-j},\ 1 \le i \le p \tag{15}$$

This AR estimator uses only the set of p reflection coeffients $\{a_q^q, a_{q+1}^{q+1}, \ldots, a_{p+q}^{p+q}\}$ and the corresponding prediction error filters. See also [5].

Points 1 and 2 above are derived in [1], [2] by equating the corresponding coefficients of (3) and an alternative representation of the prediction error filter of order n. The filter is written as a linear combination involving the previous observation data, the previous one-step predictors, the AR coefficients, and the coefficients of a transient Kalman gain. Alternatively, (15) clearly results from line $p+q$ of (9).

MA Component

1. Obtain the AR coefficients as above, form

$$\tilde{\mathbf{K}}_0 = [\Omega_p(p) \ldots \Omega_2(2)\ \Omega_1(1)]^T \qquad (16)$$

by solving the system of linear equations

$$a_i^n + \sum_{j=1}^{n} a_{i-j}^{n-j} \Omega_j(n) = a_i,\ 1 \le i \le n,\ 1 \le n \le p$$

and evaluate the initial smoothing vector

$$\mathbf{L}_0 = \tilde{\mathbf{K}}_0 + [-a_p \ldots - a_2\ -a_1]^T .$$

2. Solve sequentially the fast filtering equations,

$$\tilde{\mathbf{K}}_n = \frac{\tilde{\mathbf{K}}_{n-1} - \mathbf{H}\mathbf{L}_{n-1}\mathbf{J}\mathbf{L}_{n-1}}{1 - (\mathbf{H}\mathbf{L}_{n-1})^2} \qquad (17)$$

$$\mathbf{L}_n = \mathbf{J}\mathbf{L}_{n-1} - \mathbf{H}\mathbf{L}_{n-1}\tilde{\mathbf{K}}_n, \qquad (18)$$

with

$$\mathbf{J} = \left[\begin{array}{c|c} \begin{matrix} 0 & & \\ 1 & \ddots & \\ & \ddots & 0 \\ & & 1 \end{matrix} & \begin{matrix} 0 \\ \\ 0 \end{matrix} \end{array}\right], \mathbf{H} = [0 \ldots 0\ 1] \qquad (19)$$

When convergence is attained (i.e., $\mathbf{L}_n \to 0$), the MA coefficients are the last q elements of $\tilde{\mathbf{K}}_n$. The vector $\tilde{\mathbf{K}}_n$ is the time-varying Kalman gain for a particular state space representation of (1), [2]. See also [4]. For full details on the algorithm derivation, see [2].

Algorithms' Implementation

The two algorithms use the Burg technique to estimate the coefficients of the prediction error filters of increasing orders directly from the data. The estimated values of the innovation filters coefficients used in algorithm 1 are obtained through a recursive inversion of $\widehat{\mathbf{W}_N^{-1}}$. This matrix is computed from (6) by replacing $\{a_i^n\}$ by $\{\widehat{a_i^n}\}$.

Algorithms' Comparison

The similarities of the algorithms are their starting point based on the reflection coefficients sequence directly estimated from the data, the use of increasing order prediction error filters and the fact that both MA estimation procedures have an asymptotic character. The main differences between them are the following:

i) algorithm 1 is dual, i.e., the AR and MA components are obtained using the same type of operations;

ii) the AR and MA estimation procedures are decoupled in algorithm 1, whereas in algorithm 2 the MA estimation depends on the AR estimated values;

iii) algorithm 2 does not use the innovation filters' coefficients;

iv) algorithm 1 can use overdimensioned systems;

v) the AR estimation in algorithm 1 may use lower order reflection coefficients than it does in algorithm 2;

vi) in algorithm 1, the assymptotics of the MA procedure are in the order, N, of the highest reflection coefficient used, a_N^N. In algorithm 2, this is in the iterative procedure propagating the Kalman gain. This last method does not use the a_N^N for $N > p+q$;

Simulation Results

We present the results of 3 simulated experiments. Figures 1, 2 and Table 2 show the true pole/zero pattern ($\times$-true poles, $\circ$-true zeros) and the mean estimated pattern (#-estimated poles, $\diamond$-estimated zeros) over 100 Monte-Carlo runs and 512 points on each sample. Diagrams a) in all figures refer to algorithm 2, while the remaining refer to the results of algorithm 1 for successively higher number, N, of reflection coefficients retained. Table 1 shows the experimental parameter values used.

Model	Poles	Zeros	σ^2
ARMA(4,2)	$0.95e^{\pm j36^\circ}$ $0.9\ e^{\pm j54^\circ}$	$0.7e^{\pm j108^\circ}$	1
ARMA(4,4)	$0.95e^{\pm j72^\circ}$ $0.95e^{\pm j107^\circ}$	$0.856e^{\pm j70.4^\circ}$ $0.856e^{\pm j109.6^\circ}$	1.5195
ARMA(6,4)	$0.85e^{\pm 45^\circ}$ $0.9e^{\pm 90^\circ}$ $0.85e^{\pm 135^\circ}$	$0.95\ e^{\pm 67.5^\circ}$ $0.95e^{\pm 112.5^\circ}$	1

Table 1-True pole/zero location

Figure 1 illustrates the assymptotics of algorithm 1 with N, showing the existence of an optimal N. For $N = p+q = 6$, both algorithms use the same number of reflection coefficients, the assymptotics of algorithm 2 leads to better zero estimates (c.f., Fig.1-a) and b)). For $N = 9$, algorithm 1 uses an overdimensioned system to estimate the Ω_i leading to an overall better zero estimate (c.f., Fig. 1-a) and

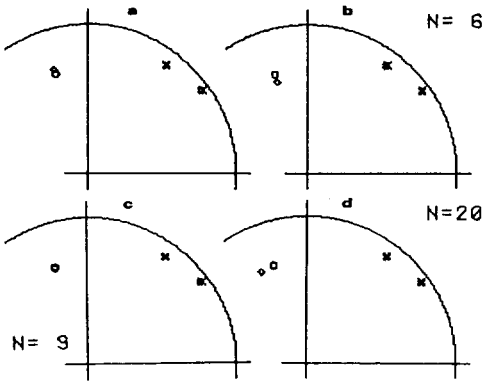

Figure 1: Simulation results for ARMA(4,2)

c)). Finally, for $N = 20$, errors in the estimation of higher order reflection coefficients produce bias and degradation in the estimation of the zeros of algorithm 1 (c.f., Fig. 1-c) and d)). This process is also studied in [5].

In Table 2 we display the mean estimated pole/zero locations for the ARMA(4,4) process.

Algorithm	Estimated Poles	Estimated Zeros
1	$0.972e^{\pm j73.0^\circ}$	$0.857e^{\pm j74^\circ}$
N=8	$0.943e^{\pm j105.1^\circ}$	$0.807e^{\pm j104.2^\circ}$
1	$0.934e^{\pm j70.3^\circ}$	$0.812e^{\pm j66.8^\circ}$
N=9	$0.937e^{\pm j109.3^\circ}$	$0.823e^{\pm j113.4^\circ}$
2	$0.979e^{\pm j73.5^\circ}$	$0.843e^{\pm j71.3^\circ}$
	$0.934e^{\pm j103.8^\circ}$	$0.804e^{\pm j106.5^\circ}$

Table 2-Simulation results for ARMA(4,4)

We conclude that the MA assymptotic procedure of algorithm 2 does not seem to improve on the zero locations when compared with those obtained with algorithm 1 for $N = p + q = 8$ and $N = 9$.

For the ARMA(6,4) process, algorithm 1 can use to its advantage the reflection coefficients at least till $N = 30$. For this value of N, this algorithm performs better than the assymptotics of algorithm 2, see Fig. 2.

Conclusions

Although the paper presented only a preliminary set of comparisons between the two algorithms, it seems plausible to conclude that when algorithm 1 can use a significant degree of overdimension, the resulting statistical stability leads to better MA estimation (c.f., examples 1 and 3). However, when this is not the case, like in example 2, this study is inconclusive.

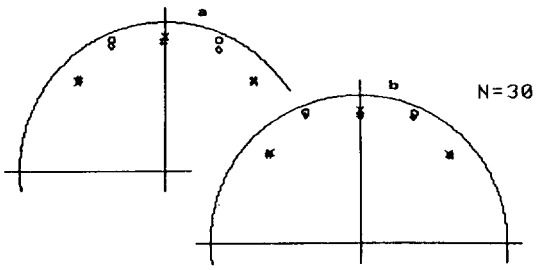

Figure 2: Simulation results for ARMA(6,4)

Besides further more exaustive comparison studies addressing also issues like bias and variance, we will in the future pursue a combined MA procedure where the starting values for algorithm 2 are provided by algorithm 1.

References

[1] G. Alengrin, J. Zerubia, "A Method to Estimate the Parameters of an ARMA Model," IEEE Tr. AC, Vol.32, No.12, Dec. 1987.

[2] G. Alengrin, J. Zerubia, "Estimation of ARMA parameters and Kalman filtering," IFAC Symp. on Identification, Beijing, P.R.China, Aug. 1988.

[3] R.S.Bucy, "Identification and Filtering," Mathematical Systems Theory, Vol.16, pp.307-317, Dec. 1983.

[4] C. J. Demeure, L. L. Scharf, "Linear Statistical Models for Stationary Sequences and Related Algorithms for Cholesky Factorization of Topelitz Matrices," IEEE Tr. on ASSP, Vol.35, no.1, Jan. 1987.

[5] S. Li, B. W. Dickinson, "Application of the Lattice Filter to Robust Estimation of AR and MA Models," IEEE Tr. on ASSP, Vol.36, No.4, Apr. 1988.

[6] J.M.F.Moura, M.I.Ribeiro, "Parametric Spectral Estimation for ARMA Processes," 3rd Workshop on Spectrum Estimation and Modeling, Boston, Nov. 1986.

[7] J.M.F.Moura, M.I.Ribeiro, "Linear Estimation of ARMA Processes," 4th Workshop on Spectrum Estimation and Modeling, Minneapolis, Aug. 1988.

[8] M.I.Ribeiro, J.M.F.Moura, "Dual Estimation of the Poles and Zeros of an ARMA(p,q) Process," Techn. Rep. No.LIDS-P-1521, Massachusets Institute of Technology, 1985.

[9] M.I.Ribeiro, J.M.F.Moura, "ARMA Estimation and Order Selection: a Dual Linear Algorithm," in preparation.