A NEW ESTIMATION APPROACH FOR AR MODELS IN PRESENCE OF NOISE

Roberto Diversi * **Umberto Soverini** * **Roberto Guidorzi** *

* *Dipartimento di Elettronica, Informatica e Sistemistica*
Università di Bologna
Viale del Risorgimento 2, 40136, Bologna, Italy
e–mail: {*rdiversi, usoverini, rguidorzi*}*@deis.unibo.it*

Abstract: This paper considers the problem of estimating the parameters of an autoregressive (AR) process in presence of additive white noise and proposes a new identification method, based on theoretical results originally developed in errors–in–variables contexts. This approach allows to estimate the AR parameters, the driving noise variance and the variance of the additive noise in a congruent way in that these estimates assure the positive definiteness of the autocorrelation matrix. The performance of the proposed algorithm is compared with that of bias–compensated least–squares methods by means fo Monte Carlo simulations. The results show the effectivenesss of the new method also in presence of high amounts of noise.

Keywords: System identification, parameter estimation, autoregressive models.

1. INTRODUCTION

Autoregressive (AR) models are commonly used in a wide range of signal processing applications, like spectrum estimation, speech analysis, noise cancellation and digital comunications (Haykin, 1991). A considerable attention has been dedicated, in the literature, to the problem of estimating the AR parameters from signals corrupted by white noise. This is, in fact, a very common situation. A correct description of AR plus noise models requires also the introduction of zeros so that the estimates obtained with classical AR identification methods are poor, particularly for low signal–to–noise ratio conditions (Kay, 1980; Paliwal, 1986).

Since noisy AR processes can be described by means of ARMA models, the usual approach for solving this problem is to use standard ARMA identification techniques for recovering their autoregressive part (Pagano, 1974). This can be done, for instance, by means of modified Yule–Walker equations (Kay, 1988), the maximum–likelihood method (Tong, 1975) and the prediction error method (Nehorai and Stoica, 1988).

Another class of methods is based on the bias–compensation technique. In this case the noise variance is assumed as known or is estimated. This information is then used to correct the estimates given by an AR identification method. In many speech enhancement applications, for example, the white noise variance can be estimated from preceding silent portions of speech (silent frames), when present. In many other contexts, however, this procedure cannot be applied and the noise variance estimation constitutes an essential part of the identification problem.

In the last years a large variety of bias–compensated least–squares (BCLS) techniques have been proposed (Sakai and Arase, 1979; Zheng, 1999; Jin *et al.*, 2002). These methods are based on iterative procedures where, at each step, the current estimate of the noise variance is used to improve the estimate of the AR parameters and *vice versa*. A comparative analysis of various BCLS algorithms is reported in (Jia *et al.*, 2003).

This paper introduces a novel approach based on the theoretical results originally developed in errors–in–variables contexts in (Beghelli *et al.*, 1990). This approach relies, in particular, on the properties of the family of solutions of the dynamic Frisch scheme and on the shift property of time–invariant dynamic systems. The method allows to estimate the AR parameters, the driving noise variance and the variance of the additive noise in a congruent way since these estimates assure the positive definiteness of the autocorrelation matrix.

The effectiveness of the proposed algorithm has been tested by means of Monte Carlo simulations. The results show that this approach yields better estimates than those obtained with BCLS methods, especially in presence of low signal–to–noise ratios.

The contents are organized as follows. Section 2 defines the noisy AR identification problem. Section 3 concerns some asymptotic properties of noisy AR models that are at the basis of the new identification method presented in Section 4. In Section 5 the performance of the proposed algorithm is compared with that of BCLS methods by means of Monte Carlo simulations. Some conclusions are finally reported in Section 6.

2. STATEMENT OF THE PROBLEM

Consider a noisy autoregressive model of order n described by the equations

$$x(t) = \alpha_1\, x(t-1) + \cdots + \alpha_n\, x(t-n) + e(t), \quad (1)$$
$$y(t) = x(t) + w(t), \quad (2)$$

where $x(t)$ is the output of the noise–free AR model, driven by the input $e(t)$ while $y(t)$ is the available observation, affected by the noise process $w(t)$. The following assumptions will be assumed in the sequel.

A1. $e(t)$ and $w(t)$ are zero–mean white processes, mutually uncorrelated, with unknown variances σ_e^{2*} and σ_w^{2*} respectively.
A2. $e(t)$, $x(t)$ and $w(t)$ are ergodic processes.
A3. The system order, n, is known.

By defining the vectors

$$\varphi_x(t) = [\, x(t-n)\ \ldots\ x(t-1)\ x(t)\,]^T, \quad (3)$$
$$\varphi_y(t) = [\, y(t-n)\ \ldots\ y(t-1)\ y(t)\,]^T, \quad (4)$$
$$\varphi_w(t) = [\, w(t-n)\ \ldots\ w(t-1)\ w(t)\,]^T, \quad (5)$$

and the parameter vector

$$\theta^* = \left[\, \alpha_n\ \cdots\ \alpha_1\ -1 \right]^T, \quad (6)$$

it is possible to represent model (1)–(2) in the form

$$\left(\varphi_x^T(t) - [\,0\ \ldots\ 0\ e(t)\,]\right)\, \theta^* = 0, \quad (7)$$

$$\varphi_y(t) = \varphi_x(t) + \varphi_w(t), \quad (8)$$

that will be used in the subsequent analysis.

The problem that will be considered is the following.

Problem 1. Estimate the AR parameters $\alpha_1, \ldots, \alpha_n$ and the variances σ_e^{2*}, σ_w^{2*} starting from the available measurements $y(1), y(2), \ldots, y(N)$, generated by model (1)–(2) under assumptions A1–A3.

3. ASYMPTOTIC PROPERTIES OF NOISY AR MODELS

Define the following covariance matrices

$$\Sigma_n = E\,[\,\varphi_y(t)\,\varphi_y^T(t)\,], \quad (9)$$
$$\hat{\Sigma}_n = E\,[\,\varphi_x(t)\,\varphi_x^T(t)\,] - \mathrm{diag}\,[\underbrace{0\ \ldots\ 0}_{n}\ \ \sigma_e^{2*}\,]. \quad (10)$$

From (7), (8) and assumption A1 it follows that

$$\hat{\Sigma}_n\, \theta^* = 0, \quad (11)$$
$$\Sigma_n = E\,[\,\varphi_x(t)\,\varphi_x^T(t)\,] + E\,[\,\varphi_w(t)\,\varphi_w^T(t)\,], \quad (12)$$
$$E\,[\,\varphi_w(t)\,\varphi_w^T(t)\,] = \sigma_w^{2*}\, I_{n+1}, \quad (13)$$

where $E[\cdot]$ denotes mathematical expectation. In particular, relation (11) can be obtained premultiplying (7) by $\varphi_x(t)$ and applying the operator $E[\cdot]$, taking into account that $E[x(t)\,e(t)] = E[e^2(t)] = \sigma_e^{2*}$. By combining (10) and (12) it is finally possible to write

$$\Sigma_n = \hat{\Sigma}_n + \tilde{\Sigma}_n^*, \quad (14)$$

where

$$\tilde{\Sigma}_n^* = \begin{bmatrix} \sigma_w^{2*} & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_w^{2*} & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \sigma_w^{2*} & 0 \\ 0 & \cdots & \cdots & 0 & (\sigma_w^{2*} + \sigma_e^{2*}) \end{bmatrix} \quad (15)$$
$$= \mathrm{diag}\,[\sigma_w^{2*}\, I_n, \sigma_s^{2*}],$$

with $\sigma_s^{2*} = \sigma_w^{2*} + \sigma_e^{2*}$.

Consider now the family of non-negative definite diagonal matrices $\tilde{\Sigma}_n = \mathrm{diag}\,[\sigma_w^2\, I_n, \sigma_s^2]$ such that

$$\Sigma_n - \tilde{\Sigma}_n \geq 0. \quad (16)$$

This set can be described in a way similar to that reported in (Beghelli *et al.*, 1990) with reference to errors–in–variables (EIV) models. The following theorem can be easily derived by considering noisy AR instead than EIV models.

Theorem 1. The set of all matrices $\tilde{\Sigma}_n$ satisfying relation (16) defines the points $P = (\sigma_s^2, \sigma_w^2)$ of a convex curve $\mathcal{S}(\Sigma_n)$ belonging to the first quadrant of the noise plane $\mathcal{R}^2$ and whose concavity faces the origin. Every point $P = (\sigma_s^2, \sigma_w^2)$ of $\mathcal{S}(\Sigma_n)$ satisfies the relation

$$\hat{\Sigma}_n(P) = \Sigma_n - \text{diag}\;[\sigma_w^2\, I_n, \sigma_s^2] \geq 0 \tag{17}$$

and can be associated with a coefficients vector $\theta(P)$ satisfying the relation

$$\hat{\Sigma}_n(P)\,\theta(P) = 0. \tag{18}$$

Figure 1 shows a typical shape of $\mathcal{S}(\Sigma_n)$. Note that the points (σ_s^2, σ_w^2) of the curve with $\sigma_s^2 \leq \sigma_w^2$ (dotted line) are non admissible because they do not satisfy the condition $\sigma_e^2 = \sigma_s^2 - \sigma_w^2 > 0$. The set of admissible solutions (continuous line) is thus delimited by the straight line $\sigma_w^2 = \sigma_s^2$.

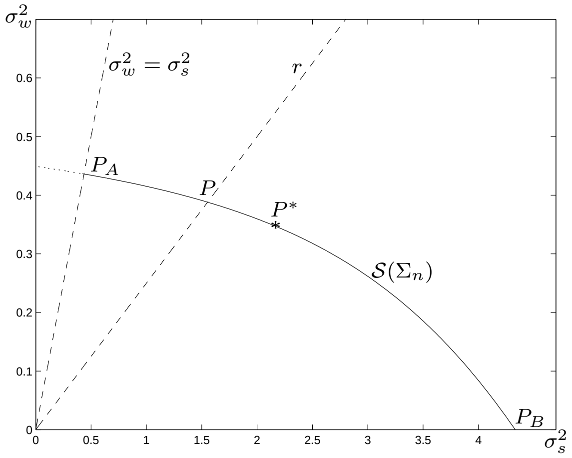

Fig. 1. Typical shape of $\mathcal{S}(\Sigma_n)$. All admissible solutions lie on the continuos line.

Remark 1. The intersection of $\mathcal{S}(\Sigma_n)$ with the σ_s^2 axis is the point $P_B = (\sigma_{s\,\max}^2, 0)$ given by the least squares solution

$$\sigma_{s\,\max}^2 = \frac{\det\,(\Sigma_n)}{\det\,(\Sigma_n')}, \tag{19}$$

where Σ_n' is obtained from Σ_n by deleting its $(n+1)$–th row and column. The intersection of $\mathcal{S}(\Sigma_n)$ with the straight line $\sigma_w^2 = \sigma_s^2$ is the point $P_A = (\sigma_{w\max}^2, \sigma_{w\max}^2)$, given by the eigenvector solution

$$\sigma_{w\max}^2 = \min\;\text{eig}\,(\Sigma_n). \tag{20}$$

Since point P_A corresponds to $\sigma_e^2 = 0$ it is not a solution for Problem 1.

Because of (11) and (14), the point $P^* = (\sigma_s^{2*}, \sigma_w^{2*})$ associated with the actual noise variances belongs to $\mathcal{S}(\Sigma_n)$ and the AR model associated with P^* is characterized by the true coefficients, i.e. $\theta(P^*) = \theta^*$. In this asymptotic context, the determinaton of P^* on $\mathcal{S}(\Sigma_n)$ leads to the solution of Problem 1.

The locus $\mathcal{S}(\Sigma_n)$ of solutions defined by (16) can be parameterized on the basis of the following result (Guidorzi and Pierantoni, 1995).

Theorem 2. Let $\xi = (\xi_1, \xi_2)$ be a generic point of the first quadrant of $\mathcal{R}^2$ and r the straight line from the origin through ξ (see Fig. 1). Its intersection with $\mathcal{S}(\Sigma_n)$ is the point $P = (\sigma_s^2, \sigma_w^2)$ defined by

$$\sigma_s^2 = \frac{\xi_1}{\lambda_M}, \qquad \sigma_w^2 = \frac{\xi_2}{\lambda_M}, \tag{21}$$

where

$$\lambda_M = \max \text{eig}\Big(\Sigma_n^{-1}\,\text{diag}\big[\xi_2\, I_n, \xi_1\big]\Big)\;. \tag{22}$$

This theorem allows to associate a solution with every straight line from the origin lying in the first quadrant.

Remark 2. In some applications the ratio $\eta = \sigma_e^{2*}/\sigma_w^{2*}$ is *a priori* known (Zheng, 2001). In this case the point P^* can be easily determined by means of Theorem 2 taking $\xi = (1+\eta, 1)$. In fact

$$\begin{aligned}\tilde{\Sigma}_n^* &= \text{diag}\;[\sigma_w^{2*}\, I_n, \sigma_w^{2*} + \sigma_e^{2*}]\\ &= \sigma_w^{2*}\;\text{diag}\;[I_n, 1+\eta],\end{aligned} \tag{23}$$

so that P^* is the intersection between $\mathcal{S}(\Sigma_n)$ and the straight line from the origin with slope $1/(1+\eta)$.

4. AR IDENTIFICATION

As pointed out in Section 3, the solution of Problem 1 requires the determination of the point P^* on $\mathcal{S}(\Sigma_n)$. Define, for this purpose, the vectors

$$\bar{\varphi}_x(t) = [\,\varphi_x^T(t)\;\; x(t+1)\,]^T, \tag{24}$$

$$\bar{\varphi}_y(t) = [\,\varphi_y^T(t)\;\; y(t+1)\,]^T, \tag{25}$$

$$\bar{\varphi}_w(t) = [\,\varphi_w^T(t)\;\; w(t+1)\,]^T. \tag{26}$$

From (1) it follows that

$$\left(\bar{\varphi}_x^T(t) - [0\;\ldots\;0\;\, e(t+1)]\right)\,\bar{\theta}^* = 0, \tag{27}$$

where

$$\bar{\theta}^* = \big[\,0\,\alpha_n\;\cdots\;\alpha_1\;\,-1\,\big]^T = \big[\,0\;\;\theta^{*T}\,\big]^T. \tag{28}$$

Define also the covariance matrices

$$\Sigma_{n+1} = E[\,\bar{\varphi}_y(t)\,\bar{\varphi}_y^T(t)\,], \tag{29}$$

$$\hat{\Sigma}_{n+1} = E[\,\bar{\varphi}_x(t)\,\bar{\varphi}_x^T(t)\,] - \text{diag}\;[\underbrace{0\;\ldots\;0}_{n+1}\;\;\sigma_e^{2*}\,]. \tag{30}$$

Since $\bar{\varphi}_y(t) = \bar{\varphi}_x(t) + \bar{\varphi}_w(t)$, from (27) and Assumption A1 it is easy to obtain

$$\hat{\Sigma}_{n+1}\,\bar{\theta}^* = 0, \tag{31}$$

$$\Sigma_{n+1} = \hat{\Sigma}_{n+1} + \tilde{\Sigma}_{n+1}^*, \tag{32}$$

with $\tilde{\Sigma}_{n+1}^* = \text{diag}\;[\sigma_w^{2*}\, I_{n+1}, \sigma_s^{2*}]$.

Making reference to Σ_{n+1} we can now introduce the curve $\mathcal{S}(\Sigma_{n+1})$, belonging to the first quadrant of the noise plane $P = (\sigma_s^2, \sigma_w^2)$, whose shape is similar to that of $\mathcal{S}(\Sigma_n)$. Every point $P = (\sigma_s^2, \sigma_w^2)$ of this curve satisfies the condition

$$\hat{\Sigma}_{n+1}(P) = \Sigma_n - \text{diag}\,[\sigma_w^2\, I_{n+1}, \sigma_s^2] \geq 0. \quad (33)$$

It is also possible to prove (Beghelli *et al.*, 1990) that $\mathcal{S}(\Sigma_{n+1})$ lies under $\mathcal{S}(\Sigma_n)$. It can be easily verified that, because of (31) and (32), P^* belongs to both $\mathcal{S}(\Sigma_n)$ and $\mathcal{S}(\Sigma_{n+1})$.

In this asymptotic context, the determination of the common point P^* leads to the solution of Problem 1. However, when the length N of the sequences is finite, Σ_n and Σ_{n+1} must be replaced by the sample quantities Σ_n^N, Σ_{n+1}^N. In this case, $\mathcal{S}(\Sigma_n^N)$ and $\mathcal{S}(\Sigma_{n+1}^N)$ do no longer exhibit any common point so that it is necessary to introduce a suitable and consistent criterion to select a single model on $\mathcal{S}(\Sigma_n^N)$. The criterion that will be proposed in the following is based on the shift property of time–invariant dynamic systems described by relation (31).

Let $P' = (\sigma_s^{2\,\prime}, \sigma_w^{2\,\prime})$ and $P'' = (\sigma_s^{2\,\prime\prime}, \sigma_w^{2\,\prime\prime})$ be the intersections of a line from the origin with $\mathcal{S}(\Sigma_n)$ and $\mathcal{S}(\Sigma_{n+1})$ respectively, so that

$$\frac{\sigma_w^{2\,\prime}}{\sigma_s^{2\,\prime}} = \frac{\sigma_w^{2\,\prime\prime}}{\sigma_s^{2\,\prime\prime}}. \quad (34)$$

Define then the cost function

$$J(P', P'') = \|\hat{\Sigma}_{n+1}(P'')\, v(P')\|_2^2 = v^T(P')\hat{\Sigma}_{n+1}^2(P'')\, v(P'), \quad (35)$$

where

$$v(P') = [0\ \theta^T(P')\,]^T. \quad (36)$$

This function exhibits the following properties:

i) $J(P', P'') \geq 0$
ii) $J(P', P'') = 0 \Leftrightarrow P' = P'' = P^*$.

It is thus possible to solve Problem 1 searching for the solution that minimizes (35).

The following consistent algorithm for identifying noisy AR processes from finite sequences of data can thus be considered.

Algorithm 1.

(1) Compute the estimates of Σ_n and Σ_{n+1} given by the sample quantities

$$\Sigma_n^N = \frac{1}{N-n}\sum_{t=n+1}^{t=N} \varphi_y(t)\,\varphi_y^T(t),$$

$$\Sigma_{n+1}^N = \frac{1}{N-n-1}\sum_{t=n+1}^{t=N-1} \bar{\varphi}_y(t)\,\bar{\varphi}_y^T(t).$$

(2) Start from a generic point (a generic direction) $\xi = (\xi_1, \xi_2)$, $\xi_1 \geq \xi_2$ in the first quadrant of $\mathcal{R}^2$ and compute, by means of (21)–(22), the corresponding points $P' = (\sigma_s^{2\,\prime}, \sigma_w^{2\,\prime})$, $P'' = (\sigma_s^{2\,\prime\prime}, \sigma_w^{2\,\prime\prime})$ on $\mathcal{S}(\Sigma_n^N)$ and $\mathcal{S}(\Sigma_{n+1}^N)$.

(3) Compute $\hat{\Sigma}_n(P')$, $\hat{\Sigma}_{n+1}(P'')$ and $\theta(P')$ by means of the relations

$$\hat{\Sigma}_n(P') = \Sigma_n^N - \text{diag}\,[\sigma_w^{2\,\prime}\, I_n,\ \sigma_s^{2\,\prime}],$$

$$\hat{\Sigma}_{n+1}(P'') = \Sigma_{n+1}^N - \text{diag}\,[\sigma_w^{2\,\prime\prime}\, I_{n+1},\ \sigma_s^{2\,\prime\prime}],$$

$$\hat{\Sigma}_n(P')\,\theta(P') = 0.$$

(4) Compute the cost function $J(P', P'')$.

(5) Search on $\mathcal{S}(\Sigma_n^N)$ for the point $P^\circ = (\sigma_s^{2\circ}, \sigma_w^{2\circ})$ associated with the minimum of $J(P', P'')$.

(6) Estimate the driving noise variance as

$$\sigma_e^{2\circ} = \sigma_s^{2\circ} - \sigma_w^{2\circ}.$$

The ergodicity property A2 and property ii) assure the consistency of the proposed procedure since

$$\lim_{N\to\infty} \Sigma_n^N = \Sigma_n, \quad (37)$$

$$\lim_{N\to\infty} \Sigma_{n+1}^N = \Sigma_{n+1}, \quad (38)$$

$$\lim_{N\to\infty} \min_{\substack{P'\in\mathcal{S}(\Sigma_n^N)\\ P''\in\mathcal{S}(\Sigma_{n+1}^N)}} J(P', P'') = 0, \quad (39)$$

and relation (39) holds only for $P' = P'' = P^*$.

5. SIMULATION RESULTS

The performance of Algorithm 1 has been tested by means of numerical simulations whose results have been compared with those of the bias–compensated least–squares method proposed by Zheng (Zheng, 1999), which, according to the author, gives better results than previous traditional methods. Consider the following 4th–order AR model (Jia *et al.*, 2003)

$$x(t) = 2.4\,x(t-1) - 3.03\,x(t-2) + 1.986\,x(t-3) - 0.6586\,x(t-4) + e(t),$$

where $e(t)$ is white noise with unknown variance $E[e^2(t)] = \sigma_e^{2*} = 1$. Two Monte Carlo simulations of 100 independent runs have been performed with signal–to–noise ratios (SNR), defined as

$$\text{SNR} = 20\,\log_{10}\sqrt{\frac{E[x^2(t)]}{E[w^2(t)]}} = 20\,\log_{10}\sqrt{\frac{E[x^2(t)]}{\sigma_w^{2*}}},\quad \text{(dB)}$$

of 20 dB and 10 dB. Differently from (Jia *et al.*, 2003), the number of samples has been limited to $N = 1000$.

Table 1. True and estimated values of parameters and variances for Algorithm 1 and the BCLS method, SNR = 20 dB.

	α_1	α_2	α_3	α_4	σ_e^{2*}	σ_w^{2*}
true	2.4	-3.03	1.986	-0.6586	1	0.3973
Alg.1	2.3940 ± 0.0635	-3.0244 ± 0.1312	1.9828 ± 0.1259	-0.6656 ± 0.0509	1.0254 ± 0.1868	0.3924 ± 0.0283
BCLS	2.3945 ± 0.0639	-3.0253 ± 0.1320	1.9832 ± 0.1266	-0.6658 ± 0.0512	1.0249 ± 0.1842	0.3933 ± 0.0281

Table 2. True and estimated values of parameters and variances for Algorithm 1 and the BCLS method, SNR = 10 dB. * The values reported for the BCLS method have been averaged over 97 runs because of its lack of convergence in three runs.

	α_1	α_2	α_3	α_4	σ_e^{2*}	σ_w^{2*}
true	2.4	-3.03	1.986	0.6586	1	3.9730
Alg.1	2.3239 ± 0.4648	-2.8904 ± 0.9159	1.8578 ± 0.8577	-0.6170 ± 0.3288	1.6099 ± 1.9908	3.8376 ± 0.4464
BCLS*	0.9815 ± 0.0727	-0.4349 ± 0.0918	-0.3532 ± 0.0667	0.1500 ± 0.0241	10.5626 ± 1.3009	1.4053 ± 0.6292

In every run, a gaussian white noise sequence $w(\cdot)$ has been generated by means of the function `randn` of MATLAB and added to the AR output $x(\cdot)$.

The results are summarized in Tables 1 and 2 that report the true values of parameters and variances, the mean values of their estimates and the corresponding standard deviations.

The good selectivity of cost function (35) can be observed in Figure 2 that reports the values of $J(P', P'')$ versus the noise variance σ_w^2 along $\mathcal{S}(\Sigma_n^N)$ in a typical run of the Monte Carlo simulation.

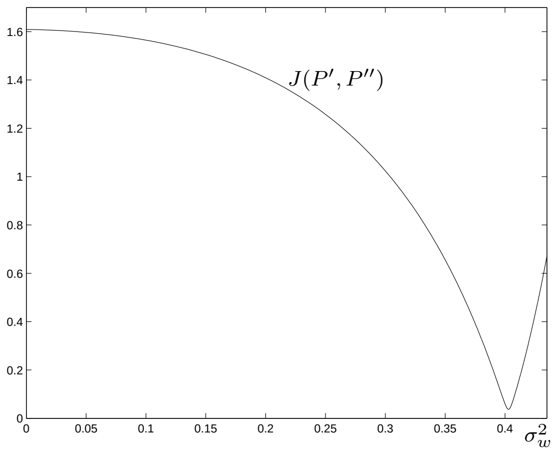

Fig. 2. Values of $J(P', P'')$ along $\mathcal{S}(\Sigma_n^N)$ in a typical run of the Monte Carlo simulation, SNR = 20 dB.

It can be observed from Table 1 that, for high SNR, Algorithm 1 and the BCLS method give similar results. In these cases the BCLS method can be preferable because of its simpler implementation. Table 2 shows, on the contrary, that in presence of high amounts of noise, the BCLS algorithm has convergence problems and is significantly outperformed by Algorithm 1.

6. CONCLUSIONS

In this paper, a new identification method for identifying autoregressive models in presence of additive white noise has been proposed. This approach relies, in particular, on the properties of the family of solutions of the dynamic Frisch scheme and on the shift property of time–invariant dynamic systems.

The effectiveness of the proposed algorithm has been tested by means of Monte Carlo simulations which show that this approach yields better estimates than those obtained with bias–compensated least–squares methods, especially in presence of low signal–to–noise ratios.

REFERENCES

Beghelli, S., R. Guidorzi and U. Soverini (1990). The Frisch scheme in dynamic system identification. *Automatica*, 26, 171–176.

Guidorzi, R. and M. Pierantoni (1995). A new parametrization of Frisch scheme solutions. *Proceedings of the XII International Conference on Systems Science*, pp. 114–120, Wroclaw, Poland.

Haykin, S. (1991). *Adaptive Filter Theory*. Prentice–Hall, Englewood Cliffs, New Jersey.

Jia, L.-J., S. Kanae, Z.-J. Yang and K. Wada (2003). On bias compensation estimation for noisy AR process. *Proceedings of the 42nd IEEE Conference on Decision and Control*, pp. 405–410, Maui, Hawaii, USA.

Jin, C.-Z., L.-J. Jia, Z.-J. Yang and K. Wada (2002). On convergence of a BCLS algorithm for noisy autoregressive process estimation. *Proceedings of the 41st IEEE Conference on Decision and Control*, pp. 4252–4257, Las Vegas, Nevada, USA.

Kay, S. M. (1980). Noise compensation for autoregressive spectral estimates. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28, 292–303.

Kay, S. M. (1988). *Modern Spectral Estimation*. Prentice–Hall, Englewood Cliffs, New Jersey.

Nehorai, A. and P. Stoica (1988). Adaptive algorithms for constrained ARMA signals in the presence of noise. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36, 1282–1291.

Pagano M. (1974). Estimation of models of autoregressive signal plus white noise. *The Annals of Statistics*, 2, 99–108.

Paliwal, K. K. (1986). A noise–compensated long correlation matching method for AR spectral estimation of noisy signals. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1369–1372, Tokyo, Japan.

Sakai, H. and M. Arase (1979). Recursive parameter estimation of an autoregressive process disturbed by white noise. *International Journal of Control*, 30, 949–966.

Tong, H. (1975). Autoregressive model fitting with noisy data by Akaike's information criterion. *IEEE Transactions on Information Theory*, 21, 476–480.

Zheng, W. X. (1999). A least–squares based method for autoregressive signals in the presence of noise. *IEEE Transactions on Circuits and Systems–II: Analog and Digital Signal Processing*, 46, 81–85.

Zheng, W. X. (2001). Unbiased parameter identification for noisy autoregressive signals. *Proceedings of the 2001 International Symposium on Circuits and Systems*, pp. 121–124, Sydney, Australia.