IDENTIFICATION OF NONCAUSAL NONMINIMUM PHASE AR MODELS USING HIGER-ORDER STATISTICS

H. Tora and D. M. Wilkes

Vanderbilt University
Department of Electrical and Computer Engineering
Nashville, TN, 37235, USA

ABSTRACT

In this paper, we address the problem of estimating the parameters of a noncausal autoregressive (AR) signal from estimates of the higher-order cumulants of noisy observations. The proposed family of techniques uses both 3rd-order and 4th-order cumulants of the observed output data. Consequently, at low SNR, they provide superior performance to methods based on autocorrelations. The measurement noise is assumed to be Gaussian and may be colored. The AR model parameters here are directly related to the solution of a generalized eigenproblem. The performance is illustrated by means of simulation examples.

1. INTRODUCTION

There are several motivations behind the use of higher order statistics in several areas of system identification and signal processing. First, higher order cumulants are blind to any kind of Gaussian process. Hence, when the processed signal is non-Gaussian and the additive noise is Gaussian, the noise will vanish in the cumulant domain [9]. Thus, a greater degree of noise immunity is possible. Second, cumulants are useful for identifying nonminimum phase systems or reconstructing nonminimum phase signals if the input signals are non-Gaussian. That is because cumulants preserve the phase information of the signal [9]. Third, cumulants are useful for detecting and characterizing the properties of nonlinear systems. In this paper, the first two properties are exploited.

Various methods have been suggested for MA, AR, and ARMA model identification based on higher-order statistics [1,2,3,4,5,6,7,8,9,10]. Since most of them have used second-order statistics along with higher order statistics, their performance has been sensitive to additive Gaussian noise. Techniques that use second-order information can be particularly sensitive to additive colored Gaussian noise. In this study, third- and fourth-order statistics are used to identify noncausal AR models. The additive noise therefore does not strongly affect the performance of the method.

This paper is organized as follows. In Section 2 the notation and model assumptions are presented. In Section 3 the proposed family of techniques is derived, and in Section 4 some results are presented. Finally, in Section 5 the paper is summarized.

2. MODEL ASSUMPTIONS

2.1 Signal Model

The noiseless discrete-time process $y_s(t)$ satisfies the following stochastic difference equation

$$\sum_{i=-p}^{q} a(i) y_s(t-i) = x(t) \tag{1}$$

$$y(t) = y_s(t) + w(t) \tag{2}$$

where the $a(i)$, $i=-p$, …, q, are the model coefficients to be estimated, $y_s(t)$ is the output sequence, $w(t)$ is the noise sequence, and $x(t)$ is a non-Gaussian, i.i.d. input sequence, with zero mean, $E\{x(t)\}=0$, variance, $\sigma_x^2 = E\{x^2(t)\} \neq 0$, non-zero skewness, $\gamma_{3x} = E\{x^3(t)\} \neq 0$, and non-zero kurtosis, $\gamma_{4x} = E\{x^4(t)\} - 3E^2\{x^2(t)\} \neq 0$.

The system is assumed to be non-minimum phase and exponentially stable with a transfer function given by

$$H(z) = \sum_{i=-\infty}^{\infty} h(i) z^{-i} = \frac{1}{A(z)} = \frac{1}{\sum_{i=-p}^{q} a(i) z^{-i}}$$

$$= \frac{1}{\prod_{i=1}^{p}(1-u_i z)\prod_{i=1}^{q}(1-v_i z^{-1})} \tag{3}$$

From (3), the output sequence can also be expressed in the time domain as

$$y_s(t) = \sum_{i=-\infty}^{\infty} x(i) h(t-i) \tag{4}$$

where $h(t)$ is the impulse response of the AR model.

The objective of this paper is to estimate the AR parameters $a(i)$, $i=-p$, ..., q, based on estimates of the higher-order cumulants of the noisy observation $y(t)$.

2.2 Preliminaries

Both 4th-order and 3rd-order cumulants are exploited to estimate the parameters of the non-causal AR system. For a zero-mean stationary process $\{y(t)\}$, it can be shown [9] that the third order and fourth order cumulants are given by

$$C_{3y}(\tau_1,\tau_2)=E\{y(t)y(t+\tau_1)y(t+\tau_2)\} \tag{5}$$

$$\begin{aligned} C_{4y}(\tau_1,\tau_2,\tau_3) &= E\{y(t)y(t+\tau_1)y(t+\tau_2)y(t+\tau_3)\} \\ &-E\{y(t)y(t+\tau_1)\}E\{y(t+\tau_2)y(t+\tau_3)\} \\ &-E\{y(t)y(t+\tau_2)\}E\{y(t+\tau_1)y(t+\tau_3)\} \\ &-E\{y(t)y(t+\tau_3)\}E\{y(t+\tau_1)y(t+\tau_2)\} \end{aligned} \tag{6}$$

Because for a Gaussian process, the kth -order cumulants vanish for $k \geq 3$, the higher-order cumulants of $y_s(t)$ and the noisy $y(t)$ are identical. The insensitivity of the cumulants to the presence of additive colored Gaussian noise is an important reason for employing cumulants, even for cases where the autocorrelation sequence is sufficient.

3. A FAMILY OF ALGORITHMS

3.1 The Basic Idea

It is well-known [9] that the higher-order cumulants are not affected by Gaussian noise whereas the 2nd-order autocorrelation estimates suffer from such noise. Since our approach for identification of non-causal AR systems uses 3rd- and 4th-order cumulants, it is robust to additive Gaussian noise and more robust than other techniques based on the autocorrelation sequence, in the sense of signal-to-noise ratio (SNR). Our simulation examples show that good identification is achieved for low SNR cases. We, therefore, ignore the noise effects in our derivations. Our starting point is equation (1) with $y_s(t)$ replaced by $y(t)$, i.e. $y_s(t)=y(t)$.

Multiplying both sides of (1) by $y(t-k)y(t-l)y(t-m)$ and taking the expected value, we obtain

$$\begin{aligned} &\sum_{i=-p}^{q} a(i)E\{y(t-i)y(t-k)y(t-l)y(t-m)\} \\ &\quad = E\{x(t)y(t-k)y(t-l)y(t-m)\} \end{aligned} \tag{7}$$

Now, by defining λ_k as

$$\lambda_k = \left(\frac{\gamma_{4x}}{\gamma_{3x}}\right)h(-k) \tag{8}$$

the fundamental expression of our method can be shown to be

$$\begin{aligned} &\sum_{i=-p}^{q} a(i)C_{4y}(i-k,i-l,i-m) \\ &= \lambda_k \sum_{i=-p}^{q} a(i)C_{3y}(i-l,i-m) \end{aligned} \tag{9}$$

This can be written in matrix form as

$$\mathbf{C}_{4yk}\mathbf{a}=\lambda_k \mathbf{C}_{3y}\mathbf{a} \tag{10}$$

where

$\mathbf{a}=[a(-p),a(-p+1),...,a(q-1),a(q)]^{\mathbf{T}}$.

Since the matrices are not square, the equation is multiplied by $\mathbf{C}_{3y}^{\mathbf{T}}$ on the left-hand side to obtain

$$\mathbf{C}_{3y}^{\mathbf{T}}\mathbf{C}_{4yk}\mathbf{a}=\lambda_k \mathbf{C}_{3y}^{\mathbf{T}}\mathbf{C}_{3y}\mathbf{a} \tag{11}$$

It is obvious that the vector $\mathbf{a}$ is the generalized eigenvector corresponding to the eigenvalue λ_k. The cumulant matrices can be directly estimated from output and eigenvector problem can be solved uniquely. We obtain $p+q+1$ eigenvectors, consequently, we have to choose which one among them corresponds to the true solution. To do that, we generate estimated input signals by using each eigenvector as an FIR filter, and calculate the kurtosis values for each input signal estimate. It has been shown [11] that the eigenvalue-eigenvector pair that maximizes the kurtosis of these input estimates is the best estimate of the AR model.

3.2 Using More Information

In fact, we do not know at the beginning which value of k should be chosen. To overcome that problem, a set of k values is employed to form the cumulant matrices in (11). We solve (11) in each case and note the λ_k in each solution. Now we may form a large set of equations as

$$\begin{pmatrix} \cdot \\ \cdot \\ \mathbf{C}_{4yk_{-1}} \\ \mathbf{C}_{4yk_0} \\ \mathbf{C}_{4yk_1} \\ \cdot \\ \cdot \end{pmatrix} \mathbf{a} = \begin{pmatrix} \cdot \\ \cdot \\ \lambda_{k_{-1}}\mathbf{C}_{3y} \\ \lambda_{k_0}\mathbf{C}_{3y} \\ \lambda_{k_1}\mathbf{C}_{3y} \\ \cdot \\ \cdot \end{pmatrix} \mathbf{a} \tag{12}$$

For simplicity, this equation may be written as

$$\mathbf{C_4 a}=\mathbf{C_3 a} \tag{13}$$

or equivalently

$$(\mathbf{C}_4-\mathbf{C}_3)\mathbf{a}=\mathbf{0} \tag{14}$$

By taking the SVD of $(\mathbf{C_4}-\mathbf{C_3})$ we choose as the solution, the right singular vector with the smallest singular value (which should be zero in the ideal case).

3.3 Gradient Approaches

Since Eq. (11) is a generalized non-symmetric eigenproblem, there is always a chance of getting a complex eigenvalue-eigenvector pair, which is not acceptable for the technique presented in this paper. To overcome this handicap, we present another solution technique based on minimizing a cost function that guarantees that the solution to the AR model is always real.

From (11), we define the following error vector

$$\varepsilon_k = (\mathbf{C}_{4yk} - \lambda_k \mathbf{C}_{3y})\mathbf{a} \quad (15)$$

As a performance measure or cost function, we introduce the squared error defined as follows:

$$J_k = \|\varepsilon_k\|^2 = \varepsilon_k^T \varepsilon_k \quad (16)$$

We now search for the vector $\mathbf{a}$ for which the squared error J_k is minimum.

Substituting Eq. (15) into (16), we obtain

$$J_k = \mathbf{a}^\mathbf{T}(\mathbf{C}_{4yk}^\mathbf{T} - \lambda_k \mathbf{C}_{3y}^\mathbf{T})(\mathbf{C}_{4yk} - \lambda_k \mathbf{C}_{3y})\mathbf{a}$$

$$= \mathbf{a}^\mathbf{T}(\mathbf{C}_{4yk}^\mathbf{T}\mathbf{C}_{4yk} + \lambda_k^2 \mathbf{C}_{3y}^\mathbf{T}\mathbf{C}_{3y} - \lambda_k \mathbf{C}_{3y}^\mathbf{T}\mathbf{C}_{4yk} - \lambda_k \mathbf{C}_{4yk}^\mathbf{T}\mathbf{C}_{3y})\mathbf{a}$$

$$= \mathbf{a}^\mathbf{T}[\mathbf{C}_k + \lambda_k^2 \mathbf{B}_k - \lambda_k(\mathbf{A}_k + \mathbf{A}_k^\mathbf{T})]\mathbf{a} \quad (17)$$

$$= \mathbf{a}^\mathbf{T}\mathbf{D}_k(\lambda_k)\mathbf{a}$$

To determine the optimum vector $\mathbf{a}$, we differentiate the cost function J_k with respect to λ_k and then set the result equal to zero.

Differentiating Eq. (17) with respect to λ_k, we readily find that

$$\nabla_{\lambda_k} J_k = 2\lambda_k \mathbf{a}^\mathbf{T}\mathbf{B}_k\mathbf{a} - \mathbf{a}^\mathbf{T}(\mathbf{A}_k + \mathbf{A}_k^\mathbf{T})\mathbf{a} = 0 \quad (18)$$

which can be solved for λ_k to obtain

$$\lambda_k = \frac{\mathbf{a}^\mathbf{T}\left(\frac{\mathbf{A}_k + \mathbf{A}_k^\mathbf{T}}{2}\right)\mathbf{a}}{\mathbf{a}^\mathbf{T}\mathbf{B}_k\mathbf{a}} \quad (19)$$

Due to the nature of our method, the cost function J_k is a function of both λ_k and $\mathbf{a}$. Therefore, we wish to minimize J_k with respect to λ_k and $\mathbf{a}$ simultaneously.

We seek the optimum solution $\mathbf{a}$ which satisfies all the k values because we are getting different fourth-order cumulant matrices $\mathbf{C}_{4yk}$ for each value of k. To do so, we define a new cost function which considers a set of k values.

$$J = \sum_k \mathbf{a}^\mathbf{T}\mathbf{D}_k(\mathbf{a})\mathbf{a} \quad (20)$$

where $\mathbf{D}_k(\mathbf{a}) = \mathbf{C}_k + \lambda_k^2\mathbf{B}_k - \lambda_k(\mathbf{A}_k + \mathbf{A}_k^\mathbf{T})$. On the other hand, λ_k is also a function of the vector $\mathbf{a}$ as can be seen from (19).

Substituting (19) for λ_k and differentiating (20) with respect to $\mathbf{a}$, we find

$$\nabla_\mathbf{a} J = \sum_k \nabla_\mathbf{a}(\mathbf{a}^\mathbf{T}\mathbf{D}_k(\mathbf{a})\mathbf{a}) \quad (21)$$

where

$$\nabla_\mathbf{a}(\mathbf{a}^\mathbf{T}\mathbf{D}_k(\mathbf{a})\mathbf{a}) = 2\mathbf{D}_k(\mathbf{a})\mathbf{a} + [2\lambda_k\mathbf{a}^\mathbf{T}\mathbf{B}_k\mathbf{a} - \mathbf{a}^\mathbf{T}(\mathbf{A}_k + \mathbf{A}_k^\mathbf{T})\mathbf{a}] \times \left[\frac{(\mathbf{A}_k + \mathbf{A}_k^\mathbf{T})\mathbf{a}}{\mathbf{a}^\mathbf{T}\mathbf{B}_k\mathbf{a}} - \frac{\mathbf{a}^\mathbf{T}(\mathbf{A}_k + \mathbf{A}_k^\mathbf{T})\mathbf{a}}{(\mathbf{a}^\mathbf{T}\mathbf{B}_k\mathbf{a})^2}\mathbf{B}_k\mathbf{a}\right] \quad (22)$$

We employ the method of steepest descent to find the optimum vector $\mathbf{a}$. According to this method, the adjustment applied to the vector $\mathbf{a}$ at iteration n is defined by

$$\Delta\mathbf{a}(n) = -\eta\nabla_\mathbf{a} J(n) \quad (23)$$

where η is a positive step-size parameter. Given the old value of $\mathbf{a}$ at iteration n, the updated value of this $\mathbf{a}$ at the next iteration $n+1$ is computed as

$$\mathbf{a}(n+1) = \mathbf{a}(n) + \Delta\mathbf{a}(n) \quad (24)$$

This iteration method is summarized as follows:

1. Assign an initial value to $\mathbf{a}$.
2. Compute $\nabla_\mathbf{a} J(n)$ using (21) and (22).
3. Update $\mathbf{a}$ using (24).
4. Go to step 2 if $J(n) > \rho$ (some chosen tolerance).

In order to speed up the convergence and improve the likelihood of reaching the global minimum of the error surface, instead of choosing an arbitrary initial value for $\mathbf{a}$, we select a rational guess. To do that, we developed a technique from our simulation experiences as follows:

1. Select a k and a range of λ_k values.
2. Compute $\mathbf{D}_k(\lambda_k)$.
3. Find the eigenvalues and eigenvectors of $\mathbf{D}_k(\lambda_k)$.
4. Assign the eigenvector corresponding to the smallest eigenvalue to $\mathbf{a}$.
5. Compute J_k from (17) for each λ_k.
6. Find all the minima of J_k. From our experience, J_k has more than one minimum.
7. Select the minimum that maximizes the output kurtosis as the initial value of $\mathbf{a}$.

4. RESULTS

To simulate our algorithm, we generated an i.i.d. exponentially distributed random process with zero mean and finite cumulants for the input, which we convolved with the true impulse response. Zero mean, Gaussian noise (white and colored were considered) were added to produce a signal-to-noise ratio of 10dB. Colored noise was generated by passing white Gaussian noise through a first order AR filter with its pole at 0.95. To reduce the realization dependency of our simulations, we averaged over 20 Monte Carlo runs.

Example : The second-order noncausal model that we simulated is

$$H(z) = \frac{1}{(1-0.8z)(1+0.75z^{-1})}$$

with a causal pole at -0.75 and an anticausal pole at 1.25.

The results for N=1024 data points are below.

TRUE AR(2) COEFF.	EST AR(2) COEFF., N=1024, SNR=10DB, 20 MONTE CARLO RUNS WHITE	COLORED
-0.6854	-0.6860 $\pm$ 0.0383	-0.6944 $\pm$ 0.0387
0.3427	0.3678 $\pm$ 0.0903	0.3524 $\pm$ 0.0534
0.6425	0.6196 $\pm$ 0.0409	0.6232 $\pm$ 0.0396

The table below is for N=4096.

TRUE AR(2) COEFF.	EST AR(2) COEFF., N=4096, SNR=10DB, 20 MONTE CARLO RUNS WHITE	COLORED
-0.6854	-0.6868 $\pm$ 0.0140	-0.6949 $\pm$ 0.0397
0.3427	0.3562 $\pm$ 0.0477	0.3378 $\pm$ 0.0533
0.6425	0.6314 $\pm$ 0.0236	0.6311 $\pm$ 0.0295

5. CONCLUSIONS

We examined the problem of estimating the parameters of a noncausal AR signal from estimates of the higher-order cumulants of noisy observations. The family of techniques presented herein uses both 3rd-order and 4th-order cumulants of the observed data. Consequently, for additive Gaussian noise at low SNR, they provide superior performance to methods based on autocorrelations. The AR model parameters here are directly related to the solution of a generalized eigenproblem.

6. REFERENCES

[1] G.B. Giannakis, J.M. Mendel, and W. Wang, “ARMA modeling using cumulant and autocorrelation statistics”, *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Dallas, TX, Apr. 1987, pp. 61-64.

[2] M.R. Raghuveer and C.L. Nikias, “Bispectrum estimation: A parametric approach,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp.1213-1230, 1985.

[3] C.L. Nikias and H.H. Chiang, “Non-causal autoregressive bispectrum estimation and deconvolution,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp.1911-1913, 1988.

[4] J.K. Tugnait, “Fitting noncausal autoregressive signal plus noise models to noisy non-Gaussian linear process,” *Proc. Int. Conf. Acoust., Speech, Sig. Proc.*, Dallas, 1987, pp. 324-327.

[5] M. Huzii, “ Estimation of coefficients of an AR process by using a higher-order moment,” *J. Time Series Anal*., pp. 87-93, 1981.

[6] K.S. Lii and M. Rosenblatt, “Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear process,” *Ann. Stat*., pp. 1195-1208, 1982.

[7] H.H.Chiang and C.L. Nikias, “Adaptive deconvolution and identification of nonminimum phase FIR system based on cumulants,” *IEEE Trans. Automat. Contr*., vol. AC-35, Jan. 1990, pp. 37-47.

[8] G.B. Giannakis and A. Swami, “On estimating noncausal nonminimum phase ARMA models of non-Gaussian process,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38,March 1990, pp. 478-495.

[9] J.M. Mendel, “Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications,” *Proc. IEEE*, vol. 79, March 1991, pp. 278-305.

[10] M.R. Raghuveer and C.L. Nikias, “Bispectrum estimation via AR modeling,” *Signal Processing*, vol. 10, pp. 35-48, Jan. 1986.

[11] O. Shalvi and E. Weinstein, “New criteria for blind deconvolution of nonminimum phase systems (channels),” *IEEE Trans. Information Theory*, 36(2):312-321, March 1990.