In other words, when $(u, u(-N), \cdots, u(-1))$ belongs to U, asymptotically periodic inputs produce asymptotically periodic outputs with the same period. The proof of this theorem makes use of a contraction-mapping fixed-point argument.

C. Quadratic Filters

The techniques used in our omitted proofs are also useful in connection with related problems that are "more nonlinear." In particular, related results are given in [9] for the discrete-time "quadratic filter" whose output $y(0), y(1), \cdots$ satisfies

$$y(n) = \sum_{i=0}^{N} a_i u(n-i) + \sum_{i=1}^{N} b_i y(n-i) + \sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j} y(n-i)y(n-j), \qquad n \geq 0 \tag{3}$$

in which the a_i, b_i, and $c_{i,j}$ are real coefficients, $u(0), u(1), \cdots$ is the input sequence, $y(-N), \cdots, y(-1)$ and $u(-N), \cdots, u(-1)$ are initial values, and N is a positive integer. The initial values and the elements of the input and output sequences are real numbers, as in (1). In [10], conditions are presented under which bounded inputs to quadratic filters produce bounded outputs.[4] There too, it is assumed that the initial values of the output are zero. In [9], we show that the condition concerning initial values is not needed in the sense that small values of the magnitudes of the initial conditions can be accommodated by making a small reduction in the bound on the allowed inputs. More significantly, in [9], we give simple conditions (on the coefficients, input, and initial values) under which (3) has the additional stability properties that ii)–iv) of Section I are met.

REFERENCES

[1] C. Bruni, G. DiPillo, and G. Koch, "Bilinear systems: An appealing class of 'Nearly Linear' systems in theory and applications," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 334–348, Aug. 1974.

[2] R. R. Mohler and W. J. Kolodziej, "An overview of bilinear system theory and applications," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-10, pp. 683–688, Oct. 1980.

[3] W. J. Rugh, *Nonlinear System Theory*. Baltimore, MD: The John Hopkins Univ. Press, 1982.

[4] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984, p. 39.

[5] J. Lee and V. J. Mathews, "A stability condition for certain bilinear systems," *IEEE Trans. Signal Processing*, vol. 42, pp. 1871–1873, July 1994.

[6] I. W. Sandberg, "Some results on the theory of physical systems governed by nonlinear functional equations," *Bell Syst. Tech. J.*, vol. 44, no. 5, pp. 871–898, May/June 1965.

[7] A. N. Michel, "Stability: The common thread in the evolution of feedback control," *IEEE Contr. Syst.*, vol. 16, pp. 50–60, June 1996.

[8] K. K. Johnson and I. W. Sandberg, "On the stability of bilinear filters," *J. Circuits, Syst., Signal Process.*, to be published.

[9] ——, "The stability of quadratic filters," *Int. J. Circuit Theory Applicat.*, vol. 26, pp. 191–197, 1998.

[10] A. Carini and E. Mumolo, "Adaptive stabilization of recursive second order polynomial filters by means of a stability test," in *Proc. 1995 IEEE Workshop Nonlinear Signal Image Process.*, Neos Marmaras, Greece, June 1995, pp. 939–942.

[4] There is a difference between (3) and the model in [10]. There, $a_i = 0$ for $i > 1$. We have added the additional terms because their presence leads to a more useful filter.

Autoregressive Model Order Selection by a Finite Sample Estimator for the Kullback–Leibler Discrepancy

P. M. T. Broersen and H. E. Wensink

***Abstract*—The finite sample information criterion (FSIC) is introduced as an estimator for the Kullback–Leibler discrepancy of an autoregressive time series. It is derived especially for order selection in finite samples, where model orders are greater than one tenth of the sample size. It uses a theoretical expression for the ratio between the squared prediction error and the residual variance as the penalty factor for additional parameters in a model. This ratio can be found with the finite sample theory for autoregressive estimation, which is based on empirical approximations for the variance of parameters. It takes into account the different number of degrees of freedom that are available effectively in the various algorithms for autoregressive parameter estimation. The performance of FSIC has been compared with existing order selection criteria in simulation experiments using four different estimation methods. In finite samples, the FSIC selects model orders with a better objective quality for all estimation methods.**

I. INTRODUCTION

Model order selection has been a topic in time series for more than two decades. The problem is best characterized by the difference between the residual variance and the prediction error. The residual variance is minimized in estimation, and it will always decrease when more parameters are estimated, whereas the unknown prediction error reaches a minimum when all significant parameters have been included in the model, and it will increase for higher model orders. Many criteria for order selection are transformations of the residual variance: AIC [1], consistent criteria [2], [3], and finite sample equivalents for those criteria [4]. The minimum description length (MDL) criterion is equivalent with some consistent criteria [5]. Finite sample criteria are necessary if the ratio p/N of model order and sample size is greater than 0.1 because for those model orders, the empirical statistics of the AR parameters and model fit depend on the estimation method. Another type of selection criterion that is adapted to the estimation statistics is the predictive least squares (PLS) criterion [6], which calculates an estimate for the true prediction error by using only estimated models based on the past observed values. However, that method of separating data for estimation and for prediction requires many observations. It loses its accuracy for finite samples, and it cannot be used if the maximum order is as high as $N/2$. Small sample selection results have also given rise to a corrected AIC criterion: AIC_C [7]. The correction term is an asymptotical bias term in the derivation of the Kullback–Leibler discrepancy.

In this correspondence, the Kullback–Leibler concept is combined with the finite sample theory to define the finite sample information criterion (FSIC). It uses an approximation for the ratio of prediction error and residual variance. A study of the penalty as a function of the model order reveals that FSIC is very similar to AIC for Yule–Walker estimates and similar to AIC_C for Burg estimates. Typical simulation examples will show the favorable properties of FSIC.

Manuscript received July 25, 1996; revised December 18, 1997. The associate editor coordinating the review of this paper and approving it for publication was Prof. Moeness Amin.

P. M. T. Broersen is with the Department of Applied Physics, Delft University of Technology, Delft, The Netherlands.

H. E. Wensink is with Hollandse Signaalapparaten, B.V., Hengelo, The Netherlands.

Publisher Item Identifier S 1053-587X(98)04437-7.

II. Finite Sample Theory

An autoregressive process of order K is given by the parameters $1, a_1, \cdots a_K$ and independent, identically distributed zero mean innovations ε_n with variance σ_ε^2 and with finite fourth-order moments. An AR(p) model of an arbitrary order p can be fitted to the data generated by the AR(K) process; the estimates $\hat{\boldsymbol{a}}_{\boldsymbol{i}}$ constitute the $(p+1) \times 1$ parameter vector $\hat{\boldsymbol{a}}_{\boldsymbol{p}}$. The residual variance $S^2(p)$ is defined as the mean square fit of the model to the data from which the parameters have been estimated. The forecasting quality of a model is expressed by the squared error of prediction PE(p). The value of $S^2(p)$ decreases for each parameter included in the model, whereas the PE(p) of a model will only improve as long as the extra included parameters are significant.

In finite samples, where interesting model orders are greater than about $N/10$, estimation results turn out to depend on the estimation method [4]. Basic elements are the finite sample variance coefficients $v_{i,\cdot}$, where i denotes the model order, and the dot indicates one of four estimation methods:

1) Yule–Walker (YW) method;
2) method of Burg;
3) least squares method that minimizes forward and backward residuals (LSFB);
4) least squares method that minimizes forward residuals only (LSF).

The $v_{i,\cdot}$ replace the asymptotical variance $1/N$ for parameters estimated in a white noise process. They can be considered as expressions for the degrees of freedom as a function of the model order i and of the estimation method and are given by

$$\begin{aligned} v_{i,\mathrm{YW}} &= (N-i)/N(N+2) \\ v_{i,\mathrm{BURG}} &= 1/(N+1-i) \\ v_{i,\mathrm{SFB}} &= 1/(N+1.5-1.5i) \\ v_{i,\mathrm{LSF}} &= 1/(N+2-2i). \end{aligned} \tag{1}$$

For all methods, $v_{0,\cdot}$ is defined to be $1/N$ if the mean of the observations is subtracted; otherwise, it is zero. The variance coefficients $v_{i,\cdot}$ are approximations for the variance of the last parameter in an AR(i) model estimated in a white noise process. Likewise, they are good approximations for the variance of the last parameter in all AR processes above the true order. For i much smaller than N, all $v_{i,\cdot}$ are about $1/N$, which is the single value that is found in the asymptotical theory for all parameters and all estimation methods.

Simulation results and theoretical support has been given for the main formulae of the finite sample theory [4] for orders $p \geq K$

$$\begin{aligned} E\{S^2(p)\} &= \sigma_\varepsilon^2 \prod_{i=0}^{p}(1-v_{i,\cdot}) \\ E\{\mathrm{PE}(p)\} &= \sigma_\varepsilon^2 \prod_{i=0}^{2}(1+v_{i,\cdot}) \end{aligned} \tag{2}$$

where for $v_{i,\cdot}$ the expressions defined in (1) are substituted, depending on the estimation method. Those multiplicative formulae give an accurate description of the residual variance and of the prediction error for order K and higher in simulations of autoregressive processes. The accuracy of (2) is very good if the time delay with significant correlation is not greater than about half the observation length N. The products in (2) reduce to the values of the asymptotical theory $1-p/N$ and $1+p/N$, respectively, if p is less than $N/10$.

III. The Finite Sample Information Criterion FSIC(p)

Elements of information theory have been used by Akaike to derive the order selection criterion AIC. A consistent criterion with penalty $\log N$ [2], which is equal to MDL [5], and a minimal consistent criterion with penalty $2 \log\log N$ [3] followed. Those asymptotically based criteria can be described together as a generalized information criterion GIC(p, α) with α as penalty factor

$$\mathrm{GIC}(p,\alpha) = \log S^2(p) + \alpha \frac{p}{N}. \tag{3}$$

Another correction to the AIC criterion with $\alpha = 2$ is AIC_C, which has been derived from the asymptotical χ^2 distribution and gives $\alpha = 2N/(N-p-2)$ [7]. A generalized form of AIC_C is found directly from the correspondence between the definitions of Kullback–Leibler information and maximum likelihood estimation for Gaussian processes [8]. It utilizes a relation between the log likelihood L_x of data x_n that are used to estimate the parameters and the log likelihood L_y of an independent realization y_n with the same statistical properties. The Kullback–Leibler discrepancy Δ is defined as the expectation of $-2 \log L_y(\hat{\boldsymbol{a}}, \hat{\sigma}^2)$, where the expectation is taken with respect to the true values of σ^2 and the parameter vector. As the independent realization will not be available, an estimate of Δ will be made with the given data x_n. Suppose that N normally distributed data $x_1, \cdots, x_N$ are available with the estimate $\hat{a}$ for the parameters and $\hat{\sigma}^2$ for the residual variance. Then, the relation between the likelihood of y_n and x_n is given by [8]

$$\begin{aligned} &-2 \log L_y(\hat{a}, \hat{\sigma}^2) \\ &\quad = -2 \log L_x(\hat{a}, \hat{\sigma}^2) + \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 / \hat{\sigma}^2 - N. \end{aligned} \tag{4}$$

This relation is valid for orders K and higher because true parameters zero can be added beyond order K. The first term on the right-hand side can be written as

$$\begin{aligned} -2 \log L_x(\hat{a}, \hat{\sigma}^2) &= N \log 2\pi + N \log \hat{\sigma}^2 \\ &\quad + \sum_{n=1}^{N} (x_n - \hat{x}_n)^2 / \hat{\sigma}^2 \\ &= N \log 2\pi + N \log \hat{\sigma}^2 + N. \end{aligned}$$

This term can be approximated by substituting the residual variance $S^2(K)$ for $\hat{\sigma}^2$. A new approximation for the expectation for the second right-hand term in (4) can be found with the finite sample theory. Using only the first term of a Taylor expansion approximates the expectation of a quotient of two stochastic variables by the quotient of their expectations. Hence, the numerator becomes the expectation of the prediction error, and the denominator yields the residual variance; therefore, substitution of (2) gives

$$E\left\{ \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 / \hat{\sigma}^2 \right\} \approx N \prod_{i=0}^{K} \frac{1+v_{i,\cdot}}{1-v_{i,\cdot}}.$$

Together, the expectation of (4) can be approximated with

$$\begin{aligned} &E\{-2 \log L_y(\hat{a}, \hat{\sigma}^2)\} \\ &\quad \approx N \log 2\pi + N \log S^2(k) + N \prod_{i=0}^{K} \frac{1+v_{i,\cdot}}{1-v_{i,\cdot}}. \end{aligned}$$

Subtracting from this estimated Kullback–Leibler discrepancy the constants $N \log 2\pi + N$, dividing by N, and generalizing to an arbitrary model order p, we define the finite sample information criterion for autoregressive order selection as:

$$\mathrm{FSIC}(p) = \log S^2(p) + \prod_{i=0}^{p} \frac{1+v_{i,\cdot}}{1-v_{i,\cdot}} - 1. \tag{5}$$

By subtracting 1 in (5), an asymptotical series expansion of FSIC approaches AIC for $N \to \infty$ and $p < N/10$.

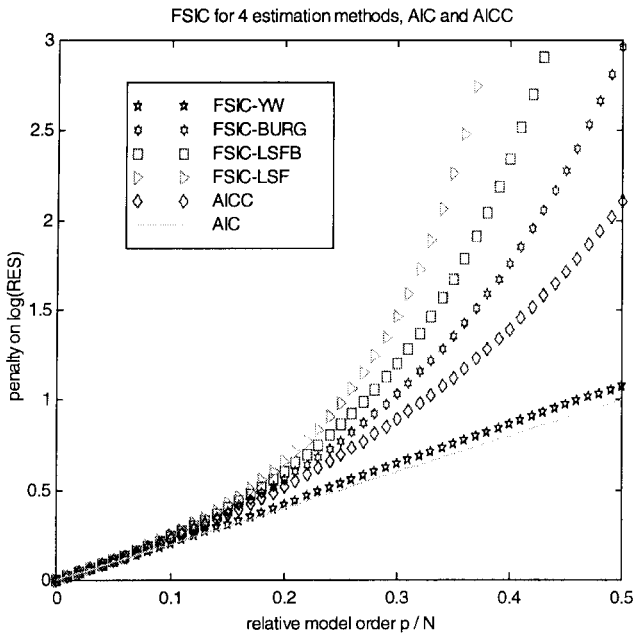

Fig. 1. Penalty for FSIC with four estimation methods and for AIC and AIC_C, as a function of the relative model order p/N.

IV. Simulation Results and Discussion

A first investigation of the properties of FSIC can be made by comparing the penalty $2p/N$ of AIC with those of FSIC for the four estimation methods for which the $v_{i,\cdot}$ are given in (1). The penalty in Fig. 1 is defined as what is added to log $S^2(p)$ in the selection criterion; therefore, $\alpha p/N$ for GIC(p,α) in (3). It has been computed for $N = 100$, but it is only a function of p/N effectively. It is seen that the penalty of FSIC for the Yule–Walker method is very similar to that of AIC. The penalty function determines which order will be selected; therefore, the performance of AIC and FSIC will be similar for Yule–Walker estimation. Likewise, FSIC has some similarity with AIC_C for the Burg method. For the least squares methods LSFB and LSF, higher penalty functions are found. It is clear that the probability of selecting a very high order will be much greater for AIC than for the criteria FSIC and AIC_C that have been derived from the finite sample Kullback–Leibler information. The higher penalties for increasing model orders of the latter methods create an insensitivity of the selection result for the maximum candidate order for selection, whereas GIC(p,α) with penalty αp is sensitive for that maximum, at least for values of α in the range from 1–5. All six penalty functions in Fig. 1 are very close for relative orders p/N less than 0.1. Differences are moderate for relative orders less than 0.2 and deviate more for still higher orders. The performance of FSIC, AIC_C, and AIC will almost be the same if the maximum candidate model order is less than about $N/10$. According to the finite sample formulae, the difference between FSIC and AIC is the greatest for the LSF estimation method.

A quality measure for evaluation of the difference between criteria is a normalized prediction error, which is denoted the model error ME(p) and defined as [9]

$$\mathrm{ME}(p) = N\left\{\frac{\mathrm{PE}(p)}{\sigma_\varepsilon^2} - 1\right\} = N\,\frac{\hat{\boldsymbol{a}}_p^T R(p)\hat{\boldsymbol{a}}_p - \sigma_\varepsilon^2}{\sigma_\varepsilon^2}. \tag{6}$$

$R(p)$ is the $(p+1)\times(p+1)$ submatrix of the infinite dimensional Toeplitz matrix R that describes the true covariance of the process that generates the data and is known only in simulations. The $(p+1)\times 1$ parameter vector $\hat{\boldsymbol{a}}_p$ is estimated in each run. The multiplication with N gives the asymptotical value $\mathrm{ME}(p) = p$ for $p \geq K$ independent of the sample size N.

TABLE I
Percentage of Selected True Order K and Average Model Error $ME(p)$ with Different Selection Criteria for an $AR(2)$ Process with $a_1 = 0.56$ and $a_2 = 0.4, N = 30, L = 15$, Average of 5000 Simulation Runs

	YW		BURG		LSFB		LSF	
	ME	% K	ME	%K	ME	%K	ME	%K
AIC=GIC(p,2)	3.7	61.4	15.6	39.8	31.3	23.5	161.7	5.5
FIC(p,2)	4.1	57.4	8.6	52.0	10.5	46.8	81.4	30.1
GIC(p,2loglogN)	3.8	62.0	10.9	50.3	23.5	35.4	155.6	9.0
MDL=GIC(p,logN)	4.7	54.1	5.9	56.4	11.5	49.8	128.8	18.1
AIC_C(p)	3.9	59.4	4.4	59.4	8.4	55.6	32.4	37.5
FSIC(p)	3.6	61.8	4.3	61.9	4.4	58.1	5.0	53.6

Table I gives a comparison between FSIC, AIC_C, and GIC(p,α) for $N = 30$ in an AR(2) process with $L = N/2$ as the maximum candidate order. The first column for every estimation method gives the model error ME as a measure for the predictive accuracy of the selected model. The second column gives the percentage of simulation runs $\%K$ in which the true order of the data generating process was selected. In both columns, the best result for each method is always found with FSIC, which demonstrates that the good quality of FSIC is established with any sensible way of comparison. As could be expected from Fig. 1, AIC is second best for Yule–Walker, and AIC_C is second best for Burg. This conclusion has also been found in simulations for many different processes with different numbers of observations. The differences between FSIC and the other methods become smaller if the maximum candidate order L is made less than 15 in this example and disappear almost completely for L equal to 2 or 3. The result for FSIC remains almost the same by taking as maximum order 8 instead of 15. AIC_C for the Burg method was also not sensitive for the maximum candidate order. Yule–Walker estimates are only slightly sensitive to that maximum order because the finite sample variance coefficients (1) are smaller than $1/N$. FIC$(p,2)$ is slightly worse than the asymptotical AIC or GIC$(p,2)$ for the Yule–Walker method but much better for all other methods. The same pattern is found in a comparison of GIC(p,α) with FIC(p,α) for different values for the penalty α like 3, 4, $2\log\log N$ or $\log N$, which is 3.40 for $N = 30$. Therefore, the finite sample criteria [4] improve the performance in comparison with their asymptotical equivalents, but they remain sensitive to the highest candidate order for selection. Table I gives the methods in the sequence of increasing $v_{i,\cdot}$ and decreasing quality. All criteria GIC(p,α) gave for all methods a worse result than FSIC if higher maximum orders are candidate for selection. Therefore, FSIC has some advantages. The finite sample selection result is the best for each estimation method, and it is independent of the maximum candidate order that is considered, which makes the selection result independent of the experimenter's prejudice.

Of course, a maximum candidate order can be prescribed for computational reasons. Although FSIC requires no maximum order, specific limits exist for each method of estimation. It is, as an example, impossible to estimate more than $N/2$ parameters from N observations with the LSF method. A simple and practical way to deal with those limits is to restrict order selection to those candidate orders for which $v_{i,\cdot}$ is less than 0.25. This value for the finite sample variance coefficient agrees with a standard deviation of 0.5 for the last parameter, which has a possible range of -1 to $+1$. Estimating parameters with so much *a priori* inaccuracy will not contribute significantly to the accurate description of the process.

Table II gives a comparison of ME for models selected with AIC, MDL, and FSIC for a number of different AR(4) processes that have been generated with four reflection coefficients β that are equal in magnitude; the parameters for the AR(4) processes are determined with the Levinson recursion. For high negative values of β, the

TABLE II
$ME(p)$ FOR SELECTION WITH AIC, MDL, AND FSIC FOR YULE–WALKER, BURG, LSFB, AND LSF ESTIMATES AS A FUNCTION OF β. $N = 30, L = 15$, AVERAGE OF 2000 RUNS OF AR(4) PROCESSES WITH PARAMETERS COMPUTED FROM THE REFLECTION COEFFICIENTS $-\beta, \beta, -\beta, \beta$

	YW			BURG			LSFB			LSF		
β	AIC	MDL	FSIC	AIC	MDL	FSIC	AIC	MDL	FSIC	AIC	MDL	FSIC
-0.8	300.4	386.2	300.5	103.3	85.3	73.3	214.3	91.1	66.2	712.1	626.9	80.1
-0.6	36.7	53.0	36.5	35.0	27.5	20.7	57.8	38.4	22.5	305.6	273.7	28.2
-0.4	14.5	19.9	14.2	23.3	17.4	13.5	40.3	24.3	13.4	225.0	189.6	15.0
-0.2	5.8	5.8	5.8	16.4	7.3	6.5	32.2	12.3	6.5	166.9	134.3	7.2
0	1.3	0.4	1.5	12.3	1.4	2.0	28.5	5.9	2.0	188.5	138.7	2.6
0.2	5.5	5.6	5.4	16.5	7.1	6.4	32.6	12.3	6.3	153.5	118.4	6.7
0.4	8.1	11.6	8.0	18.7	10.8	7.8	35.2	18.5	8.0	175.6	151.7	9.3
0.6	21.2	25.6	21.1	19.7	10.6	8.0	38.3	19.4	7.9	253.7	228.3	9.9
0.8	179.6	194.0	179.5	22.3	12.1	10.2	51.4	26.5	9.4	854.8	807.1	14.8

effective correlation length is much longer than 30 observations; therefore, the accuracy of the formulae (2) for residual variance and prediction error will be poor. As a consequence, for $\beta = -0.8$, models of the true order 4 also have a high ME: 260.3, 60.1, 60.5, and 71,1 for Yule–Walker, Burg, LSFB, and LSF, respectively. Hence, the quality of selected models also has to be poor. However, even in those difficult circumstances, the performance of FSIC remains good in comparison with AIC and MDL. The difference between the criteria is smallest for Yule–Walker estimates and most important for LSF estimates, as can be understood with the differences in the penalty function of Fig. 1. Many more processes and selection criteria have been included in the simulations, but the performance of FSIC over a range of different processes, different sample sizes, and various estimation methods is mostly the best and otherwise close to that.

Artifacts are found if the best model order is near zero or near the highest candidate order L or if the last parameter value is very significant. Order zero is best in the AR(4) processes of Table II for approximately $-0.19 < \beta < 0.19$ because for those processes, E[ME(0)] is less than E[ME(4)], which is four, asymptotically. Taking GIC(p, ∞) would give the best result then. The MDL columns in Table II show that GIC$(p, \log N)$ has a higher ME value than FSIC for all values $|\beta| \geq 0.2$. On the other side of the range, the lowest penalty functions give the best results if $E[\mathrm{ME}(p)]$ has the minimum value for the highest order that is candidate for selection, because GIC$(p, 0)$ with zero penalty will always select the highest order. Finally, if the last parameter is very significant, the highest penalty factor that does not lead to an underfitted model might be the best because it reduces the risk of overfit [9]. Apart from these artifacts, the performance of FSIC was always, for all estimation methods, all sample sizes, all our simulated processes, and all our maximum candidate orders, better than or the same as the performance of all GIC(p, α) for every value of α.

V. CONCLUDING REMARKS

The single order selection criterion FSIC performs the best for all four autoregressive estimation methods. FSIC is almost equal to AIC for Yule–Walker estimates and to AIC$_C$ for Burg. The finite sample theory extends the good performance of those two combinations of selection criterion and estimation method to the LSF and LSFB estimation methods as well, for which no particularly adapted criterion had been reported. The finite sample theory clarifies why an accurate estimate for the Kullback–Leibler discrepancy depends on the estimation method. Moreover, it gives some theoretical basis for the good performance of FSIC.

REFERENCES

[1] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inform. Theory*, B. N. Petrov and F. Csaki, Eds, Budapest, Hungary, 1973, pp. 267–281.
[2] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.
[3] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *J. R. Statist. Soc.*, vol. B 41, pp. 190–195, 1979.
[4] P. M. T. Broersen and H. E. Wensink, "On finite sample theory for autoregressive model order selection," *IEEE Trans. Signal Processing*, vol. 41, pp. 194–204, 1993.
[5] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
[6] A. K. Nandi and J. A. Chambers, "New lattice realization of the predictive least-squares order selection criterion," *Proc. Inst. Elect. Eng, F*, vol. 138, pp. 545–550, 1991.
[7] C. M. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, pp. 297–307, 1989.
[8] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1991.
[9] P. M. T. Broersen and H. E. Wensink, "On the penalty factor for autoregressive order selection in finite samples," *IEEE Trans. Signal Processing*, vol. 44, pp. 748–752, Mar. 1996.

Azimuth and Elevation Direction Finding Using Arbitrary Array Geometries

Tsung-Hsien Liu and Jerry M. Mendel

***Abstract*— An ESPRIT-based algorithm is proposed to estimate the azimuth and elevation angles of multiple independent sources. The algorithm uses cumulants and imposes no geometric constraint on the array. Only one third of the hardware is needed for our algorithm as compared with covariance-based two-dimensional (2-D) ESPRIT. Our algorithm can estimate azimuth and elevation angles of $M - 1$ sources using M sensors. Simulation results show that for several array configurations, our algorithm works well.**

***Index Terms*—Azimuth, cumulant, direction of arrival, elevation, ESPRIT, VESPA.**

Manuscript received November 11, 1996; revised July 24, 1997. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jonathon A. Chambers.

The authors are with the Signal and Image Processing Institute, Department of Electrical Engineering—Systems, University of Southern California, Los Angeles, CA 90089-2564 USA.

Publisher Item Identifier S 1053-587X(98)04440-7.