Order Selection for AR Models by Predictive Least Squares

MATI WAX

Abstract—**We present a new criterion for selecting the order of AR models which, unlike the existing criteria, is amenable to on-line or adaptive operation. It is based on the predictive least squares (PLS) principle and is implemented in a computationally efficient way by predictive lattice filters. We prove the consistency of the criterion and demonstrate its performance by computer simulations.**

I. Introduction

THE problem of fitting an autoregressive (AR) model to an observed data sequence arises in a variety of applications ranging from adaptive control to speech modeling and synthesis, equalization of communication channels, and spectral estimation. The problem is formulated as follows. Given a set of samples from a discrete time process $\{ y_t, 0 \leqq t \leqq n-1 \}$, we want to find the set of coefficients $\{ a_i \}$ that gives the *best linear predictor* of y_n from the past samples, which we denote by $y_{n|n-1}$:

$$y_{n|n-1} = \sum_{i=1}^{p} a_i y_{n-i}. \tag{1}$$

The crucial part of this problem is to determine the order of the model, namely the value of p. The popular criteria for order selection are the AIC proposed by Akaike [1] and the MDL proposed by Rissanen [10] and Schwarz [12]. Both criteria are based upon asymptotic results. Thus, strictly speaking, their applicability is justified only for large samples. Moreover, the AIC is based on the assumption that the data are Gaussian, while the MDL is based on modeling the data as Gaussian. In both cases, the applicability of the criteria to non-Gaussian data is questionable. Another deficiency of these criteria is their two-pass nature; their evaluation requires two passes over the data, and hence they are not amenable to on-line or adaptive operation.

In this paper, we present a novel criterion for order-selection which is based on the predictive least squares (PLS) principle (Rissanen [11]) and is free of the above deficiencies. It is equally well applicable to all sample sizes, large and small, to any distribution, Gaussian or other, and is a one-pass scheme. The criterion selects the order as the one for which the sum of the *true prediction errors* along the sequence is minimized. By true, we mean "honest" in the sense that the predictor coefficients are estimated only from the *past* data. Hence, the selection of the order according to this principle amounts to computing the sum of the true prediction errors for different orders and choosing the one that yields the minimum.

To compute the PLS order selection criterion, we use predictive lattice filters. These filters, introduced by Morf *et al.* [9], have many desirable properties, the key ones being their computational efficiency and their modular structure which is very amenable to VLSI implementation.

Manuscript received February 3, 1987; revised October 9, 1987.
The author is with RAFAEL, Haifa 31021, Israel.
IEEE Log Number 8719020.

II. The PLS Order Selection Criterion

The central notion in the predictive least squares (PLS) principle (Rissanen [11]) is that of the *true prediction error*. For the case of AR models, the true prediction error is defined as the error in predicting y_t from the past samples $\{ y_0, \cdots, y_{t-1} \}$. Thus, considering an AR model of order k, the true prediction error is given by

$$\hat{e}_{k,t} = y_t - \sum_{i=1}^{k} \hat{a}_{k,i,t} y_{t-i} \tag{2}$$

where $\{ \hat{a}_{k,i,t} \}$ are the predictor coefficients determined from the *past* data $\{ y_0, \cdots, y_{t-1} \}$ by the conventional least squares criterion

$$\min_{a_1, \cdots, a_k} \sum_{i=1}^{t-1} \bar{e}_{k,i}^2 \tag{3a}$$

with

$$\bar{e}_{k,i} = y_i - \sum_{j=1}^{k} a_j y_{i-j}. \tag{3b}$$

Notice that $\bar{e}_{k,i}$ is the conventional residual of y_i obtained by a predictor whose coefficients are computed from *all* the data $\{ y_0, \cdots, y_{t-1} \}$, in contrast to the true prediction error $\hat{e}_{k,i}$ where only the *past* data $\{ y_0, \cdots, y_{i-1} \}$ are used.

The PLS criterion asserts that the order should be selected as that for which the sum of the true prediction errors along the sequence is *minimized*. Restricting the maximal order to M, this gives the procedure

$$\min_{k \leq M} \text{PLS}\ (k) \tag{4a}$$

where

$$\text{PLS}\ (k) = \frac{1}{n} \sum_{t=1}^{n} \hat{e}_{k,t}^2. \tag{4b}$$

This criterion has a very strong intuitive appeal: the order is selected as that which has yielded the *best* predictions along the sequence.

III. Implementation by Predictive Lattice Filters

To compute the PLS criterion, we have to recompute the true prediction errors for every order and for every sample point along the sequence. Thus, for order k and sample point t, we have to first compute the predictor coefficients $\{\hat{a}_{k,j,t}\}$ from (3), and then compute the prediction error $\{\hat{e}_{k,t}\}$ from (2). The solution of (3) is given by the well-known formula

$$\hat{\boldsymbol{a}}_{k,t} = \left(\sum_{i=1}^{t} \boldsymbol{y}_{k,i} \boldsymbol{y}_{k,i}^T \right)^{-1} \sum_{i=1}^{t} \boldsymbol{y}_{k,i} y_i \tag{5a}$$

where $\hat{\boldsymbol{a}}_{k,t}$ is the $k \times 1$ vector

$$\hat{\boldsymbol{a}}_{k,t} = [\hat{a}_{k,1,t}, \cdots, \hat{a}_{k,k,t}]^T \tag{5b}$$

and $\boldsymbol{y}_{k,i}$ is the $k \times 1$ vector

$$\boldsymbol{y}_{k,i} = [y_{i-1}, \cdots, y_{i-1-k}]^T. \tag{5c}$$

Performing the operations implied by (4) and (5) for every order and every sample point along the sequence is computationally expensive. Fortunately, it is possible to exploit the structure of the problem to reduce substantially the computational burden. Indeed, the so-called lattice filters (Morf *et al.* [9], Lee *et al.* [7], Friedlander [4], [5], Lev-Ari *et al.* [8], and Cioffi and Kailath [3]) yield a computationally efficient solution to this problem; the kth-order prediction error for the sample point t is obtained by very simple update formulas from the $(k-1)$th prediction error at time $t-1$. Assuming, for simplicity, that the data prior to $t = 0$ is zero, referred to as the prewindowed case, and using the notation of Friedlander [4], we get the following set of recursions for the prediction errors to be performed at every time instant $t = 0, \cdots, n-1$.

Initialization:

$$e_{0,t} = r_{0,t} = y_t \tag{6a}$$

$$R_{0,t}^e = R_{0,t}^r = R_{0,t-1}^e + y_t^2 \tag{6b}$$

$$\gamma_{-1,t} = 1. \tag{6c}$$

Main loop, to be done for $k = 0$ to $\min\{M, t\} - 1$:

$$\Delta_{k+1,t} = \Delta_{k+1,t-1} + e_{k,t} r_{k,t-1} / \gamma_{k-1,t-1} \tag{7a}$$

$$\gamma_{k,t} = \gamma_{k-1,t} - r_{k,t}^2 / R_{k,t}^r \tag{7b}$$

$$e_{k+1,t} = e_{k,t} - \Delta_{k+1,t} r_{k,t-1} / R_{k,t-1}^r \tag{7c}$$

$$R_{k+1,t}^e = R_{k,t}^e - \Delta_{k+1,t}^2 / R_{k,t-1}^r \tag{7d}$$

$$r_{k+1,t} = r_{k,t-1} - \Delta_{k+1,t} e_{k,t} / R_{k,t}^e \tag{7e}$$

$$R_{k+1,t}^r = R_{k,t-1}^r - \Delta_{k+1,t}^2 / R_{k,t}^e \tag{7f}$$

$$\hat{e}_{k+1,t} = e_{k+1,t} / \gamma_{k,t-1}. \tag{7g}$$

Note that the equations, except (7g), are the conventional prewindowed recursions (Friedlander [5]). Relation (7g) relates the true and filtered prediction errors (Lev-Ari *et al.* [8], Cioffi and Kailath [3]). Formulas (6)–(7) constitute a complete set of recursions for both order and time update of the prediction errors, thus enabling efficient computation of the PLS criterion (4).

IV. Consistency

The PLS criterion we have presented above was motivated on intuitive grounds. The question remains whether this criterion yields a consistent estimator. That is, assuming the data to be generated by an autoregressive model of order p, the question is whether the criterion yields the correct order p with probability that converges to 1 as the length of the data grows to infinity. In what follows, we prove that the PLS criterion is indeed consistent.

First we introduce some notations. Let

$$|\boldsymbol{b}| = (\boldsymbol{b}^T \boldsymbol{b})^{1/2}$$

be the Euclidean norm of the vector b, and let

$$|\boldsymbol{B}| = \sup_{|\boldsymbol{b}|<1} (\boldsymbol{b}^T \boldsymbol{B}^T \boldsymbol{B} \boldsymbol{b})^{1/2}$$

be the matrix norm of $\boldsymbol{B}$.

To prove the consistency, we make the following three assumptions.

A1: The sampled data $\{y_t\}$ are generated by the pth-order univariate AR process

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + e_t \tag{8}$$

where $\{e_t\}$, referred to as the innovations, are uncorrelated $(0, \sigma^2)$ random variables, with finite eighth moment.

A2: The roots of the characteristic equation

$$z^m - a_1 z^{m-1} - \cdots - a_m = 0$$

are all less than unity in absolute value.

A3: For all t sufficiently large ($> t_0$, say)

$$|\boldsymbol{C}_{k,t} \boldsymbol{C}_k^{-1} - \boldsymbol{I}_k| < 1$$

where $\boldsymbol{C}_{k,t}$ is the sample-covariance matrix

$$\boldsymbol{C}_{k,t} = \frac{1}{t} \sum_{i=1}^{t} \boldsymbol{y}_{k,i} \boldsymbol{y}_{k,i}^T \tag{9}$$

and $\boldsymbol{C}_k$ is the covariance matrix

$$\boldsymbol{C}_k = E[\boldsymbol{y}_{k,t} \boldsymbol{y}_{k,t}^T]. \tag{10}$$

Note that A1 does not say anything about the distribution of the innovations; specifically, they need not be Gaussian. Note also that A2 is the conventional assumption for stationary AR processes. Assumption A3 is identical to that made by Bhansali [2].

The proof of the consistency of the PLS criterion is rather elaborate and involves several steps. Following Rissanen [11], observe that the value of k that minimizes the PLS criterion is identical to the value of k that minimizes a modified criterion, defined as

$$\text{MPLS}(k) = \frac{1}{n} \sum_{t=1}^{n} (\hat{e}_{k,t}^2 - e_t^2). \tag{11}$$

It turns out that it is substantially simpler to prove the consistency of the MPLS criterion rather than that of the PLS criterion. In what follows, we state several results regarding the mean and variance of the MPLS criterion which we need in order to establish the consistency. A notion central in these results is that of *order of convergence*: we say that $\delta_n = 0(\phi_n)$ if for some positive number M, $|\delta_n| \leqq M|\phi_n|$ for every n, where $|\ |$ denotes the absolute value.

Theorem 1: Under assumptions A1–A3 and for $k \geqq p$,

$$E[\text{MPLS}(k)] = \frac{1}{n} \sum_{t=1}^{n} \frac{k}{t} + 0\left(\frac{1}{n}\right).$$

Proof: See Appendix B.

Corollary 1.1: Under assumptions A1–A3, for every $\delta > 0$ and for $k \geqq p$, then for n large enough

$$(k - \delta)\frac{\log n}{n} \leqq E[\text{MPLS}(k)] \leqq (k + \delta)\frac{\log n}{n}.$$

Proof: The proof follows immediately from Theorem 1.

Corollary 1.2: Under assumptions A1–A3, for $k \geqq p$ and n large enough,

$$E[\text{MPLS}(k)] - E[\text{MPLS}(p)] \geqq (k - p)\frac{\log n}{n}.$$

Proof: The proof follows immediately from Corollary 1.1.

Theorem 2: Under assumptions A1–A3 and for $k < p$,

$$E[\text{MPLS}(k)] = \alpha_k + 0\left(\frac{1}{n^{1/2}}\right)$$

where

$$\alpha_k = \text{tr}\,[\boldsymbol{D}_k \boldsymbol{a}_p \boldsymbol{a}_p^T \boldsymbol{D}_k \boldsymbol{C}_p]$$

with tr [ ] denoting the trace of the bracketed matrice, D_k denoting the $p \times p$ diagonal matrix consisting of all ones except for the first k elements

$$\boldsymbol{D}_k = \text{diag}\,(0, \cdots, 0, 1, \cdots, 1),$$

$\boldsymbol{a}_p$ denoting the $p \times 1$ vector of the AR coefficients

$$\boldsymbol{a}_p = (a_1, \cdots, a_p)^T,$$

and $\boldsymbol{C}_p$ denoting the $p \times p$ covariance matrix defined in (10).

Proof: See Appendix C.

Corollary 2.1: For any $\delta > 0$ and for large enough n,

$$E[\text{MPLS}(k)] - E[\text{MPLS}(p)] \geqq \alpha_k - \delta.$$

Proof: The proof follows immediately from Theorem 2.

Theorem 3: Under assumptions A1–A3 and for $k \geqq p$,

$$\text{Var}\,[\text{MPLS}(k)] = 0\left(\frac{\log n}{n^2}\right).$$

Proof: See Appendix D.

Theorem 4: Under assumptions A1–A3 and for $k < p$,

$$\text{Var}\,[\text{MPLS}(k)] = 0\left(\frac{1}{n^{1/2}}\right).$$

Proof: See Appendix E.

From Theorems 1 and 2, it follows that as the length of the data grows to infinity, the mean of MPLS (k) attains its minimum for $k = p$. However, to prove the consistency of the MPLS criterion, which implies, as we have seen, the consistency of the PLS criterion, it remains to be shown that the probability of obtaining the minimum at $k = p$ goes to 1 as the sample size grows to infinity. In what follows, we use Theorems 1–4 in conjunction with the Chebyshev inequality to establish this fact.

Theorem 5: Under assumptions A1–A3 and for $n \to \infty$,

$$\Pr[\hat{k} = p] \to 1 \tag{12}$$

where Pr [ ] is the probability of the event in brackets and $\hat{k}$ is the minimizing value of MPLS (k).

Proof: For notational convenience, denote

$$\eta_k = \text{MPLS}(k).$$

Now, since $\hat{k}$ is the minimizing value of η_k, we get

$$\Pr[\hat{k} = p] = \Pr\left[\bigcup_{\substack{k=1 \\ k \neq p}}^{M} (\eta_p > \eta_k)\right] \leqq \sum_{\substack{k=1 \\ k \neq p}}^{M} \Pr[\eta_p > \eta_k]. \tag{13}$$

Thus, it suffices to show that $\Pr[\eta_p > \eta_k] \to 0$ for $k \neq p$. To this end, from Corollary 1.2, we get

$$\Pr[\eta_p > \eta_k] \leqq \Pr\left[\eta_p - \eta_k - (E\eta_p - E\eta_k) > (k-p)\frac{\log n}{n}\right] \leqq \Pr\left[\left|(\eta_p - E\eta_p) - (\eta_k - E\eta_k)\right| > (k-p)\frac{\log n}{n}\right]. \quad (14)$$

Now, for any two real numbers a and b, one can show that the following inequality holds:

$$\Pr[|a-b| > \delta] \leqq \Pr\left[|a| > \frac{\delta}{2}\right] + \Pr\left[|b| > \frac{\delta}{2}\right] \quad (15)$$

which implies that

$$\Pr\left[\left|(\eta_p - E\eta_p) - (\eta_k - E\eta_k)\right| > (k-p)\frac{\log n}{n}\right] \leqq \Pr\left[|\eta_p - E\eta_p| > (k-p)\frac{\log n}{2n}\right] + \Pr\left[|\eta_k - E\eta_k| > (k-p)\frac{\log n}{2n}\right]. \quad (16)$$

Next, by the Chebyshev inequality and Theorem 3, we get for $k > p$

$$\Pr\left[|\eta_k - E\eta_k| > (k-p)\frac{\log n}{2n}\right] \leqq \frac{E[\eta_k - E\eta_k]^2}{\left((k-p)\frac{\log n}{2n}\right)^2} = 0\left(\frac{1}{\log n}\right). \quad (17)$$

Thus, from (14) using (16), we get for $k > p$

$$\Pr[\eta_p > \eta_k] \leqq 0\left(\frac{1}{\log n}\right). \quad (18)$$

Similarly, from Corollary 2.1, we get for $k < p$

$$\Pr[\eta_p > \eta_k] \leqq \Pr[\eta_p - \eta_k - (E\eta_p - E\eta_k) > \alpha_k - \delta] \leqq \Pr\left[\left|(\eta_p - E\eta_p) - (\eta_k - E\eta_k)\right| > \alpha_k - \delta\right]. \quad (19)$$

Now, by the Chebyshev inequality and Theorem 4, we get for $k < p$

$$\Pr\left[|\eta_k - E\eta_k| > \frac{\alpha_k - \delta}{2}\right] \leqq \frac{E(\eta_k - E\eta_k)^2}{\frac{1}{4}(\alpha_k - \delta)^2} = 0\left(\frac{\log n}{n}\right) \quad (20)$$

TABLE I
THE RESULTS OF 100 MONTE CARLO RUNS OF THE DATA GENERATED BY THE FIRST-ORDER MODEL (24). THE LENGTH OF EACH DATA RECORD WAS $n = 15$

k	PLS (k)	MDL (k)	AIC (k)
1	85	84	66
2	3	8	15
3	3	4	9
4	1	0	0
5	1	1	1
6	1	0	2
7	2	1	3
8	1	2	4

while by applying the Chebyshev inequality and Theorem 3, we get

$$\Pr\left[|\eta_p - E\eta_p| > \frac{\alpha_k - \delta}{2}\right] \leqq \frac{E(\eta_p - E\eta_p)^2}{\frac{1}{4}(\alpha_k - \delta)^2} = 0\left(\frac{\log n}{n^2}\right). \quad (21)$$

Thus, from (19), using (15) together with (20) and (21), we obtain for $k < p$

$$\Pr[\eta_p > \eta_k] \leqq 0\left(\frac{\log n}{n}\right). \quad (22)$$

Combining (18) and (22), we finally get that as $n \to \infty$,

$$\sum_{\substack{k=1 \\ k \neq p}}^{M} \Pr[\eta_p > \eta_k] \to 0 \quad (23)$$

which, by (13), completes the proof.

V. SIMULATIONS RESULTS

To demonstrate the performance of the PLS criterion, we compare it to the AIC and MDL criteria in two examples.

In the first example, we generated the data with the following first-order AR:

$$y_t = 0.5y_{t-1} + e_t. \quad (24)$$

Taking 100 Monte Carlo runs of this process, each of length $n = 15$, and then selecting the order by the PLS, MDL, and AIC for each run, we obtained the following results, summarized in Table I; each row in the table represents the number of times that order was selected by the three criteria.

In the second example, we generated the data with the following second-order AR process:

$$y_t = 1.80y_{t-1} - 0.97y_{t-2} + e_t. \quad (25)$$

Taking 100 Monte Carlo runs of this process, this time of length $n = 100$, we obtained the results shown in Table II.

The two examples demonstrate the good performance

TABLE II
THE RESULTS OF 100 MONTE CARLO RUNS OF THE DATA GENERATED BY THE SECOND-ORDER MODEL (25). THE LENGTH OF EACH DATA RECORD WAS $n = 100$

k	PLS (k)	MDL (k)	AIC (k)
1	12	0	0
2	83	93	72
3	1	6	11
4	0	0	7
5	2	1	4
6	0	0	3
7	1	0	1
8	1	0	2

of the PLS criterion while bringing out the somewhat worse performance of the AIC.

APPENDIX A

In this Appendix, we present results needed in proofs outlined in Appendixes B, C, D, and E.

Lemma A.1: Assuming A1–A2, then as $t \to \infty$,

$$E\left[\left|\boldsymbol{C}_{k,t} - \boldsymbol{C}_k\right|^2\right] = 0\left(\frac{1}{t}\right). \tag{A-1}$$

Proof: See Bhansali [2].

Lemma A.2: Assuming A1–A3, then as $t \to \infty$,

$$E\left[\left|\hat{\boldsymbol{a}}_{k,t} - \boldsymbol{a}_k\right|^2\right] = 0\left(\frac{1}{t}\right). \tag{A-2}$$

Proof: See Bhansali [2].

Lemma A.3: Let $\boldsymbol{B}_t$ be a sequence of random matrices and let g_t be a sequence of positive real numbers such that

$$E\left[\left|\boldsymbol{B}_t\right|^2\right] = 0(g_t^2). \tag{A-3}$$

Then

$$\boldsymbol{B}_t = 0_p(g_t). \tag{A-4}$$

That is, for every $\epsilon > 0$, there exists a positive number M_ϵ such that the elements $\{b_{i,j,t}\}$ of the matrix $\boldsymbol{B}_t$ obey the inequality

$$\Pr\left\{\left|b_{i,j,t}\right| > M_\epsilon g_t\right\} < \epsilon. \tag{A-5}$$

Proof: See Fuller [6, p. 185].

Lemma A.4: Let g_t be a sequence of positive real numbers. Then

$$0\left(\sum_{t=1}^{n} 0(g_t)\right) = 0\left(\sum_{t=1}^{n} g_t\right). \tag{A-6}$$

Proof: Denote

$$h_n = \sum_{t=1}^{n} r_t \tag{A-7}$$

with $r_t = 0(g_t)$ so that $h_n = 0(\Sigma_{t=1}^{n} 0(g_t))$. Now, by definition,

$$h_n < M \sum_{t=1}^{n} g_t \tag{A-8}$$

implying that $h_n = 0(\Sigma_{t=1}^{n} g_t)$, which completes the proof.

Lemma A.5: Assuming that A1–A3 hold and that $k > p$ and $t > i$, then

$$E[\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,i}\boldsymbol{y}_{k,i}^T\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,t}\boldsymbol{y}_{k,t}^T e_{i+1}^2] = \sigma^2\boldsymbol{I}_k + 0(\rho^{t-i}) \tag{A-9}$$

where ρ is a positive scalar smaller than 1 and $\boldsymbol{I}_k$ is the $k \times k$ identity matrix.

Proof: Using the well-known first-order vector representation for an autoregressive process, we can rewrite (8) as

$$\boldsymbol{y}_{k,t} = \boldsymbol{A}_k\boldsymbol{y}_{k,t-1} + \boldsymbol{e}_{k,t} \tag{A-10a}$$

where $\boldsymbol{A}_k$ is the $k \times k$ companion matrix

$$\boldsymbol{A}_k = \begin{bmatrix} a_1 & \cdots & a_p & 0 & \cdots & 0 \\ 1 & & & & & \\ 0 & & & & & 0 \\ 0 & & & 0 & 1 & 0 \end{bmatrix} \tag{A-10b}$$

and $\boldsymbol{e}_{k,t}$ is the $k \times 1$ vector

$$\boldsymbol{e}_{k,t} = [e_{k,t-1} \quad 0 \quad \cdots \quad 0]^T. \tag{A-10c}$$

Thus, dropping the subscript k to simplify the notation, we can express y_t as

$$\boldsymbol{y}_t = \boldsymbol{A}^{t-i}\boldsymbol{y}_i + \sum_{j=i+1}^{t} \boldsymbol{A}^{t-j}\boldsymbol{e}_j. \tag{A-11}$$

Recalling that we have assumed the characteristic roots of the matrix $\boldsymbol{A}$ to lie inside the unit circle, it follows that $\boldsymbol{A}^t = 0(\rho^t)$ as $t \to \infty$, and hence, with some straightforward calculations,

$$\begin{aligned} E[\boldsymbol{C}^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T e_{i+1}^2] &= \sigma^2\boldsymbol{C}^{-1}E[\boldsymbol{y}_i\boldsymbol{y}_i^T]\boldsymbol{C}^{-1} \\ &\quad \cdot \sum_{j=i+2}^{t} \boldsymbol{A}^{t-j}\boldsymbol{H}(\boldsymbol{A}^{t-j})^T + 0(\rho^{t-i}) \end{aligned} \tag{A-12a}$$

where

$$\boldsymbol{H} = E[\boldsymbol{e}_j\boldsymbol{e}_j^T]. \tag{A-12b}$$

Now

$$\boldsymbol{C} = \sum_{j=0}^{\infty} \boldsymbol{A}^j\boldsymbol{H}(\boldsymbol{A}^j)^T \tag{A-13}$$

implying that

$$\sum_{j=i+2}^{t} \boldsymbol{A}^{t-j}\boldsymbol{H}(\boldsymbol{A}^{t-j})^T = \boldsymbol{C} + 0(\rho^{t-i}) \tag{A-14}$$

and therefore

$$E[\boldsymbol{C}^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T e_{i+1}^2] = \sigma^2\boldsymbol{I} + 0(\rho^{t-i}). \tag{A-15}$$

Lemma A.6: Assuming A1–A3 hold and that $t > s > j > i$, then

$$E\left[\operatorname{Tr}\left[\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,i}\boldsymbol{y}_{k,i}^T\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,s}\boldsymbol{y}_{k,s}^T\right]\right.$$
$$\left.\cdot \operatorname{Tr}\left[\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,j}\boldsymbol{y}_{k,j}^T\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,t}\boldsymbol{y}_{k,t}^T\right]e_{i+1}^2e_{j+1}^2\right.$$
$$= \sigma^4 \operatorname{Tr}^2[I] + 0(\rho^{j-i}) + 0(\rho^{s-j}) + 0(\rho^{t-s}). \quad \text{(A-16)}$$

Proof: The proof is omitted to save space.

Lemma A.7: Assuming A1–A3 hold and that $k > p$ and $t > j > i$, then

$$E\left[\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,i}\boldsymbol{y}_{k,i}^T\boldsymbol{C}_k^{-1}\boldsymbol{y}_{k,t}\boldsymbol{y}_{k,t}^Te_{i+1}e_{j+1}\right] + 0(\rho^{t-i}). \quad \text{(A-17)}$$

Proof: Using the first-order vector representation (A-10) and some straightforward calculations, again dropping the subscript k to simplify the notation, we get

$$E\left[\boldsymbol{C}^{-1}\boldsymbol{y}_i\boldsymbol{y}_j^T\boldsymbol{C}^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}e_{j+1}\right]$$
$$= 2E\left[\boldsymbol{C}^{-1}\boldsymbol{y}_i\boldsymbol{y}_j^T\boldsymbol{C}^{-1}\boldsymbol{A}^{t-i}\boldsymbol{A}^{t-j}e_{i+1}^2e_{j+1}^2\right]$$
$$= 2E\left[\boldsymbol{C}^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}^{-1}(\boldsymbol{A}^{j-i})^T\boldsymbol{C}^{-1}\boldsymbol{A}^{2t-i-j}e_{i+1}^2e_{j+1}^2\right]$$
$$= 0(\rho^{t-i}). \quad \text{(A-18)}$$

Appendix B

From (2) and (5), using (8) and (10), we get

$$\hat{e}_{k,t} = e_t - v_{k,t} \quad \text{(B-1)}$$

where

$$v_{k,t} = \frac{1}{t}\boldsymbol{y}_{k,t}\boldsymbol{C}_{k,t}^{-1}\sum_{i=1}^{t}\boldsymbol{y}_{k,i}e_{i+1}. \quad \text{(B-2)}$$

Thus, from (11), we have

$$\text{MPLS}(k) = \frac{1}{n}\sum_{t=1}^{n}[2e_tv_{k,t} + v_{k,t}^2]. \quad \text{(B-3)}$$

Taking expectations of both sides, observing that e_t is independent of $v_{k,t}$, we get

$$E[\text{MPLS}(k)] = \frac{1}{n}\sum_{t=1}^{n}E[v_{k,t}^2]. \quad \text{(B-4)}$$

Now, from (B-2), dropping the subscript k to simplify the notation, we obtain

$$E[v_{k,t}^2] = \frac{1}{t^2}E\left[\left[\boldsymbol{y}_t^T\boldsymbol{C}_t^{-1}\sum_{i=1}^{t}\boldsymbol{y}_ie_{i+1}\right]^2\right]$$
$$= \frac{1}{t^2}\sum_{i=1}^{t}\operatorname{Tr}\left[E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}^2]\right]$$
$$+ \frac{1}{t^2}\sum_{i=1}^{t}\sum_{j=i+1}^{t}$$
$$\cdot \operatorname{Tr}\left[E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_i\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}e_{j+1}]\right]$$
$$+ \frac{1}{t^2}\sum_{j=1}^{t}\sum_{i=j+1}^{t}$$
$$\cdot \operatorname{Tr}\left[E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_i^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{j+1}e_{i+1}]\right]. \quad \text{(B-5)}$$

By Lemmas A.1, A.3, and A.5, we obtain

$$E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}^2]$$
$$= E[\boldsymbol{C}^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}^2] + 0\left(\frac{1}{t^{1/2}}\right)$$
$$= \sigma^2\boldsymbol{I} + 0(\rho^{t-i}) + 0\left(\frac{1}{t^{1/2}}\right). \quad \text{(B-6)}$$

Thus,

$$\frac{1}{t^2}\sum_{i=1}^{t}\operatorname{Tr}\left[E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}^2]\right]$$
$$= \frac{k\sigma^2}{t} + 0\left(\frac{1}{t^{3/2}}\right). \quad \text{(B-7)}$$

Similarly, by Lemmas A.1, A.3, and A.7, for $j > i$, we get

$$E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_i\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^{-1}e_{i+1}e_{j+1}] = 0(\rho^{t-i}) + 0\left(\frac{1}{t^{1/2}}\right) \quad \text{(B-8)}$$

and therefore

$$\frac{1}{t^2}\sum_{i=1}^{t}\sum_{j=i+1}^{t}$$
$$\cdot \operatorname{Tr}\left[E[\boldsymbol{C}_t^{-1}\boldsymbol{y}_i\boldsymbol{y}_j\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^Te_{i+1}e_{j+1}]\right] = 0\left(\frac{1}{t^2}\right). \quad \text{(B-9)}$$

Substituting (B-7) and (B-9) into (B-3), we get

$$E[v_{k,t}^2] = k\sigma^2\frac{1}{t} + 0\left(\frac{1}{t^{3/2}}\right) \quad \text{(B-10)}$$

implying that

$$E[\text{MPLS}(k)] = \frac{k\sigma^2}{n}\sum_{t=1}^{n}\frac{1}{t} + 0\left(\frac{1}{n}\right). \quad \text{(B-11)}$$

Appendix C

From (2) and (5), we have

$$\hat{e}_{k,t} = e_t + u_{k,t} + w_{k,t} \quad \text{(C-1)}$$

where

$$u_{k,t} = \boldsymbol{y}_{k,t}^T(\boldsymbol{a}_k - \hat{\boldsymbol{a}}_{k,t}) \quad \text{(C-2)}$$

and

$$w_{k,t} = \boldsymbol{y}_{p,t}^T\boldsymbol{D}_k\boldsymbol{a}_p \quad \text{(C-3)}$$

with $\boldsymbol{D}_k$ denoting the $p \times p$ diagonal matrix whose first k

elements are zero and the rest are ones:

$$\boldsymbol{D}_k = \text{diag}\left\{0, \cdots, 0, 1, \cdots, 1\right\}. \quad \text{(C-4)}$$

Squaring (C-1) and taking expectations, observing that e_t is independent of $u_{k,t}$ and $w_{k,t}$, we get

$$E[\hat{e}_{k,t}^2 - e_t^2] = E[w_{k,t}^2] + E[u_{k,t}^2] + 2E[w_{k,t}u_{k,t}]. \quad \text{(C-5)}$$

Now, by Lemmas A.2 and A.3, we have

$$E[w_{k,t}^2] = \text{Tr}\,[\boldsymbol{D}_k\boldsymbol{a}_p\boldsymbol{a}_p^T\boldsymbol{D}_k\boldsymbol{C}_p] + 0\left(\frac{1}{t}\right)$$

$$E[u_{k,t}^2] = 0\left(\frac{1}{t}\right)$$

and

$$E[u_{k,t}w_{k,t}] = 0\left(\frac{1}{t^{1/2}}\right)$$

so that

$$E[\hat{e}_{k,t}^2 - e_t^2] = \text{Tr}\,[\boldsymbol{D}_k\boldsymbol{a}_p\boldsymbol{a}_p^T\boldsymbol{D}_k\boldsymbol{C}_p] + 0\left(\frac{1}{t^{1/2}}\right)$$

implying, from (8), that

$$E[\text{MPLS}\,(k)] = \text{Tr}\,[\boldsymbol{D}_k\boldsymbol{a}_p\boldsymbol{a}_a^T\boldsymbol{D}_k\boldsymbol{C}_p] + 0\left(\frac{1}{n^{1/2}}\right).$$

Appendix D

Squaring both sides of (B-3) and taking expectations, observing that e_s is independent of $v_{k,s}$ and $v_{k,t}$ for $s > t$, we get

$$\begin{aligned} E\left[\left[\text{MPLS}\,(k)\right]^2\right] &= \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}E[v_{k,s}^2v_{k,t}^2] \\ &\quad + \frac{4}{n^2}\sum_{t=1}^{n}E[e_t^2v_{k,t}^2] \\ &\quad - \frac{2}{n^2}\sum_{s=1}^{n}\sum_{t=s+1}^{n}E[e_sv_{k,s}v_{k,t}^2] \\ &\quad - \frac{2}{n^2}\sum_{t=1}^{n}\sum_{s=t+1}^{n}E[e_tv_{k,t}v_{k,s}^2]. \end{aligned} \quad \text{(D-1)}$$

Now, using (B-2) and dropping the subscript k to simplify the notation, for $t > s$, we get

$$\begin{aligned} E[v_s^2v_t^2] &= \frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{j=1}^{t}E\left[\text{Tr}\left(\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T\right]\right. \\ &\quad \cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}^2e_{j+1}^2] \\ &\quad + \frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{j=1}^{s}E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\quad \cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}^2e_{j+1}^2] \\ &\quad + \frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{\substack{h=1\\ i\neq h}}^{s}\sum_{j=1}^{t}E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_h^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\quad \cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}e_{h+1}e_{j+1}^2] \\ &\quad + \frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{\substack{j=1\\ j\neq l}}^{t}\sum_{l=1}^{t}e\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\quad \cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_l^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}^2e_{j+1}e_{l+1}] \\ &\quad + \frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{h=1}^{s}\sum_{j=1}^{t}\sum_{l=1}^{t} \\ &\quad \cdot E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_h^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\quad \left.\left.\cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_l^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t]e_{i+1}e_{h+1}e_{j+1}e_{l+1}\right]. \right. \end{aligned} \quad \text{(D-2)}$$

Now, by some straightforward calculations using Lemmas A.1, A.2, A.3, A.4, and A.7, we get

$$\begin{aligned} &\frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{j=1}^{t}E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\qquad \left.\cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}^2e_{j+1}^2\right] \\ &\quad = k^2\sigma^4\frac{1}{st} + 0\left(\frac{1}{st}\rho^{t-s}\right) + 0\left(\frac{1}{s^{3/2}t}\right) \end{aligned} \quad \text{(D-3)}$$

and

$$\begin{aligned} &\frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{j=1}^{s}E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{u}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\qquad \left.\cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}^2e_{j+1}^2\right] \\ &\quad = k^2\sigma^4\frac{1}{t^2} + 0\left(\frac{1}{t^2}\rho^{t-s}\right) + 0\left(\frac{1}{s^{1/2}t^2}\right). \end{aligned} \quad \text{(D-4)}$$

By similar calculations, we have

$$\begin{aligned} &\frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{\substack{h=1\\ i\neq h}}^{s}\sum_{j=1}^{t}E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\qquad \left.\cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^T]e_{i+1}e_{h+1}e_{j+1}\right] \\ &\quad = 0\left(\frac{1}{s^2t^2}\right) \end{aligned} \quad \text{(D-5)}$$

$$\begin{aligned} &\frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{\substack{j=1\\ j\neq l}}^{t}\sum_{l=1}^{t}E\left[\text{Tr}\,[\boldsymbol{C}_s^{-1}\boldsymbol{y}_i\boldsymbol{y}_i^T\boldsymbol{C}_s^{-1}\boldsymbol{y}_s\boldsymbol{y}_s^T]\right. \\ &\qquad \left.\cdot \text{Tr}\,[\boldsymbol{C}_t^{-1}\boldsymbol{y}_j\boldsymbol{y}_j^T\boldsymbol{C}_t^{-1}\boldsymbol{y}_t\boldsymbol{y}_t^t]e_{i+1}e_{j+1}e_{l+1}\right] \\ &\quad = 0\left(\frac{1}{s^2t^2}\right) \end{aligned} \quad \text{(D-6)}$$

and

$$\frac{1}{s^2t^2}\sum_{i=1}^{s}\sum_{\substack{h=1\\ i\neq h}}^{s}\sum_{j=1}^{t}\sum_{\substack{l=1\\ j\neq l}}^{t} E\left[\operatorname{Tr}\left[\boldsymbol{C}^{-1}\mathbf{y}_i\mathbf{y}_i^T\boldsymbol{C}_s^{-1}\mathbf{y}_s\mathbf{y}_s^T\right]\cdot \operatorname{Tr}\left[\boldsymbol{C}_t^{-1}\mathbf{y}_j\mathbf{y}_j^T\boldsymbol{C}_t^{-1}\mathbf{y}_t\mathbf{y}_t^T\right]e_{i+1}e_{h+1}e_{j+1}e_{l+1}\right] = 0\left(\frac{1}{s^2t^2}\right). \quad \text{(D-7)}$$

Substituting (D-3)–(D-7) into (D-2) and using Lemma A.4, we get

$$\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=t+1}^{n} E[v_{k,s}^2v_{k,t}^2] = \frac{k^2\sigma^4}{n^2}\sum_{s=1}^{n}\sum_{t=s+1}^{n}\frac{1}{st} + 0\left(\frac{\log n}{n^2}\right). \quad \text{(D-8)}$$

By straightforward and similar calculations, we also get

$$E[e_{k,s}v_{k,s}v_{k,t}^2] = 0\left(\frac{1}{st}\rho^{t-s}\right)$$

implying, by Lemma A.4, that

$$\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=t+1}^{n} E[e_{k,t}v_{k,s}v_{k,t}^2] = 0\left(\frac{\log n}{n^2}\right). \quad \text{(D-9)}$$

Finally, recalling that $e_{k,t}$ is independent of $v_{k,t}$, from Corollary 1.1, we get

$$\frac{1}{n^2}\sum_{t=1}^{n} E[e_{k,t}^2v_{k,t}^2] = \frac{\sigma^2}{n^2}\sum_{t=1}^{n} E[v_{k,t}^2] = 0\left(\frac{\log n}{n^2}\right). \quad \text{(D-10)}$$

Substituting (D-9)–(D-12) into (D-1), we obtain

$$E\left[\left(\text{MPLS }(k)\right)^2\right] = \left(\frac{k\sigma^2}{n}\sum_{t=1}^{n}\frac{1}{t}\right)^2 + 0\left(\frac{\log n}{n^2}\right) \quad \text{(D-11)}$$

and therefore

$$\text{Var}\left[\text{MPLS }(k)\right] = \left(E\left[\text{MPLS }(k)\right]\right)^2 - E\left[\left(\text{MPLS }(k)\right)^2\right] = 0\left(\frac{\log n}{n^2}\right). \quad \text{(D-12)}$$

Appendix E

From (C-1), invoking Lemmas A.2, A.3, and A.4 and some straightforward calculations, we get

$$E\left[\left[\text{MPLS }(k)\right]^2\right] = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n} E\left[(\hat{e}_{k,t}^2 - e_t^2)(\hat{e}_{k,s}^2 - e_s^2)\right] = \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n} E[w_{k,t}^2w_{k,s}^2] + 0\left(\frac{1}{n}\right). \quad \text{(E-1)}$$

Now, taking $t > s$ and invoking Lemma A.5, we obtain

$$E[w_{k,t}^2w_{k,s}^2] = \operatorname{Tr}\left[\boldsymbol{D}_k\mathbf{a}_p\mathbf{a}_p^T\boldsymbol{D}_k\boldsymbol{C}_p\right] + 0(\rho^s) + 0(\rho^{t-s}) \quad \text{(E-2)}$$

where ρ is a positive constant smaller than 1. Thus, by straightforward application of Lemma A.4, we get

$$E\left[\left[\text{MPLS }(k)\right]^2\right] = \left[\operatorname{Tr}\left[\boldsymbol{D}_k\mathbf{a}_p\mathbf{a}_p^T\boldsymbol{D}_k\boldsymbol{C}_p\right]\right]^2 + 0\left(\frac{1}{n}\right) \quad \text{(E-3)}$$

and hence, by Theorem 2,

$$\text{Var}\left[\text{MPLS }(k)\right] = 0\left(\frac{1}{n^{1/2}}\right). \quad \text{(E-4)}$$

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. Caski, Eds., 1973, pp. 267–281.

[2] R. J. Bhansali, "Effects of not knowing the order of an autoregressive process on the mean squared error of prediction—I," *J. Amer. Statist. Ass.*, vol. 76, pp. 588–597, 1981.

[3] J. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 304–337, 1984.

[4] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, pp. 829–867, 1982.

[5] ——, "Lattice methods for spectral estimation," *Proc. IEEE*, vol. 70, pp. 990–1017, 1982.

[6] W. A. Fuller, *Introduction to Statistical Time Series*. New York: Wiley, 1976.

[7] D. Lee, M. Morf, and B. Friedlander, "Recursive least-squares ladder estimation algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 627–641, 1981.

[8] H. Lev-Ari, T. Kailath, and J. Cioffi, "Least-squares adaptive lattice and transversal filters," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 222–236, 1984.

[9] M. Morf, A. Vieira, and D. Lee, "Lattice forms for identification and speech processing," in *Proc. ICASSP*, 1977, pp. 1074–1078.

[10] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[11] ——, "A predictive least-squares principle," *IMA J. Math. Contr. Inform.*, vol. 3, nos. 2–3, pp. 211–222, 1986.

[12] G. Schwarz, "Estimation of the dimension of the model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

Mati Wax received the B.Sc. and the M.Sc. degrees from the Technion, Haifa, Israel, in 1969 and 1975, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, in 1985, all in electrical engineering.

From 1969 to 1973 he was in the Israeli Army where he developed communication systems. In 1974 he was with A.E.L., Israel, where he was engaged in the development of microwave components and subsystems. In 1975 he joined RAFAEL, where he worked on the development of communication systems and position location techniques. From 1980 to 1983 he was at Stanford University, Stanford, CA, where he was involved in research in the area of signal processing. During 1984 he was a Visiting Scientist at the IBM Research Laboratories, San Jose, CA, where he did research in pattern recognition and image compression. Since 1985 he has been with RAFAEL, heading the Center for Signal Processing.

Dr. Wax is the recipient of the 1985 Senior Paper Award of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING.