approach to adaptive filtering in subbands is also described, where an expected improvement in convergence speed is observed. Some practical aspects of the subband implementation are also briefly discussed.

References

[1] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 309–316, Apr. 1986.

[2] W. B. Mikhael, F. H. Wu, L. G. Kazovsky, G. S. Kang, and L. J. Fransen, "Adaptive filters with individual adaptation of parameters," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 677–686, July 1986.

[3] F. F. Yassa, "Optimality in the choice of convergence factor for gradient based adaptive algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 48–59, Jan. 1987.

[4] J. B. Evans and B. Liu, "Variable step size methods for the LMS adaptive algorithm," in *Proc. IEEE Int. Symp. Circuits Syst.* (Philadelphia, PA), May 1987, pp. 422–425.

[5] T. J. Shan and T. Kailath, "Adaptive algorithm with an automatic gain control feature," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 122–127, Jan. 1988.

[6] S. Karni and G. Zeng, "A new convergence factor for adaptive filters," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 1011–1012, July 1989.

[7] S. S. Narayan, A. M. Peterson, and M. J. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 609–615, June 1983.

[8] D. F. Marshall, W. K. Jenkins, and J. J. Murphy "The use of orthogonal transforms for improving performance of adaptive filters," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 474–484, Apr. 1989.

[9] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[10] P. S. R. Diniz and J. E. Cousseau, "Optimal convergence factor for Gauss-Newton algorithms and its application to an adaptive parallel realization," in *Proc. IEEE Int. Telecommun. Symp.*, Rio de Janeiro, Brazil, Sept. 1990, pp. 420–424.

[11] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[12] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* (New York, NY), 1988, pp. 1572–1575.

[13] M. J. Smith and T. P. Barnwell, III, "Exact reconstitution techniques for tree-structured subband coders," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 434–441, June 1986.

Order Selection of Autoregressive Models

Petar M. Djurić and Steven M. Kay

***Abstract*—This correspondence addresses the problem of order determination of autoregressive models by Bayesian predictive densities. A criterion is derived employing noninformative prior densities of the model parameters. The form of the obtained criterion coincides with that of Rissanen in [16]. Simulation results are presented which demonstrate the good performance of the criterion, and comparisons with four other popular approaches verify its superiority in many cases.**

I. Introduction

It is a common practice in various scientific and engineering disciplines to represent observed discrete-time random processes by autoregressive (AR) models. The determination of the order and the estimation of the parameters of the models is then of major interest. Usually, the parameter estimation is carried out by least squares or maximum likelihood procedures. For model order determination, there exists a long list of approaches, ranging from the classical ones based on the estimated residuals of the fitting model [5] to those founded on information [3] and coding theory [14] or Bayesian analysis [13], [17].

In this correspondence we derive a criterion that is based on Bayesian predictive densities according to models and data [8]. From a decision theoretic point of view, when the errors for under- or overparametrization of a model have equal costs (importance), this approach under certain additional assumptions will yield a criterion that will minimize the overall probability of error. Noninformative priors of the model parameters will be employed in the derivation to account for the lack of knowledge regarding the model parameters and to allow "the data to speak for themselves." It should be noted that Rissanen has come to basically the same result using the concept of stochastic complexity [16]. The criterion is checked by Monte Carlo simulations and compared to other approaches which are common in the recent literature.

Manuscript received April 5, 1990; revised September 3, 1991.

P. M. Djurić is with the Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794.

S. M. Kay is with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI 02881.

IEEE Log Number 9202793.

II. Order Selection

Consider the sequence of samples $y_{1,N}^T = [\, y[1]y[2] \cdots y[N]\,]$ generated according to the AR model

$$y[n] = -a_{1p}y[n-1] - a_{2p}y[n-2] - \cdots - a_{pp}y[n-p] + e[n]$$

where p is the order of the model, $a_{1p}, a_{2p}, \cdots, a_{pp}$ the coefficients of the model, and $e[n]$ a Gaussian random variable. The coefficients of the model and the order p are unknown. In addition, we assume that

$$\mathcal{E}(e[n]) = 0$$

$$\mathcal{E}(e[n]e[m]) = \sigma^2\delta_{mn}$$

where $\mathcal{E}(\cdot)$ is the expectation operator, δ_{mn} the Kronecker delta function, and σ^2 the unknown noise variance; and

$$A(z) = 1 + a_{1p}z + a_{2p}z^2 + \cdots + a_{pp}z^p \neq 0, \qquad |z| \leq 1.$$

The problem is to select the optimal AR model. Optimality will be defined according to the overall probability of selection error. If it is assumed that the cost for incorrect selection is one and for correct selection zero, and the *a priori* probabilities of the considered models are all equal, then the error is minimized when the selection is carried out according to the predictive densities of the models. These densities are defined as densities of "future" data, conditioned on the assumption that the examined model is true [2]. Since the parameters of the model are usually unknown, they are estimated only from "past" data. If we formally write the predictive density of $y[n]$ as $f(\, y[n]|\, \mathbf{y}_{1,n-1}, k)$, where $\mathbf{y}_{1,n-1}$ is the vector of the past data, and k the order of the assumed model, the criterion will take the form

$$\hat{p} = \arg\left\{ \min_{k \in \{0,1,\cdots,P\}} \left(\sum_{n=2}^{N} - \ln f(\, y[n]|\mathbf{y}_{1,n-1}, k) \right) \right\} \quad (1)$$

where P is the maximum model order. Note that the model with zeroth order is included in the set of models.

We first outline the derivation of $f(y[n]|y_{1,n-1}, k)$ for $k > 0$ using the Bayesian approach. Namely,

$$f(y[n]|y_{1,n-1}, k > 0) = \int_{\sigma}\int_{a_k} f(y[n]|y_{1,n-1}, a_k, \sigma, k) \cdot f(a_k, \sigma | y_{1,n-1}, k)\, da_k\, d\sigma \quad (2)$$

where

$$f(y[n]| y_{1,n-1}, a_k, \sigma, k > 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} (y[n] + a_{1k}y[n-1] + a_{2k}y[n-2] + \cdots + a_{kk}y[n-k])^2 \right\} \quad (3)$$

and $f(a_k, \sigma | y_{1,n-1}, k)$ is the *a posteriori* density of the parameters a_k and σ after $y_{1,n-1}$ has been observed. Clearly, we are unable to write (3) for $n \leq k$ because we do not know the values of $y[0]$, $y[-1]$, $y[-2]$, etc., of the AR process. This entails that the exact evaluation of (2) will be impossible. Therefore, our ultimate goal is to find a reasonable approximation of (2). Since we are reluctant to assume anything about the unobserved samples, we consider $y[1]$, $y[2]$, $\cdots$, $y[k]$ to be initial conditions for the kth model. Moreover, in the derivation of $f(y[n]|y_{1,n-1}, k)$ we assume that $n > 2k + 1$, for reasons that will become clear below. In order to carry out the integration in (2), we need to find the form of $f(a_k, \sigma | y_{1,n-1}, k)$. According to Bayes' theorem and our assumptions

$$f(a_k, \sigma | y_{1,n-1}, k > 0) \propto f(y_{k+1,n-1}| y_{1,k}, a_k, \sigma, k > 0) \cdot f(a_k, \sigma | k > 0) \quad (4)$$

where $f(a_k, \sigma | k > 0)$ is the *a priori* density function of the parameters, and

$$f(y_{k+1,n-1}|y_{1,k}, a_k, \sigma, k > 0) = \frac{1}{(2\pi\sigma^2)^{n-1-k/2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=k+1}^{n-1} (y[j] + a_{1k}y[j-1] + a_{2k}y[j-2] + \cdots + a_{kk}y[j-k])^2 \right\}. \quad (5)$$

The *a priori* density function of the parameters should reflect our state of knowledge about these parameters. When very little is known about them, a usual route is to employ the noninformative priors [11]. For our problem [19, ch. 7]

$$f(a_k, \sigma | k > 0) \propto \frac{1}{\sigma}. \quad (6)$$

When (5) and (6) are substituted into (4), and the resulting $f(a_k, \sigma | y_{1,n-1}, k > 0)$ plugged into (2), the integration yields

$$f(y[n]| y_{1,n-1}, k > 0) = \frac{1}{\sqrt{2\pi}} \frac{|\boldsymbol{H}_k^T[n-2]\boldsymbol{H}_k[n-2]|^{1/2}}{|\boldsymbol{H}_k^T[n-1]\boldsymbol{H}_k[n-1]|^{1/2}} \frac{\Gamma\left(\frac{n-2k}{2}\right)}{\Gamma\left(\frac{n-2k-1}{2}\right)} \cdot \frac{\left(\frac{y_{k+1,n-1}^T \boldsymbol{P}_k^{\perp}[n-2] y_{k+1,n-1}}{2}\right)^{(n-2k-1)/2}}{\left(\frac{y_{k+1,n}^T \boldsymbol{P}_k^{\perp}[n-1] y_{k+1,n}}{2}\right)^{(n-2k)/2}} \quad n > 2k+1 \quad (7)$$

where

$$\boldsymbol{H}_k[l] = \begin{pmatrix} y[k] & y[k-1] & \cdots & y[1] \\ y[k+1] & y[k] & \cdots & y[2] \\ \vdots & \vdots & \vdots & \vdots \\ y[l] & y[l-1] & \cdots & y[l-k+1] \end{pmatrix} \quad l > 2k$$

$$\boldsymbol{P}_k^{\perp}[l] = \boldsymbol{I} - \boldsymbol{H}_k[l](\boldsymbol{H}_k^T[l]\boldsymbol{H}_k[l])^{-1}\boldsymbol{H}_k^T[l]$$

and $\Gamma(x)$ is the gamma function defined by

$$\Gamma(x) \stackrel{\text{def}}{=} \int_0^{\infty} t^{x-1} \exp\{-t\}\, dt, \qquad x > 0.$$

Equation (7) is the form of the predictive density that is substituted into (2) to obtain the criterion for the model of order k when $k > 0$. The evaluation of (7) is only possible for $n > 2k + 1$ and the summation in (2) runs from $n = 2$ to N. A discussion of how to approximate the first $2k$ terms will be given below.

If $k = 0$, the procedure for obtaining $f(y[n]|y_{1,n-1}, k = 0)$ is similar. The result is

$$f(y[n]|y_{1,n-1}, k = 0) = \frac{1}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\left(\frac{y_{1,n-1}^T y_{1,n-1}}{2}\right)^{(n-1)/2}}{\left(\frac{y_{1,n}^T y_{1,n}}{2}\right)^{(n/2)}} \quad n > 1. \quad (8)$$

Clearly, when $k = 0$, the summation in (1) runs from $n = 2$ to $n = N$, i.e.,

$$J_0 = -\sum_{n=2}^{N} \ln f(y[n]|y_{1,n-1}, k = 0). \quad (9)$$

This reflects our unwillingness to write the predictive density of the next sample unless we have a minimum number of samples (observations) to estimate the parameters of the model. Since for $k = 0$ the only unknown is σ^2, it follows that the procedure may start from $n = 2$.

The first order AR model has two unknowns, a_{11} and σ^2. We deduce that three observations is the least number of samples necessary to make a unique estimate of these parameters. Thus $n = 4$ is the time index of the first sample for which we can write the predictive density. Again, note that we did not assume zero initial conditions to allow the derivation of the predictive density of $y[n]$ for $n < 4$. If we had proceeded with those assumptions, it would

have entailed the use of different amounts of *a priori* information for different models, an approach which for short sequences might affect the selection procedure significantly. So for $k = 1$ we write

$$J_1' = -\sum_{n=4}^{N} \ln f(y[n]|y_{1,n-1}, k = 1). \tag{10}$$

The comparison between (9) and (10) is not appropriate because the two criteria have predictive density functions for different numbers of samples. A natural solution to this inconvenience is to change (10) to

$$J_1 = -\sum_{n=2}^{3} \ln f(y[n]|y_{1,n-1}, k = 0) - \sum_{n=4}^{N} \ln f(y[n]|y_{1,n-1}, k = 1).$$

We proceed in a similar fashion with the AR models of higher order. For example, for $k = 2$ we have

$$J_2 = -\sum_{n=2}^{3} \ln f(y[n]|y_{1,n-1}, k = 0) - \sum_{n=4}^{5} \ln f(y[n]|y_{1,n-1}, k = 1) - \sum_{n=6}^{N} \ln f(y[n]|y_{1,n-1}, k = 2).$$

Needless to say, the evaluation of J_k can be carried out recursively in time.

III. Brief Review of Other Approaches

Akaike [3] used the same assumptions employed in the derivation of (7) and (8) and suggested the following model selection criterion:

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,P\}} \left(\ln \hat{\sigma}_k^2 + \frac{2k}{N}\right)\right\} \tag{11}$$

where $\hat{\sigma}_k^2$ is again the maximum likelihood estimate of the noise variance σ^2. With the increase of k, the first term in (11) will monotonically decrease, but the second term will increase to account for the increase in variance due to the estimation of extra parameters. Similar in philosophy is the minimum description length (MDL) criterion given by [14], [17]

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,P\}} \left(\ln \hat{\sigma}_k^2 + \frac{k}{N}\ln N\right)\right\}. \tag{12}$$

It can be shown that for large N and using asymptotic approximations, (1) becomes identical to (12).

A third criterion that falls into this category of criteria is the predictive minimum description length (PMDL) [16]. It is similar in form to (1), i.e.,

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,P\}} \left(\sum_n -\ln f(y[n]|y_{1,n-1}, \hat{\boldsymbol{a}}_k[n-1], \hat{\sigma}_k[n-1], k)\right)\right\} \tag{13}$$

where $f(y[n]|y_{1,n-1}, \hat{\boldsymbol{a}}_k[n-1], \hat{\sigma}_k[n-1], k)$ is the quasi-likelihood predictive density of the data [10], and $\hat{\boldsymbol{a}}_k[n-1]$ and $\hat{\sigma}_k[n-1]$ are the maximum likelihood estimates of $\boldsymbol{a}_k$ and σ from the first $n-1$ samples. (In the statistical literature the terms maximum likelihood plug-in forecast density [7] and estimative density are also used [1].) A clear distinction must be made between the predictive density used in (2) and the quasi-likelihood predictive density in (13). The former has been shown to be better for discrimination purposes than the latter [1]. (13) can be rewritten as

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,P\}} \left(\sum_{n=k+1}^{N-1} \left(\ln \hat{\sigma}_k^2[n] + \frac{\hat{e}_k^2[n]}{\hat{\sigma}_k^2[n]}\right)\right)\right\}$$

where

$$\hat{e}_k[n] = y[n] + \hat{a}_{1k}[n-1]y[n-1] + \cdots + \hat{a}_{kk}[n-1]y[n-k].$$

In the predictive least squares (PLS) criterion the probabilistic assumptions about the data are relaxed. The criterion is based on accumulated squared prediction errors [15], or

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,P\}} \left(\sum_{n=1}^{N} (y[n] - \hat{y}_k[n])^2\right)\right\} \tag{14}$$

where

$$\hat{y}_k[n] = -\hat{a}_{1k}[n-1]y[n-1] - \hat{a}_{2k}[n-1]y[n-2] - \cdots - \hat{a}_{kk}[n-1]y[n-k]. \tag{15}$$

Finally, we want to introduce a criterion similar to (14). Instead of validating the model by accumulated squared prediction errors, we shall employ the accumulated absolute values of the prediction errors, or

$$\hat{p} = \arg\left\{\min_{k\in\{0,1,\cdots,P\}} \left(\sum_{n=1}^{N} |y[n] - \hat{y}_k[n]|\right)\right\} \tag{16}$$

where $\hat{y}_k[n]$ is given by (15). We shall refer to (16) as the predictive least absolute value (PLAV) criterion. The parameters $\hat{\boldsymbol{a}}_k$ can be estimated by employing the least absolute value criterion [4], or the least squares criterion using lattice filters [18], [9]. In the simulations presented in the next section, the latter approach was used.

IV. Simulation Results

In this section the results of eight Monte Carlo experiments are presented. The performance of the six methods was assessed on 6 different autoregressive processes:

$$y[n] = 0.5y[n-1] + e[n] \tag{17}$$

$$y[n] = 0.7y[n-1] + e[n] \tag{18}$$

$$y[n] = 1.8y[n-1] - 0.97y[n-2] + e[n] \tag{19}$$

$$y[n] = 1.37y[n-1] - 0.56y[n-2] + e[n] \tag{20}$$

$$y[n] = 1.352y[n-1] - 1.338y[n-2] + 0.662y[n-3] - 0.240y[n-4] + e[n] \tag{21}$$

and

$$y[n] = 2.760y[n-1] - 3.809y[n-2] + 2.654y[n-3] - 0.924y[n-4] + e[n]. \tag{22}$$

Equations (17) and (18) were used in [18] and [12, eqs. (21) and (22)]. The noise variance was always $\sigma^2 = 1$ except in the last experiment when it was $\sigma^2 = 0.1$. The maximum model order P was 8. Each experiment was repeated 1000 times. The number of samples was varied between 20 and 100.

Table I shows the results obtained from the analysis of (17) with $N = 40$. MDL performed the best and PLS the worst. In [18] much better results were reported for PLS, even for sequences that had

TABLE I
COMPARISON RESULTS FOR EXAMPLE (17). THE SEQUENCES HAD 40 SAMPLES

$p = 1, a_{11} = 0.5, \sigma^2 = 1, N = 40$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	222	360	389	31	107	324
1	747	602	592	622	805	627
2	28	33	16	124	57	33
3	3	5	3	58	19	14
4	0	0	0	40	5	1
5	0	0	0	32	4	0
6	0	0	0	34	2	1
7	0	0	0	28	1	0
8	0	0	0	31	1	0

TABLE II
COMPARISON RESULTS FOR EXAMPLE (17). THE SEQUENCES HAD 100 SAMPLES

$p = 1, a_{11} = 0.5, \sigma^2 = 1, N = 100$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	10	111	189	0	1	93
1	969	857	798	702	955	873
2	18	30	12	127	37	30
3	3	2	1	63	5	4
4	0	0	0	30	1	0
5	0	0	0	24	1	0
6	0	0	0	18	0	0
7	0	0	0	17	0	0
8	0	0	0	19	0	0

fewer samples. The reason for PLS's better performance was the absence of the zeroth order model in the examined set of models.

In the second set of experiments, the same process was used except that the sequences had $N = 100$ samples. Now PDC yielded the correct order most often. Table II shows the results. In these two experiments PLAV slightly outperformed PLS.

Table III gives the results obtained from the analysis of (18) on sequences $N = 40$. Again, the highest percentage of correct selections was achieved by PDC. AIC had the fewest correct selections, but also the fewest underparametrizations.

In Tables IV–VI the results of the analysis of second-order processes are shown. Table IV shows the performances obtained from sequences $N = 40$ and generated according to (19). We see that this time PLS and PLAV outperformed the MDL criterion. Tables V and VI give the corresponding results from sequences generated by (20) with $N = 40$ and $N = 80$, respectively. Comparing Tables V and VI, we deduce that with an increase in the number of samples, the performance rankings of the criteria may quickly change. In Table VI we see that the only criterion which did not underestimate the model order was AIC.

Finally, in Tables VII and VIII we can see the results obtained when the criteria were applied to sequences generated according to (21) and (22). The processes (21) and (22) have broad-band and narrow-band spectral characteristics, respectively. In the broad-band case only, AIC, MDL, and PDC made correct selections in more than 50% of the cases. When the narrow-band process was analyzed, satisfactory performance was achieved only by PDC and PLAV. Note also the difference in performance between PLAV and PLS.

We have to be very careful in the interpretation of these results and not overemphasize their significance. We have investigated the performance of the criteria when the sequences represented finite low-order autoregressive processes. In practice, unfortunately, the orders might be finite and large or even infinite. However, certain guidelines are necessary, and simulation results similar to ours may provide useful information about the characteristics of the examined methods. We have drawn the following conclusions from the experiments.

1) The best performance was achieved by PDC when the sequences represented narrow-band processes, or their lengths were big enough (how big depended on the nature of the process; the more narrow-band the process, the shorter the sequence length required to outperform the rest of the methods).
2) For wide-band processes, MDL yielded the best results.
3) PLAV outperformed PLS in every experiment.
4) AIC had the lowest number of underparametrizations.

The performance of PMDL was quite mediocre. Such a perfor-

TABLE III
COMPARISON RESULTS FOR EXAMPLE (18)

$p = 1, a_{11} = 0.7, \sigma^2 = 1, N = 40$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	25	103	184	1	5	112
1	950	855	795	682	911	819
2	22	35	21	103	58	50
3	2	6	0	59	11	9
4	1	1	0	36	7	6
5	0	0	0	28	2	2
6	0	0	0	29	2	2
7	0	0	0	25	2	0
8	0	0	0	37	2	0

TABLE IV
COMPARISON RESULTS FOR EXAMPLE (19)

$p = 2, a_{12} = 1.8, a_{22} = 0.97, \sigma^2 = 1, N = 40$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	0	7	26	0	0	5
1	0	12	24	0	0	10
2	969	929	916	613	863	800
3	28	51	32	150	90	116
4	3	1	2	81	32	38
5	0	0	0	49	9	20
6	0	0	0	42	2	5
7	0	0	0	28	1	3
8	0	0	0	37	3	3

TABLE V
COMPARISON RESULTS FOR EXAMPLE (20). THE SEQUENCES HAD 40 SAMPLES

$p = 2, a_{12} = -1.37, a_{22} = 0.56, \sigma^2 = 1, N = 40$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	0	5	39	0	0	29
1	136	307	373	10	39	272
2	840	651	573	647	872	623
3	21	36	15	114	60	53
4	2	1	0	79	13	9
5	1	0	0	46	10	7
6	0	0	0	41	4	3
7	0	0	0	25	2	1
8	0	0	0	38	0	3

TABLE VI
COMPARISON RESULTS FOR EXAMPLE (20). THE SEQUENCES HAD 80 SAMPLES

$p = 2, a_{12} = -1.37, a_{22} = 0.56, \sigma^2 = 1, N = 80$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	0	6	36	0	0	10
1	7	121	238	0	1	92
2	973	839	713	700	946	849
3	19	31	13	112	44	41
4	1	3	0	56	5	3
5	0	0	0	48	2	2
6	0	0	0	35	2	2
7	0	0	0	24	0	1
8	0	0	0	25	0	0

TABLE VII
COMPARISON RESULTS FOR EXAMPLE (21)

$p = 4, a_{14} = -1.352, a_{24} = 1.338, a_{34} = -0.662$
$a_{44} = 0.240, \sigma^2 = 1, N = 100$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	0	21	99	0	0	11
1	0	25	96	0	0	11
2	100	407	452	5	28	104
3	386	325	233	114	324	407
4	505	214	118	625	613	405
5	8	8	2	120	31	42
6	1	0	0	57	3	12
7	0	0	0	31	1	6
8	0	0	0	48	0	2

TABLE VIII
COMPARISON RESULTS FOR EXAMPLE (22)

$p = 4, a_{14} = -2.760, a_{24} = 3.809, a_{34} = -2.654$
$a_{44} = 0.924, \sigma^2 = 0.1, N = 20$

k	PDC	PLAV	PLS	AIC	MDL	PMDL
0	0	2	18	0	0	8
1	0	2	27	0	0	1
2	0	185	617	0	0	12
3	3	7	11	0	1	19
4	940	752	307	90	158	71
5	44	42	19	60	75	83
6	7	9	1	359	416	261
7	3	1	0	296	243	232
8	3	0	0	195	107	313

mance stems from the nature of the quasi-likelihood predictive density. This density does not precisely penalize for overparametrization since it does not take into account that the parameters of the model used to determine its form are not true, but rather are estimated from data.

As a final note we emphasize that the comparison between PLAV and PLS with the rest of the methods is not fair because the underlying assumptions for their use are different. To employ PLAV or PLS we do not require knowledge of the probability density function of the data, while for the other approaches this information is of fundamental importance.

V. CONCLUSION

A criterion for order selection of AR models has been derived based on Bayesian predictive densities. Its performance was assessed by extensive simulations and compared to other methods. The simulations show that this approach often yields better results than its competitors. Also, as an alternative to the PLS approach, the PLAV criterion was introduced. In the numerical simulations it systematically outperformed PLS.

REFERENCES

[1] J. Aitchison, "Goodness of prediction fit," *Biometrika*, vol. 62, pp. 547–554, 1975.
[2] J. Aitchison and I. R. Dunsmore, *Statistical Prediction Analysis*. New York: Cambridge University Press, 1975.
[3] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.
[4] P. Bloomfield and W. L. Steiger, *Least Absolute Deviations*. Boston: Birkhäuser, 1983.
[5] G. E. P. Box and J. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day, 1970.
[6] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, 1973.
[7] A. P. Dawid, "Present position and potential developments: Some personal views, statistical theory, the prequential approach," *J. Roy. Stat. Soc. A*, vol. 147, pp. 278–292, 1984.
[8] P. M. Djurić and S. M. Kay, "Predictive probability as a criterion for model selection," in *Proc. IEEE ICASSP* (Albuquerque, NM), 1990, pp. 2415–2418.
[9] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, pp. 829–867, 1982.
[10] S. Geisser and W. F. Eddy, "A predictive approach to model selection," *J. Amer. Stat. Ass.*, vol. 74, pp. 153–160, 1979.
[11] H. Jeffreys, *Theory of Probability*. New York: Oxford University Press, 1967.
[12] S. Kay, *Modern Spectral Estimation, Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[13] R. L. Kashyap, "Optimal choice of AR and MA parts in autoregressive moving average models," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 4, pp. 99–104, 1982.
[14] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–478, 1978.
[15] J. Rissanen, "Order estimation by accumulated prediction errors," *J. Appl. Prob.*, vol. 23A, pp. 55–61, 1986.
[16] J. Rissanen, "Stochastic complexity," (with discussion), *J. Roy. Stat. Soc. B*, vol. 49, pp. 223–239, and 252–265, 1987.
[17] G. Schwarz, "Estimating the dimension of the model," *Ann. Stat.*, vol. 6, pp. 461–464, 1978.
[18] M. Wax, "Order selection for AR models by predictive least squares," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, 1985.
[19] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley, 1971.

The Bispectrum of Complex Signals: Definitions and Properties

Ismail I. Jouny and Randolph L. Moses

***Abstract*—This correspondence is concerned with the definition and properties of the bispectrum of complex-valued signals. The symmetry**

Manuscript received February 7, 1991; revised August 27, 1991. This work was supported by the Department of the Navy, Office of Naval Research, under Contract N00014-80K-0202.

I. I. Jouny is with the Department of Electrical Engineering, Lafayette College, Easton, PA 18042.

R. L. Moses is with the Department of Electrical Engineering, Ohio State University, Columbus, OH 43210.

IEEE Log Number 9202797.