

- [3] J. M. Van Campenhout and T. M. Cover, "Maximum entropy and conditional probability," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 483-489, 1981.
- [4] I. Csiszar, T. M. Cover, and B. S. Choi, "Conditional limit theorems under Markov conditioning," *IEEE Trans. Inform. Theory*, pp. 788-801, 1987.
- [5] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meet. Soc. Exploration Geophysicists*, 1967; also in *Modern Spectrum Analysis*, D. G. Childers, Ed., 1968, pp. 34-41.
- [6] B. S. Choi and T. M. Cover, "An information theoretic proof of Burg's maximum entropy spectrum," *Proc. IEEE*, vol. 72, pp. 1094-1095, 1984.
- [7] B. S. Choi, "On the relation between the maximum entropy probability density function and the autoregressive model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1659-1661, 1986.
- [8] L. Breiman, "The individual ergodic theorem of information theory," *Ann. Math. Stat.*, vol. 28, pp. 809-814, 1957. (Correction, vol. 31, pp. 809-810, 1960.)
- [9] A. El Gamal and T. M. Cover, "Multiple user information theory," *Proc. IEEE*, vol. 68, pp. 1466-1483, 1980.
- [10] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.
- [11] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907-939, 1982.
- [12] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380-1419, 1981.
- [13] IBM Corporation, *IBM System/360 Scientific Subroutine Package*. New York: IBM System Reference Library, 1968.
- [14] W. J. Kennedy and J. E. Gentle, *Statistical Computing*. New York: Marcel Dekker, 1980.
- [15] C. Hastings, *Approximations for Digital Computers*. Princeton, NJ: Princeton University, 1955.
- [16] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive moving average time series models," *J. Amer. Stat. Ass.*, vol. 65, pp. 1509-1526, 1970.
- [17] R. Weinstock, *Calculus of Variations with Applications to Physics and Engineering*. New York: Dover, 1974.
- [18] E. A. Robinson, *Statistical Communication and Detection with Special Reference to Digital Data Processing of Radar and Seismic Signals*. London: Griffin, 1967.

Sliding Windows and Lattice Algorithms for Computing QR Factors in the Least Squares Theory of Linear Prediction

CÉDRIC J. DEMEURE AND LOUIS L. SCHARF

Abstract—In this correspondence we pose a sequence of linear prediction problems that differ a little from those previously posed. The solutions to these problems introduce a family of "sliding" window techniques into the least squares theory of linear prediction. By using these techniques we are able to QR factor the Toeplitz data matrices that arise in linear prediction. The matrix Q is an orthogonal version of the data matrix and the matrix R is a Cholesky factor of the experimental correlation matrix. Our QR and Cholesky algorithms generate generalized reflection coefficients that may be used in the usual ways for analysis, synthesis, or classification.

Manuscript received October 28, 1988; revised June 20, 1989. This work was supported by the Office of Naval Research, Arlington, VA, under Contract N00014-85-K-0256.

The authors are with the Electrical and Computer Engineering Department, University of Colorado, Boulder, CO 80309.

IEEE Log Number 8934016.

I. INTRODUCTION

In this correspondence we pose a sequence of least squares problems from the theory of linear prediction. The problems are a little different from those originally posed in the paper by Morf *et al.* [6]. The solutions to these problems produce QR factorizations of the data matrices that are usually associated with the covariance, prewindowed, postwindowed, and correlation methods of linear prediction. Our results apply to forward, backward, or forward-backward linear prediction. In this paper, we treat only the forward covariance problem. All of the others are treated in [15].

By QR factorization we mean the computation of an orthogonal matrix Q and an upper triangular matrix R such that a data matrix Y is written as $Y = QR$. The factorization may be used to solve the overdetermined system of equations $Ya = b$ by $R\hat{a} = Q^T b$.

Our approach is to use the generalized Levinson recursions derived by Friedlander *et al.* [1] to derive generalized recursions for computing the orthogonal matrix Q in the QR factorization of any of the Toeplitz (or concatenation of Toeplitz) matrices that can arise in linear prediction. These recursions generalize those first discovered by Cybenko [10] for the correlation method of linear prediction. We then use these recursions to derive generalized Schur recursions for Cholesky factoring any of the close-to-Toeplitz covariance matrices that can arise in linear prediction. Our procedures are generalizations of those reported in [4] for solving least squares problems in the correlation method of linear prediction. Our results differ from the recent results of Cybenko [16], in the sense that a direct QR factorization $Y = QR$ is obtained here, not an "inverse" factorization $YA = Q$.

II. LEAST SQUARES PROBLEMS IN LINEAR PREDICTION

Let $y = [y_0, y_1, \dots, y_{N-1}]^T$ denote an N sample snapshot of the stationary time series $\{y_t\}$. From this snapshot, we would like to identify an autoregressive or whitening model $A_n(z)$ for the time series $\{y_t\}$. This model takes the form

$$\sum_{i=0}^n a_i^n y_{t-i} = u_t^n; \quad a_0^n = 1 \quad (2.1)$$

where $\{u_t^n\}$ is an error sequence. The model may also be written as the predictor model

$$y_t = \hat{y}_t^n + u_t^n \quad \hat{y}_t^n = -\sum_{i=1}^n a_i^n y_{t-i} \quad (2.2)$$

where \hat{y}_t^n is the n th order prediction of y_t . The squared error between y_t and the one-step ahead predictor \hat{y}_t^n is

$$\sigma_n^2 = \sum_{t \in T_n} (u_t^n)^2 \quad (2.3)$$

The choice of T_n , the index set, determines the choice between various methods of linear prediction. What differentiates our methods from the methods in [6] is the way the set T_n varies with n .

The procedure for identifying a model $A_n(z)$ will be to form a sequence of predictions of the form

$$\hat{y}_t^p = -\sum_{i=1}^p a_i^p y_{t-i}; \quad a_0^p = 1 \quad (2.4)$$

and to let the predictor order range from $p = 0$ to $p = n$.

A. Sliding Windows

We shall be interested in the index sets T_n for which the time index t satisfies the condition

$$t \in T_n \Leftrightarrow k \leq t + (n - p) \leq l. \quad (2.5)$$

As the predictor order increases from $p = 0$ to $p = n$, a window of constant length $(l - k + 1)$ moves from left to right across the data set, as illustrated in Fig. 1. The indexes k and l may be chosen

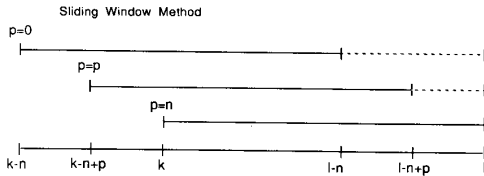


Fig. 1. Illustration of the sliding window for prediction orders increasing from $p = 0$ to $p = n$.

to select among the various methods of linear prediction (just as in [6]). The data values outside the range $[0, \dots, N - 1]$ are set to zero. In the *covariance method* of linear prediction, $k = n$ and $l = N - 1$. Fig. 2 illustrates the difference between the covariance method of linear prediction studied in [6] and our "sliding window" technique for the covariance method, by showing for each order the number of prediction errors used in determining the least squares solution. Note that the prediction errors for the final order n are the same for the two techniques, so the *final* answer for the filter $A_n(z)$ will be the same.

The choice of the index set is of utmost importance, as it determines the kind of structure the final algorithm will inherit. For example, in the covariance method of linear prediction [6], the maximum number of prediction errors (not using data outside the range $[0, N - 1]$) is used for each value of p . This leads to a decreasing number of prediction errors as p increases (see Fig. 2). In this case both a time and an order update are necessary in the algorithm. In our sliding window method, the number of prediction errors is only dependent on the maximum order n and not on the intermediate order p . This leads to an order update algorithm, *with no time update*. Such algorithms are sometimes called *nested*. There is a small computational advantage in our technique, but more importantly the structure of the algorithm is quite different and the algorithm has a square root or *QR* version. No other algorithm does.

B. Error Equations

Let us write out the error equations, over the sliding window just defined, for the p th order predictor:

$$\begin{bmatrix} y_{l-n} & \cdots & y_{l-1} & y_l \\ \vdots & \ddots & \vdots & \vdots \\ \vdots & \ddots & y_{l-1} & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ \vdots & \ddots & y_{l-n} & \vdots \\ y_k & \cdots & \vdots & \vdots \\ \vdots & \ddots & \vdots & \vdots \\ y_{k-n} & \cdots & y_{k-1} & y_k \end{bmatrix} \begin{bmatrix} a_p^p \\ \vdots \\ a_1^p \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} u_{l-n+p}^p \\ \vdots \\ u_{l-n+p-1}^p \\ \vdots \\ u_{k-n+p+1}^p \\ \vdots \\ u_{k-n+p}^p \end{bmatrix} \quad (2.6)$$

The compact notation is

$$Y \begin{bmatrix} \mathbf{a}^p \\ \mathbf{0} \end{bmatrix} = Y A^p = U^p \quad (2.7)$$

$$A^p = \begin{bmatrix} \mathbf{a}^p \\ \mathbf{0} \end{bmatrix}$$

where Y is the data matrix, \mathbf{a}^p is the whitening vector of order p , and U^p contains the error sequence of order p . This scheme may be reproduced for $p = 0$ to $p = n$ to obtain the set of equations $Y A = U$, where Y is the Toeplitz data matrix just defined, and the matrices A and U are given by

$$A = [A^0, \dots, A^p, \dots, A^n]$$

$$U = [U^0, \dots, U^p, \dots, U^n] \quad (2.8)$$

Any particular choice for k and l slices the infinite Toeplitz matrix built on the infinite data sequence $\dots, 0, y_0, y_1, \dots, y_{N-1}$,

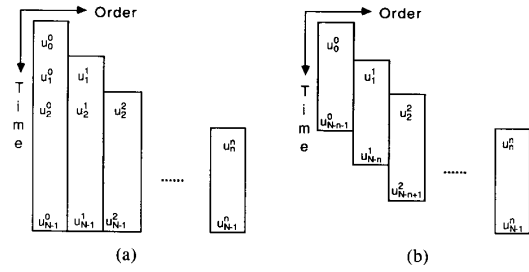


Fig. 2. Prediction errors for order-increasing predictors. (a) Classical technique for the covariance method, (b) sliding window technique for the covariance method.

$0, \dots$ by removing some of the equations, or equivalently by removing some of the prediction error values on the right-hand side of (2.6). This is illustrated in Fig. 3. Such a matrix representation is not possible in the covariance method of linear prediction studied in [6], because each order p requires a different definition for the data matrix Y .

C. Least Squares

The least squares solution for \mathbf{a} in $Y\mathbf{a} \approx \mathbf{0}$, with the constraint $a_0 = 1$ is

$$C\mathbf{a} = Y^T Y \mathbf{a} = \sigma^2 [0, \dots, 0, 1]^T \quad (2.9)$$

When this solution is written out for orders $p = 0, 1, \dots, n$, the result is

$$C A = Y^T Y A = H D^2 \quad (2.10)$$

where $H = (A^T)^{-1}$ is a lower triangular matrix and D^2 is the diagonal matrix containing the prediction errors $D^2 = \text{diag}[\sigma_0^2, \sigma_1^2, \dots, \sigma_n^2]$. These are the normal equations for the least squares theory of linear prediction when errors are defined with respect to the sliding windows of (2.5).

D. QR Factor

Writing $Y A = Q$ and using the least squares equations for A we have $Q^T Q = D^2$, an orthogonal factorization for the data matrix Y . By solving the right sequence of least squares problems, we *QR* factor a Toeplitz data matrix and produce Cholesky factors of the experimental covariance matrix and its inverse. We can think of *QR* factoring the data matrix as a "square root" method of factoring the experimental correlation matrix (and its inverse). We can also think of the inverse *QR* factor $Y A = U$, the Cholesky factor $A^T C A = D^2$, and the Cholesky factor $C = H D^2 H^T$ as three different ways of characterizing the matrix A , which contains order-increasing prediction filters.

III. FACTORING C^{-1} INTO ITS CHOLESKY FACTORS

The problem of factoring C^{-1} is the problem of finding A in the diagonalization

$$A^T C A = D^2 = \text{diag}[\sigma_0^2, \sigma_1^2, \dots, \sigma_n^2] \quad (3.1)$$

This equation may be written as $C A = H D^2$, with $H = (A^{-1})^T$, and read out as follows:

$$C_i \mathbf{a}^i = \sigma_i^2 [0, \dots, 0, 1]^T \quad (3.2)$$

Here C_i is the $(i + 1)$ by $(i + 1)$ top left submatrix of C . When i is incremented to $i + 1$ then, of course, a new column and a new row are added to C_i . If the resulting matrix C_{i+1} has a simple recursive dependence on C_i , then there is reason to hope for a recursive dependence of \mathbf{a}^{i+1} on \mathbf{a}^i . This was the insight of Friedlander *et al.* [1].

The matrix C is $Y^T Y$, $J Y^T Y J$, or $Y^T Y + J Y^T Y J$ depending on the choice between forward, backward, or forward + backward linear prediction [15]. With Y Toeplitz, this means the $(n + 1)$ by

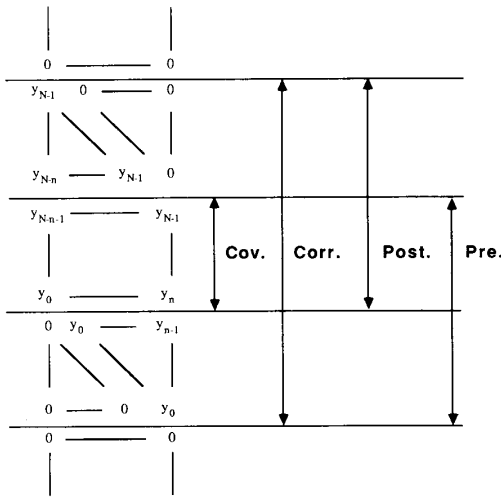


Fig. 3. Slicing the Toeplitz data matrix to obtain the sliding window techniques for the different methods of linear prediction.

$(n + 1)$ symmetric, positive definite matrix is close to Toeplitz. The shifted difference matrix [1] is the $n \times n$ matrix:

$$\delta[C]_{i,j} = c_{i+1,j+1} - c_{i,j}. \quad (3.3)$$

The rank of $\delta[C]$ is the displacement rank α . The decomposition of $\delta[C]$ may be written $\delta[C] = E\Sigma E^T$, where E is an $n \times \alpha$ matrix and Σ is an $\alpha \times \alpha$ diagonal signature matrix, containing $+1$ or -1 on its diagonal. The fundamental equation used in the derivation of fast algorithms is the update for the matrix C_i , using E_i , which consists of the first $(i + 1)$ rows of E :

$$C_{i+1} = \begin{bmatrix} c_{0,0} & c_{0,1} \cdots c_{0,i+1} \\ c_{1,0} & \\ \vdots & \\ c_{i+1,0} & C_i + E_i \Sigma E_i^T \end{bmatrix}. \quad (3.4)$$

The idea is to correct a Toeplitz approximation for C_{i+1} with a low rank matrix $E_i \Sigma E_i^T$. Note that $C_0 = c_{0,0}$ and $C_n = C$. When C is Toeplitz then $c_{i,j} = c_{|i-j|}$, which means that $\delta[C]$ is zero. $Y^T Y$ has a displacement rank equal to zero in the correlation method of linear prediction, one in the prewindowed and postwindowed methods of linear prediction, two in the covariance method of linear prediction, and four in the forward-backward covariance method of linear prediction.

The generalized Levinson algorithm for the columns of the upper triangular matrix A may be derived using the difference matrix to get [1]:

$$\begin{aligned} \hat{A}^{i+1} &= \hat{A}^i + Z A^i K_{i+1}^T \\ A^{i+1} &= Z A^i + \hat{A}^i \hat{K}_{i+1} \end{aligned} \quad \text{for } i = 0 \text{ to } n-1. \quad (3.5)$$

The matrix Z in these equations is the delay matrix:

$$Z = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ & & & 0 \\ & I & & \vdots \\ & & & 0 \end{bmatrix}. \quad (3.6)$$

These recursions are initialized by

$$\begin{aligned} A^0 &= [1, 0, \cdots, 0]^T & \sigma_0^2 &= c_{0,0} \\ \hat{A}^0 &= [A^0, \mathbf{0}, \cdots, \mathbf{0}] & \Delta_0 &= \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & -\sigma_0^{-2} \Sigma \end{bmatrix}. \end{aligned} \quad (3.7)$$

The vector K_{i+1} , an $(\alpha + 1) \times 1$ vector which generalizes the usual scalar reflection coefficient k_{i+1} , may be computed by [1]

$$\sigma_i^2 K_{i+1}^T = -(Z A^i)^T \begin{bmatrix} c_{0,0} & \mathbf{0}^T \\ c_{1,0} \\ \vdots \\ E \Sigma \\ c_{n,0} \end{bmatrix} \quad (3.8)$$

and \hat{K}_{i+1} is the solution to $\Delta_i \hat{K}_{i+1} = K_{i+1}$. The prediction error σ_i^2 , and the error matrix Δ_i are updated as follows:

$$\begin{aligned} \sigma_{i+1}^2 &= \sigma_i^2 (1 - K_{i+1}^T \Delta_i^{-1} K_{i+1}) \\ \sigma_{i+1}^2 \Delta_{i+1} &= \sigma_i^2 (\Delta_i - K_{i+1} K_{i+1}^T). \end{aligned} \quad (3.9)$$

Note that, as only Δ_i^{-1} is used in the equations, it might be more numerically advantageous to use the alternate update equation

$$\Delta_{i+1}^{-1} = \frac{\sigma_{i+1}^2}{\sigma_i^2} \Delta_i^{-1} + \hat{K}_{i+1} \hat{K}_{i+1}^T$$

with the initialization

$$\Delta_0 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & -\sigma_0^2 \Sigma \end{bmatrix}.$$

IV. ORTHOGONALIZATION OF THE DATA MATRIX Y

Using the recursions for the columns of A , we find the corresponding recursions for the columns of the orthogonal matrix U , using the QR equation $Y A = U$. For the correlation method of linear prediction, this procedure [4] produces the algorithm of Cybenko [10]. We extend this procedure to the covariance method in this section.

Let's now extend the data matrix Y with a Toeplitz pattern to produce the following Toeplitz matrix W :

$$W = \begin{bmatrix} y_{N-1} & 0 & 0 & \cdots & 0 \\ y_{N-2} & y_{N-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ y_{N-n} & \cdots & y_{N-2} & y_{N-1} & 0 \\ \hline y_{N-n-1} & \cdots & y_{N-3} & y_{N-2} & y_{N-1} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ y_0 & & & y_{n-1} & y_n \end{bmatrix}. \quad (4.1)$$

The data matrix Y has inverse QR factor $Y A = U$. Therefore the matrix W has the factor

$$W A = \begin{bmatrix} \mathcal{U} \\ U \end{bmatrix} \quad (4.2)$$

which means the orthogonal matrix U is embedded in a larger matrix. Define the i th column of $W A$ as follows:

$$W A^i = \begin{bmatrix} \mathcal{U}^i \\ U^i \end{bmatrix}. \quad (4.3)$$

Similarly define

$$W \hat{A}^i = \begin{bmatrix} \hat{\mathcal{U}}^i \\ \hat{U}^i \end{bmatrix}. \quad (4.4)$$

We reproduce these equations for $(i + 1)$ and use the generalized Levinson recursions (3.5) for A^i and \hat{A}^i to get coupled recursions for computing the vectors U^i and \hat{U}^i :

$$\begin{aligned} W A^{i+1} &= \begin{bmatrix} \mathcal{U}^{i+1} \\ U^{i+1} \end{bmatrix} = Z \begin{bmatrix} \mathcal{U}^i \\ U^i \end{bmatrix} + \begin{bmatrix} \hat{\mathcal{U}}^i \\ \hat{U}^i \end{bmatrix} \hat{K}_{i+1} \\ W \hat{A}^{i+1} &= \begin{bmatrix} \hat{\mathcal{U}}^{i+1} \\ \hat{U}^{i+1} \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{U}}^i \\ \hat{U}^i \end{bmatrix} + Z \begin{bmatrix} \mathcal{U}^i \\ U^i \end{bmatrix} K_{i+1}^T. \end{aligned} \quad (4.5)$$

These recursions perform the computation of the orthogonal matrix Q column by column in the QR factorization of the Toeplitz data matrix Y . To derive these recursions we used the fact that the last element of A^i is equal to zero for $i < n$, in which case $ZWA^i = WZA^i$ for our definition of W .

Here the shifted difference matrix $\delta[C]$ has rank 2, and

$$E = \begin{bmatrix} y_0 & \cdots & y_{n-1} \\ y_{N-n} & \cdots & y_{N-1} \end{bmatrix}^T \quad \text{and} \quad \Sigma = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (4.6)$$

The recursions are initialized as follows:

$$\begin{aligned} \mathcal{U}^0 &= [y_{N-1}, \dots, y_{N-n}]^T \\ U^0 &= [y_{N-n-1}, \dots, y_1, y_0]^T \quad \begin{bmatrix} \hat{\mathcal{U}}^0 \\ \hat{U}^0 \end{bmatrix} = \begin{bmatrix} \mathcal{U}^0 & \mathbf{0} & \mathbf{0} \\ U^0 & \mathbf{0} & \mathbf{0} \end{bmatrix} \\ \sigma_0^2 &= c_{0,0} = (U^0)^T U^0 \quad \Delta_0 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & -\sigma_0^{-2} \Sigma \end{bmatrix}. \end{aligned} \quad (4.7)$$

The three components of the reflection coefficient K_{i+1} may be computed using

$$\begin{aligned} \sigma_i^2 K_{i+1}(1) &= -[(U^0)^T, 0] \begin{bmatrix} \mathcal{U}^i(n) \\ U^i \end{bmatrix} \\ \sigma_i^2 K_{i+1}(2) &= U^i(N-n) \\ \sigma_i^2 K_{i+1}(3) &= -\mathcal{U}^i(n) \end{aligned} \quad (4.8)$$

which comes from the use of (3.8). Only one inner product is necessary per update. Note that $U^i(N-n)$ is the last value contained in the vector U^i , and that $\mathcal{U}^i(n)$ is the last value of the vector \mathcal{U}^i .

An alternate formula for K_{i+1} (and so for K_{i+1}) may also be obtained using the orthogonality property of the columns of U :

$$[0(\hat{U}^i)^T] \begin{bmatrix} \mathcal{U}^i \\ \hat{U}^i \end{bmatrix} \hat{K}_{i+1} = -[0(\hat{U}^i)^T] Z \begin{bmatrix} \mathcal{U}^i \\ U^i \end{bmatrix} \quad (4.9)$$

where we have used (4.5) and the fact that U^{i+1} is orthogonal to U^i and \hat{U}^i . This formula is computationally more demanding, but it improves the sensitivity to roundoff errors.

V. FACTORING C INTO ITS CHOLESKY FACTORS

The LU factorization of C may be written $C = HD^2H^T$. In the Toeplitz case, the LeRoux-Gueguen algorithm [2] may be used to compute H directly. In all of the linear prediction cases, or in other words when $C = Y^T Y$, where Y is Toeplitz or a concatenation of Toeplitz matrices, the recursions for the columns of H are easily deduced from the QR recursions by simply premultiplying the recursions in equation (4.5) by $[0Y^T]$. This was first done in [4] for the Toeplitz case and generalized in [15] for the general close-to-Toeplitz case.

A more general approach to this derivation consists of using the generalized Levinson recursions directly to deduce recursions for the columns of H . Then there is no need for C to equal $Y^T Y$. The resulting algorithm is a generalization of the vector version of the Leroux-Gueguen algorithm derived in [3], and is similar to the algorithm presented in [17]. In order to derive this algorithm, define H^i and \hat{H}^i :

$$H^i = CA^i \quad \hat{H}^i = C\hat{A}^i - \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & E \end{bmatrix} \Delta_i \sigma_i^2. \quad (5.1)$$

Coupled recursions may then be obtained by premultiplying (3.5) by C to get:

$$\begin{aligned} \hat{H}^{i+1} + \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & E \end{bmatrix} \Delta_{i+1} \sigma_{i+1}^2 &= \hat{H}^i + \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & E \end{bmatrix} \Delta_i \sigma_i^2 + CZA^i K_{i+1}^T \\ H^{i+1} = CZA^i + \hat{H}^i \hat{K}_{i+1} + \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & E \end{bmatrix} K_{i+1} \sigma_i^2. \end{aligned} \quad (5.2)$$

We obtain a formula for CZA^i by using (3.4):

$$CZA^i = ZH^i - \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & E \end{bmatrix} K_{i+1} \sigma_i^2 \quad (5.3)$$

so that together with (3.9), we get the recursions:

$$\begin{aligned} \hat{H}^{i+1} &= \hat{H}^i + ZH^i K_{i+1}^T \\ H^{i+1} &= ZH^i + \hat{H}^i \hat{K}_{i+1} \end{aligned} \quad \text{for } i = 0, \dots, n-1. \quad (5.4)$$

These recursions are initialized as follows:

$$\begin{aligned} H^0 &= [c_{0,0}, c_{1,0}, \dots, c_{n,0}]^T \quad \hat{H}^0 = \begin{bmatrix} 0 & c_{1,0} & \cdots & c_{n,0} \\ \mathbf{0} & \Sigma E^T & & \end{bmatrix}^T \\ \sigma_0^2 &= c_{0,0} \quad \Delta_0 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & -\sigma_0^{-2} \Sigma \end{bmatrix}. \end{aligned} \quad (5.5)$$

Note that H^i is the i th column of HD^2 , or similarly $H^i \sigma_i^{-2}$ is the i th column of H . The first i elements of H^i and the first $(i+1)$ rows of \hat{H}^i are equal to zero. The coefficient K_{i+1} may be read out of the recursions as $\hat{H}^{i+1}(i) = [0, \dots, 0] = \hat{H}^i(i) + K_{i+1}^T \sigma_i^2$, so that $-K_{i+1}^T \sigma_i^2$ equals the first nonzero row in \hat{H}^i . These coupled recursions include the update for $\sigma_i^2 = H^i(i)$, the first nonzero element in H^i .

If the reflection coefficients are known, it is a challenging problem to see if and how the correlation coefficients and the matrix H may be computed from them. In the correlation case, Robinson and Treitel [7] solved this problem by observing that the all-pole lattice filter has an output equal to the causal part of the correlation sequence when the input is zero and the state is initialized at $[c_{0,0}, 0, \dots, 0]$. The result originates from the work of Kunetz and d'Ereuve [14] on the propagation of planar waves in a layered system. The multichannel case was studied by Friedlander [8]. We present the generalization of these results to the close-to-Toeplitz case.

Using the following notation for the entries of $HD^2 = \{H_{i,j}\} = H^j(i)$, we take advantage of the fact that almost half of the computed variables in the algorithm in [8] are equal to zero, as $H^i(j)$ and $\hat{H}^i(j+1)$ are zero for $j < i$, to reduce the number of computations. The algorithm to compute the matrix H from the generalized reflection coefficients is then:

Initialization: $H^0(0) = c_{0,0}$.

For $j = 1, \dots, n$:

$\hat{H}^j(j) = [0, \dots, 0]$

For $i = j-1, \dots, 0$:

$\hat{H}^i(j) = \hat{H}^{i+1}(j) - H^i(j-1) K_{i+1}^T$

$H^0(j) = \hat{H}^0(j) [1, 0, \dots, 0]^T (= c_{j,0})$

For $i = 0, \dots, j-1$:

$H^{i+1}(j) = H^i(j-1) + \hat{H}^i(j) \Delta_i^{-1} K_{i+1}$.

In this algorithm $\hat{H}^i(j)$ is the j th row of \hat{H}^i , or in other words a row vector made up of $(\alpha+1)$ elements. When the algorithm is run in conjunction with the recursion for U^i , and \hat{U}^i in (4.5), and the computation of K_i in (4.8), then we have a complete QR algorithm for computing U and H in the QR factorization $Y = UH^T$. This algorithm generates the matrix H , row by row, and H^T in the QR factor $Y = UH^T$, column by column. This is the opposite of what happened in the coupled vector recursions.

VI. CONCLUSION

We have derived vector algorithms for the QR and Cholesky factorization of Toeplitz and close-to-Toeplitz matrices in all the cases of linear prediction. The same coupled recursions are used in all the algorithms, namely

$$N^{i+1} = N^i + ZM^i K_{i+1}^T$$

$$M^{i+1} = ZM^i + N^i \Delta_i^{-1} K_{i+1}.$$

The vector M^i contains the i th column of the matrix A , U , or HD^2 , depending upon which factorization is being computed. The inner products required to compute the reflection coefficients and to initialize the variables are summarized as follows:

- if $M^i = A^i$, then inner products are required for computing $c_{i,0}$ and K_i ;
- if $M^i = U^i$, then inner products are required for computing K_i only;
- if $M^i = H^i$, then inner products are required for computing $c_{i,0}$ only.

The Cholesky algorithms have complexity $n^2(\alpha + 2)$ if the experimental covariance is precomputed, and complexity $(N + n)n(\alpha + 2)$ if the experimental covariance must be computed from data. The fast algorithms for the orthogonal matrix U have complexity $Nn\alpha$, where N is the number of data values available.

REFERENCES

- [1] B. Friedlander, M. Morf, T. Kailath, and L. Ljung, "New inversion formulas for matrices classified in terms of their distance from Toeplitz matrices," *Linear Algebra Appl.*, vol. 27, pp. 31-60, 1979.
- [2] J. Le Roux and C. J. Gueguen, "A fixed point computation of partial correlation coefficients," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, pp. 257-259, June 1977.
- [3] C. J. Demeure and L. L. Scharf, "Linear statistical models for stationary sequences and related algorithms for Cholesky factorization of Toeplitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 29-42, Jan. 1987.
- [4] C. P. Rialan and L. L. Scharf, "Fast algorithms for QR and Cholesky factors of Toeplitz operators," in *IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing* (Dallas, TX), Apr. 1987, pp. 41-44; also *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 11, pp. 1740-1748, Nov. 1988.
- [5] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, no. 8, pp. 829-867, Aug. 1982.
- [6] M. Morf, B. Dickinson, T. Kailath, and A. Vieira, "Efficient solution of covariance equations for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, no. 5, pp. 429-433, Oct. 1977.
- [7] E. A. Robinson and S. Treitel, "Maximum entropy and the relationship of the partial autocorrelation to the reflection coefficients of a layered system," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 2, pp. 224-235, Apr. 1980.
- [8] B. Friedlander, "Efficient computation of the covariance sequence of an autoregressive process," in *IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing* (Boston, MA), Apr. 1983, pp. 182-185.
- [9] C. J. Demeure and L. L. Scharf, "Vector algorithms for computing QR and Cholesky factors of close-to-Toeplitz matrices," in *IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing* (Dallas, TX), Apr. 1987, pp. 1851-1854.
- [10] G. Cybenko, "A general orthogonalization technique with applications to time series analysis and signal processing," *Math. Comput.*, vol. 40, no. 161, pp. 323-336, Jan. 1983.
- [11] N. Levinson, "The Wiener rms criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261-278, 1947.
- [12] J. Durbin, "The fitting of time series models," *Revue l'Institut Int. Statistique*, vol. 28, pp. 233-243, 1960.
- [13] L. Marple, "A new autoregressive spectrum analysis algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 4, pp. 441-454, Aug. 1980.
- [14] G. Kunitz and I. d'Erceville, "Sur certaines propriétés d'une onde acoustique plane de compression dans un milieu stratifié," *Annal. Géophysique*, vol. 18, no. 4, pp. 1-9, 1982.
- [15] C. J. Demeure and L. L. Scharf, "Lattice algorithms for computing QR and Cholesky factors in the least squares theory of linear prediction," ECE Dep., Univ. of Colorado, Boulder, CO, Tech. Rep., Sept. 1987.
- [16] G. Cybenko, "Fast Toeplitz orthogonalization using inner products," *SIAM J. Sci. Stat. Comput.*, vol. 8, no. 5, pp. 734-740, Jan. 1983.
- [17] I. Gohberg, T. Kailath, I. Koltracht, and P. Lancaster, "Linear complexity parallel algorithms for linear systems of equations with recursive structure," *Linear Algebra Appl.*, vol. 88, pp. 271-315, 1987.

Time-Frequency Distributions for a Wide-Sense Stationary Random Signal

THEODORE E. POSCH

Abstract—This correspondence considers the time-frequency distribution for a wide-sense stationary random signal, and derives a simple criterion for when a bilinear time-frequency distribution gives the power spectrum of the signal.

INTRODUCTION

There have been many time-frequency distributions that have been proposed for the description of a signal in time and frequency. Among them are the Wigner [1], Page [2], Rihaczek [3], and the spectrogram, etc. The general theory of distributions was developed by Cohen [4], where an explicit method is given for generating distributions. Cohen introduced the concept of a kernel: by taking specific functions for the kernel different distributions are generated. Furthermore, by constraining the kernel in particular ways, distributions with given properties are obtained. Desirable properties of distributions and the corresponding constraints of the kernel have been studied by many, for example, Claassen and Mecklenbrauker in [5]. Recently, an important contribution has been made by Choi and Williams [6] where they developed the constraints to be imposed on the kernel so as to reduce the cross terms and have used a specific kernel to generate a new distribution with reduced interference terms.

Most of the applications of these distributions have been to deterministic signals, although there has been work on their applications to random signals. Indeed, some of the early papers on deterministic signals did consider the random case. Page [2] used his distribution to study random signals and Ackroyd [7] used the Rihaczek distribution for the study of random signals in addition to applying it to deterministic signals. Grace [8] considered the random signal case for the instantaneous power spectra. Recently, general approaches have been given by Martin [9], Martin and Flandrin [10], White [11], and White and Boashash [12]. White and Boashash [12] have developed a method for obtaining the random instantaneous frequency of a Gaussian random signal by use of the Wigner-Ville distribution. This procedure is an extension of the method developed by Boashash [13], for deterministic signals.

The purpose of this correspondence is to show that a wide class of time-frequency distributions gives the power spectrum for the case of a wide-sense stationary random signal. We will obtain a simple criteria on the Cohen kernel for determining whether that is the case for a particular distribution.

The Cohen class [4] is given by

$$C(t, \omega) = \frac{1}{2\pi} \int \int \int e^{-j\theta t - j\tau \omega + j\theta u} \phi(\theta, \tau) f\left(u + \frac{\tau}{2}\right) \cdot f^*\left(u - \frac{\tau}{2}\right) d\theta d\tau du \quad (1)$$

Manuscript received February 17, 1989; revised June 22, 1989.
The author is with Hughes Aircraft Company, Buena Park, CA 92633.
IEEE Log Number 8934017.