ARMA PARAMETER ESTIMATION: REVISITING A CEPSTRUM-BASED METHOD

Miguel A. Lagunas Hernández[1], *Petre Stoica* *[2], *Miguel A. Rojas* †[1]

[1]Centre Tecnòlogic de Telecomunicacions de Catalunya, Barcelona, Spain
[2]Uppsala University, Dept. of Information Technology, Uppsala, Sweden
Email: m.a.lagunas@cttc.es, ps@it.uu.se, marojas@cttc.es

ABSTRACT

An interesting but apparently forgotten auto-regressive moving-average (ARMA) parameter estimation method, introduced by one of us in 1984, and refined later on by others, is revisited. In the process, we provide a new simpler derivation of the method as well as an enhanced version of its cepstrum-based step. We argue that this method has an appealing advantage over Durbin's method, which is probably the most frequently used non-iterative method for ARMA parameter estimation.

Index Terms— ARMA, cepstrum, Durbin, LAST

1. INTRODUCTION

Estimating the parameters of auto-regressive moving-average (ARMA) sequences is a fundamental problem in signal processing, econometrics and statistics, particularly in spectral estimation and in time series analysis and forecasting [1, 2, 3, 4]. Of the many methods proposed for solving this estimation problem in the last sixty years or so, only a few ones have survived the "test of time". Durbin's method (DU) introduced in 1959 (see [5]), is one of these successful methods: DU has the appeal of conceptual and computational simplicity as well as of reasonable statistical performance, and it has been included in the most popular scientific software packages, such as Matlab (see, e.g. [6]).

Another method for ARMA parameter estimation was introduced in 1984 by one of us [7] and subsequently refined two years later in [8]. The derivation in [7][8] was based on the maximum entropy principle, which might not appeal to every potential user. Here we present a new derivation of LAST that is based only on simple basic facts. The original method in [7, 8] used a cepstrum estimate [9] that does not appear to have good statistical properties in small-or medium-size samples. Here we make use of an enhanced cepstrum estimation method that works well not only in large samples but also in samples of practical lengths. Finally the advantage of the method over other ARMA parameter estimation methods, such as DU, was not made clear in [7, 8]. Here we provide evidence that LAST does not suffer from the problem that affects DU.

*Work supported in part by the Swedish Research Council
†Work supported by the Mexican Government under grant PROMEP

2. PROBLEM FORMULATION AND PRELIMINARIES

Consider a univariate real-valued ARMA sequence, $\{y(t)\}_{t=1,2,\ldots}$, that satisfies the equation:

$$A\left(z^{-1}\right) y(t) = B\left(z^{-1}\right) e(t) \tag{1}$$

where z^{-1} denotes the unit delay operator, $A\left(z^{-1}\right)$ as well as $B\left(z^{-1}\right)$ are monic polynomials in z^{-1}, $A\left(z^{-1}\right) = 1 + a_1 z^{-1} + ... + a_n z^{-n}$, $B\left(z^{-1}\right) = 1 + b_1 z^{-1} + ... + b_n z^{-n}$ and $\{e(t)\}$ is a white noise sequence with mean zero and variance denoted by σ^2. The polynomials $A\left(z^{-1}\right)$ and $B\left(z^{-1}\right)$ are assumed to be co-prime and have the same degree n. We further assume that all zeros of $A\left(z^{-1}\right)$ lie strictly inside the unit circle, so that $\{y(t)\}$ is a stationary sequence. The latter assumption allows us to define the following quantities of which we will make frequent use later on:
ARMA's impulse response

$$H\left(z^{-1}\right) = \sum_{k=0}^{\infty} h_k z^{-k} = \frac{B\left(z^{-1}\right)}{A\left(z^{-1}\right)}; h_0 = 1 \tag{2}$$

ARMA's spectral density or spectrum

$$\phi(z) = \sigma^2 H(z) H\left(z^{-1}\right) \tag{3}$$

ARMA's covariance matrix of size $n+1$

$$\mathbf{R} = E\left(\begin{bmatrix} y(t) \\ . \\ y(t-n) \end{bmatrix} \begin{bmatrix} y(t) & ... & y(t-n) \end{bmatrix} \right) \tag{4}$$

ARMA's covariance sequence

$$r(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi\left(e^{jw}\right) e^{jkw} dw \tag{5}$$

Under the additional assumption that $B\left(z^{-1}\right)$ has no zeros on the unit circle (so that $\phi\left(e^{jw}\right) > 0$ strictly for $w \in [-\pi, \pi]$

we can also define the cepstral coefficients $\{c_k\}$: ARMA's cepstral sequence

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} Ln\left[\phi\left(e^{jw}\right)\right] e^{jkw} dw \tag{6}$$

see, e.g., [8, 10, 11, 12]. Note from (3) and (6) that

$$\sum_{k=-\infty}^{\infty} c_k z^{-k} = Ln(\sigma^2) + Ln(H(z)) + Ln(H(z^{-1})) \tag{7}$$

which implies that

$$\frac{c_0}{2} + \sum_{k=1}^{\infty} c_k z^k = \frac{Ln(\sigma^2)}{2} + Ln(H(z)) \tag{8}$$

Formal differentiation of (8) with respect to z yields the equation:

$$\sum_{k=1}^{\infty} k c_k z^{k-1} = \frac{\sum_{k=1}^{\infty} k h_k z^{k-1}}{\sum_{k=0}^{\infty} h_k z^k} \tag{9}$$

or, equivalently (using $h_k = 0$ for $k < 0$)

$$\sum_{k=1}^{\infty} k h_k z^{k-1} = \sum_{k=1}^{\infty}\left(\sum_{i=1}^{\infty} i c_i h_{k-i}\right) z^{k-1} \tag{10}$$

From (10) we obtain the well-known equation that relates the impulse response $\{h_k\}$ and the cepstral coefficients $\{c_k\}$ ([7, 8, 11])

$$h_k = \frac{1}{k}\sum_{i=1}^{k} i c_i h_{k-i}; k = 1, 2, ... \tag{11}$$

Following the previous preparations, we can state the problem of interest: given a sequence of observations $\{y(t)\}_{t=1}^{N}$, estimate the ARMA parameters σ^2, $\{a_k\}$ and $\{b_k\}$. We assume that the ARMA order n is given[13, 14]. There are numerous solutions to the ARMA parameter estimation problem, that have been proposed in the last 60 years or so, see [1, 2, 3, 5] for some examples. But probably the simplest method proposed for ARMA parameter estimation is the DU [3, 5]. This method consists of two main steps:

a. Fitting a long AR model to $\{y(t)\}_{t=1}^{N}$, and computing the residual sequence of this model, let us say $\{\hat{e}(t)\}$.

b. Estimating σ^2 as the sample variance of $\{\hat{e}(t)\}$, and a_k,b_k from (1) with $\{e(t)\}$ replaced by $\{\hat{e}(t)\}$.

A problem of DU is that the estimation of the order of the long AR model, let us say m, in step (a) is not a simple task. This is specially true for ARMA sequences for which the polynomial $B(z^{-1})$ has zeros close to the unit circle, in which case a rather large m might be needed for a reasonable fit in step (a). A large m value will lead to a shorter estimated white noise sequence $\{\hat{e}(t)\}$ and therefore to a potentially worsened accuracy of the estimates computed in step (b). Presumably to avoid this type of problem, some implementations of DU simply use a fixed value of m, such m equal to four times n (see e.g. [6]).

3. LAST

In this section we will rely on ideas in [7, 8] to derive LAST, an ARMA parameter estimation method that, while being as simple conceptually and computationally as DU, does not suffer from the above mentioned problem of the latter.

Let $\mathbf{a} = \begin{bmatrix} 1 & a_1 & ... & a_n \end{bmatrix}^T$, $\mathbf{b} = \begin{bmatrix} 1 & b_1 & ... & b_n \end{bmatrix}^T$ and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & .. & 0 \\ h_1 & 1 & .. & 0 \\ .. & .. & .. & .. \\ h_n & .. & h_1 & 1 \end{bmatrix} \tag{12}$$

A simple calculation shows that

$$E\left(\begin{bmatrix} e(t) \\ .. \\ e(t-n) \end{bmatrix}\begin{bmatrix} y(t) & .. & y(t-n) \end{bmatrix}\right) = \sigma^2\mathbf{H} \tag{13}$$

and that (using, once again, $h_k = 0$ for $k < 0$)

$$B(z) = H(z)A(z) = \sum_{i=0}^{\infty}\sum_{p=0}^{n}\left(h_{i-p}a_p\right) z^i \tag{14}$$

which implies that

$$\mathbf{b} = \mathbf{H}\mathbf{a} \tag{15}$$

From (13) and (15) along with (1) we obtain:

$$\begin{aligned} 0 &= E\left[A\left(z^{-1}\right)y(t) - B\left(z^{-1}\right)e(t)\right]^2 \\ &= \begin{bmatrix} \mathbf{a}^T & -\mathbf{b}^T \end{bmatrix}\begin{bmatrix} \mathbf{R} & \sigma^2\mathbf{H}^T \\ \sigma^2\mathbf{H} & \sigma^2\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ -\mathbf{b} \end{bmatrix} \\ &= \mathbf{a}^T\left(\mathbf{R} - \sigma^2\mathbf{H}^T\mathbf{H}\right)\mathbf{a} \end{aligned} \tag{16}$$

Because the matrix $\begin{bmatrix} \mathbf{R} & \sigma^2\mathbf{H}^T \\ \sigma^2\mathbf{H} & \sigma^2\mathbf{I} \end{bmatrix}$ is positive semi-definite, we must have (below $\mathbf{R}^{-1/2}$is a symmetric square root of $\mathbf{R}^{-1}$):

$$\begin{aligned} &\mathbf{R} - \sigma^2\mathbf{H}^T\mathbf{H} \geq \mathbf{0} \Leftrightarrow \mathbf{I} - \sigma^2\mathbf{R}^{-1/2}\mathbf{H}^T\mathbf{H}\mathbf{R}^{-1/2} \geq 0 \\ &\Leftrightarrow \sigma^2 \leq \frac{1}{\lambda_{\max\left(\mathbf{R}^{-1/2}\mathbf{H}^T\mathbf{H}\mathbf{R}^{-1/2}\right)}} = \frac{1}{\lambda_{\max\left(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T\right)}} \end{aligned} \tag{17}$$

where the symbol λ_{max} stands for the maximum eigenvalue. Furthermore, the matrix in (16), $\mathbf{R} - \sigma^2\mathbf{H}^T\mathbf{H}$, must evidently be singular. From this observation and from (17) we obtain the following expression for σ^2:

$$\sigma^2 = \frac{1}{\lambda_{\max\left(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^T\right)}} \tag{18}$$

Making use of (16), once more, it follows that:

$\mathbf{a}$ = the eigenvector of the matrix $\mathbf{R} - \sigma^2 \mathbf{H}^T \mathbf{H}$ associated with its minimum eigenvalue, with the first element normalized to one. (19)

The derived expressions for σ^2, $\mathbf{a}$ and $\mathbf{b}$ (see (18),(19) and (15)) as functions of $\mathbf{R}$ and $\mathbf{H}$, can be used to estimate these parameters as explained in the sequel.
First we estimate $\mathbf{R}$ by means of the forward-backward averaging method (see, e.g. , [3][15]). Specifically, let

$$\psi = \begin{bmatrix} 0 & .. & 1 \\ .. & .. & .. \\ 1 & .. & 0 \end{bmatrix} \tag{20}$$

denote the reversal matrix of appropriate dimensions, and let $\tilde{\mathbf{R}}$ be the following sample covariance matrix:

$$\tilde{\mathbf{R}} = \frac{1}{N-n} \sum_{t=n+1}^{N} \begin{bmatrix} y(t) \\ .. \\ y(t-n) \end{bmatrix} \begin{bmatrix} y(t) & ... & y(t-n) \end{bmatrix} \tag{21}$$

Then we use

$$\hat{\mathbf{R}} = \left(\tilde{\mathbf{R}} + \psi \tilde{\mathbf{R}} \psi \right) / 2 \tag{22}$$

as an estimate of $\mathbf{R}$. Note that (22) is centro-symmetric but not Toeplitz, unlike $\mathbf{R}$; yet there is empirical evidence that the estimate of $\mathbf{R}$ in (22) has better accuracy in small or medium sized samples than Toeplitz estimates, $[\hat{r}(k-p)]_{k,p=1}^{n+1}$, of $\mathbf{R}$.
Next, we estimate $\mathbf{H}$ by exploiting the relationship between $\{h_k\}$ and the cepstral coefficients $\{c_k\}$, see (11). To do so, we need to estimate $\{c_k\}$ first in the following way (see the definition of $\{c_k\}$ in (6) for motivation). Let $\hat{\phi}(e^{jw})$ denote an estimate of the ARMA spectrum, $\phi(e^{jw})$, and let IFFT denote the inverse fast Fourier transform. Then:

$$\{\hat{c}_k\} = IFFT \left\{ \left[Ln \left(\hat{\phi} \left(e^{jw_p} \right) \right) \right] \right\} \tag{23}$$

where $w_p = \frac{2\pi}{N}$ for $p = 1, N$. Quite often, $\hat{\phi}(e^{jw})$ in (23) is obtained using the periodogram method (see, e.g., [1, 2, 3, 12]). However, we have observed empirically that the cepstral coefficient estimates obtained using (23) and the periodogram method, and therefore the corresponding LAST estimates of the ARMA parameters, in general have satisfactory accuracy only in relatively large samples. To obtain more accurate estimates of these parameters for practical sample lengths, we recommend the use of an estimate $Ln\left(\phi\left(e^{jw}\right)\right)$ suggested in [16], in lieu of $Ln\left(\hat{\phi}\left(e^{jw}\right)\right)$, namely:

$$Ln\left[\hat{\phi}\left(e^{jw}\right)\right] = \frac{1}{M} \begin{bmatrix} 1 & .. & e^{jMw} \end{bmatrix} Ln\left[\hat{\Gamma}\right] \begin{bmatrix} 1 \\ e^{-jw} \\ .. \\ e^{-jMw} \end{bmatrix} \tag{24}$$

where $\hat{\Gamma}$ is a sample covariance matrix defined similarly to $\hat{R}$ in (22), but of dimensions $(M+1, M+1)$. In (24), $Ln\left[\hat{\Gamma}\right]$ denotes the matrix logarithm of $\hat{\Gamma}$, which is computed as:

$$Ln\left[\hat{\Gamma}\right] = \mathbf{U} \begin{bmatrix} Ln(\lambda_1) & .. & 0 \\ .. & .. & .. \\ 0 & .. & Ln(\lambda_{M+1}) \end{bmatrix} \mathbf{U}^H \tag{25}$$

where $\mathbf{U}diag(\lambda_1 ... \lambda_{M+1})\mathbf{U}^H$ is the eigen-decomposition of $\hat{\Gamma}$. To guarantee that $\lambda_k > 0$ (for $k = 1, M+1$), the value of M in the definition of $\hat{\Gamma}$ should not be chosen larger than $2N/3$. Our experience suggests that, in general, $M = N/2$ is a good choice. The proposed ARMA parameter estimation method is summarized next. Note that from now on this technique is called LAST and the previous one that uses eq.(23) is refered to as old LAST.

LAST

Step 1. Compute cepstral coefficient estimates $\{\hat{c}_k\}$ using (23) and (24). Insert $\{\hat{c}_k\}$ in (11) to obtain estimates $\left\{\hat{h}_k\right\}_{k=1}^{n}$ of $\{h_k\}_{k=1}^{n}$. Use $\left\{\hat{h}_k\right\}_{k=1}^{n}$ to build an estimate $\hat{\mathbf{H}}$ of $\mathbf{H}$.

Step 2. Use $\hat{\mathbf{H}}$ from Step 1 and $\hat{\mathbf{R}}$ from (22) in (18) to compute an estimate $\hat{\sigma}^2$ of σ^2. Obtain an estimate $\hat{\mathbf{a}}$ of $\mathbf{a}$ as the normalized minimum eigenvector of $\hat{\mathbf{R}} - \hat{\sigma}^2 \hat{\mathbf{H}}^{\mathbf{T}} \hat{\mathbf{H}}$(see (19)). Finally, estimate $\mathbf{b}$ as $\hat{\mathbf{b}} = \hat{\mathbf{H}}\hat{\mathbf{a}}$

Observe that LAST does not require the solution of an additional order estimation problem (besides the ARMA order estimation one), and therefore that the proposed method does not suffer from what appears to be an inherent drawback of DU. In particular, unlike the DU estimates, the LAST estimates of the ARMA parameters can be shown to be consistent under weak conditions (recall that the consistency of the DU estimate depends on the selection of the long AR model order; on the other hand the consistency of the LAST estimates, for instance the ones based on (23) and the periodogram method, follows easily from the consistency of $\hat{\mathbf{R}}$ and of $\{\hat{c}_k\}$).

4. NUMERICAL EXAMPLE AND CONCLUDING REMARKS

The simulations that have been carried out showed that the performance of LAST [17] can be much better than that of the old version of LAST and of a version of the method of [18]. For the sake of conciseness, in this paper we compare LAST only with the Durbin's method (DU), which is probably the most frequently used non-iterative method for ARMA parameter estimation. Fig. 1 shows that the Durbin's spectral estimates have relatively poor performance when spectral nulls have to be detected (the simulated ARMA is given by $a_1 = 1.0$, $a_2 = -1.5291$, $a_3 = 1.4512$, $a_4 = -0.7280$, $a_5 = 0.2267$, $b_1 = 1.0$, $b_2 = 1.8794$, $b_3 = 2.5321$, $b_4 = 1.8794$, $b_5 = 1.0$) this is due to the long predictor step which requires

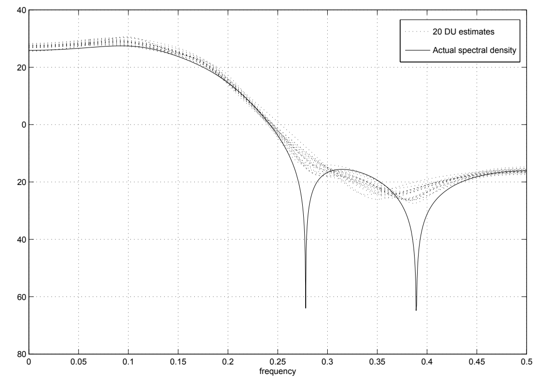

Fig. 1. Actual spectral density and 20 superimposed DU estimates

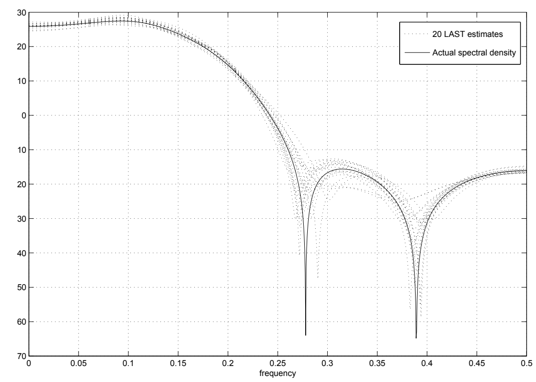

Fig. 2. Actual spectral density and 20 superimposed LAST estimates

very large orders to cope with these low energy bands. In contrast, LAST shows better performance in this region due to the sensitivity of the new cepstral estimate to low energy frequency bands see in Fig. 2. In results not shown here, we have observed that for close pole-zero pairs, LAST degrades more gracefully than DU. In the extreme case of pure lines in white noise LAST was much better than DU for any data record length. In all cases, LAST exhibits lower bias than DU and for large samples it has similar variance to DU. In summary, LAST appears to be useful technique for ARMA spectral estimation with similar performance to Durbin's method in cases with no spectral notches, but with hardly any competitor when deep spectral nulls have to be detected and located.

5. REFERENCES

[1] P.J. Brockwell and R.A. Davis, "Time series: Theory and methods," *Springer Series in Statistics*, 1987.

[2] M.B. Priestley, "Spectral analysis and time series," *Academic Press*, 1981.

[3] P. Stoica and R. Moses, "Spectral analysis of signals," *Englewood Cliffs, NJ. Prentice Hall*, 2005.

[4] C.I. Byrnes, P. Enqvist, and A. Lindquist, "Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach," *SIAM J. Control and Optimization*, vol. 41, pp. 23–59, Jan 2002.

[5] J. Durbin, "Efficient estimation of parameters in moving average models," *Biometrika 46*, 1959.

[6] Ljung, "System identification toolbox," *Mathworks*.

[7] M.A. Lagunas et al., "Arma model maximum entropy power spectral estimation," *IEEE Trans. On Acoustics Speech and Signal Processing*, vol. ASSP-32, No. 5, pp. 984–990, October 1984.

[8] B.R. Musicus and A.M. Kabel, "Maximum entropy pole-zero estimation," *ICASSP-86,paper 27.12, Tokyo*, 1986.

[9] C.I. Byrnes, P. Enqvist, and A. Lindquist, "Cepstral coefficients covariance lags and pole-zero models for finite data strings," *IEEE Trans. Signal Processing SP-50*, vol. 41, pp. 677–693, April 2001.

[10] B. Boguert, M.J. Healey, and J.W. Tukey, "The frecuency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and shape cracking," *Ed. New York:Wiley*, 1963.

[11] A. Oppenheim and R.W. Schafer, "Discrete-time signal processing," *N.J. Prentice Hall*, 1989.

[12] P. Stoica and N. Sandgren, "Smoothed nonparametric spectral estimation via cepstrum thresholding," *IEEE Signal Processing Magazine*, pp. 34–45, November 2006.

[13] P. Stoica and Y. Selen, "Model-order selection: A review of information criteria rules," *IEEE Signal Processing Magazine*, pp. 36–47, July 2004.

[14] J.J. Fuchs, "Arma order estimation via matrix perturbation theory," *IEEE Trans. On Automatic Control*, vol. AC-32, pp. 358–361, April 1987.

[15] M. Jansson and P. Stoica, "Forward-only and forward-backward sample covariance. a comparative study," *Signal Processing*, vol. 77, pp. 235–245, 1999.

[16] V.F. Pisarenko, "On the estimation of spectra by means of non-linear functions of the variance matrix," *Geophys. J. Roy. Astronom. Soc*, vol. 28, pp. 511–531, 1972.

[17] Miguel A. Lagunas and P. Stoica, "Matlab code for last," *http://www.mathworks.com/matlabcentral/fileexchange/*.

[18] A. Kizilkaya and A.H. Kayran, "Arma model parameter estimation based on the equivalent ma approach," *Digital Signal Processing, Elseiver*, vol. 16, pp. 679–681, 2006.