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 $Rb = Q' b$.

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 $y = \{y_0, y_1, \ldots, y_n\}$ 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 $A(z)$ for the time series $\{y_t\}$. This model takes the form

$$\sum_{i=0}^n a_i^2 y_{t-i} = a_0^2, \quad a_0^2 = 1 \tag{2.1}$$

where $\{a_i^2\}$ is an error sequence. The model may also be written as the predictor model

$$y_t = \hat{y}_t + a_0^2, \quad \hat{y}_t = -\sum_{i=1}^n a_i^2 y_{t-i} \tag{2.2}$$

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

$$\sigma_t^2 = \sum_{i=L} a_i^2 \tag{2.3}$$

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 $A(z)$ will be to form a sequence of predictions of the form

$$\hat{y}_t = -\sum_{i=1}^n a_i^2 y_{t-i}; \quad a_0^2 = 1 \tag{2.4}$$

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

$$t \in T_n \Leftarrow (n - p) \leq t \leq n \tag{2.5}$$

As the predictor order increases from $p = 0$ to $p = n$, a window of constant length $(k + 1)$ moves from left to right across the data set, as illustrated in Fig. 1. The indexes $k$ and $l$ may be chosen...
to select among the various methods of linear prediction (just as in [6]). The data values outside the range \([0, \cdots, N-1]\) 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:

\[
\begin{bmatrix}
1 & x_1 & x_1 & \cdots & \cdots & x_1 & x_{1+p} & x_{1+p} & \cdots & x_{1+p} \\
1 & x_2 & x_2 & \cdots & \cdots & x_2 & x_{2+p} & x_{2+p} & \cdots & x_{2+p} \\
\vdots & \vdots & \vdots & \cdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\
1 & x_n & x_n & \cdots & \cdots & x_n & x_{n+p} & x_{n+p} & \cdots & x_{n+p} \\
1 & x_{n+1} & x_{n+1} & \cdots & \cdots & x_{n+1} & x_{n+1+p} & x_{n+1+p} & \cdots & x_{n+1+p} \\
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_1 \\
y_2 \\
y_2 \\
y_n \\
y_n \\
\end{bmatrix}
= \begin{bmatrix}
a_1 \\
a_1 \\
a_1 \\
a_1 \\
a_1 \\
\end{bmatrix}
\begin{bmatrix}
y_{1-p} \\
y_{1-p} \\
y_{2-p} \\
y_{2-p} \\
y_{n-p} \\
\end{bmatrix}
\]

The compact notation is

\[
Y a^p = Y A^p = U^p
\]

where \(Y\) is the data matrix, \(a^p\) 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 \(Y A^p = U^p\), where \(Y\) is the Toeplitz data matrix just defined, and the matrices \(A\) and \(U\) are given by

\[
U = [U^0, \cdots, U^p, \cdots, U^n].
\]

Any particular choice for \(k\) and \(l\) slices the infinite Toeplitz matrix built on the infinite data sequence \(\cdots, 0, y_0, y_1, \cdots, y_{n-1}, \cdots\)

\[
A^p = \begin{bmatrix}
a^p \\
0
\end{bmatrix}
\]

C. Least Squares

The least squares solution for \(a\) in \(Ya = 0\), with the constraint \(a_0 = 1\) is

\[
Ca = Y^T Y a = e^T [0, \cdots, 0, 1]^T.
\]

When this solution is written out for orders \(p = 0, 1, \cdots, n\), the result is

\[
CA = Y^T Y A = HD^2
\]

where \(H = (A^T)^{-1}\) is a lower triangular matrix and \(D^2\) is the diagonal matrix containing the prediction errors \(D^2 = \text{diag} \{\sigma_0, \sigma_1^2, \cdots, \sigma_n^2\}\). 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 \(YA = Q\) 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 \(YA = U\), the Cholesky factor \(A^T CA = D^2\), and the Cholesky factor \(C = HD^2H^T\) 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

\[
A^T CA = D^2 = \text{diag} \{\sigma_0^2, \sigma_1^2, \cdots, \sigma_n^2\}.
\]

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

\[
C a^p = \sigma_p [0, \cdots, 0, 1]^T
\]

Here \(C_i\) is the \((i + 1)\)th 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 \(a_{i+1}\) on \(a_i\). This was the insight of Friedlander et al. [1].

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

$$\sigma_{1}^2K_{i+1} = -(ZA^T)\Sigma$$

(3.8)

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

$$\sigma_{1}^2\Delta_{i+1} = \sigma_{1}^2(1 - K_{i+1}^T\Delta_{i}^{-1}K_{i+1})$$

(3.9)

Note that, as only $\Delta_{i}$ is used in the equations, it might be more numerically advantageous to use the alternate update equation

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

with the initialization

$$\Delta_{0} = \begin{bmatrix} 1 & 0^T \\ 0 & -\sigma_{1}^2 \Sigma \end{bmatrix}$$

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 $YA = U$. For the correlation method of linear prediction, this procedure is [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 $W$:

$$W = \begin{bmatrix} y_{n-1} & 0 & \cdots & 0 \\
y_{n-2} & y_{n-1} & \cdots & 0 \\
y_{n-3} & \vdots & \ddots & \vdots \\
y_{n-k} & \vdots & \ddots & 0 \\
y_{0} & y_{n-1} & y_{n-2} & y_{n-3} \\
y_{1} & \vdots & \ddots & \vdots \\
y_{n-2} & \vdots & \ddots & \vdots \\
y_{n-1} & \vdots & \ddots & y_{0} \end{bmatrix}$$

(4.1)

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

$$WA = \begin{bmatrix} \hat{U} \\ U \end{bmatrix}$$

(4.2)

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

$$WA^{i} = \begin{bmatrix} \hat{U}^{i} \\ U^{i} \end{bmatrix}$$

(4.3)

Similarly define

$$WA^{i + 1} = \begin{bmatrix} \hat{U}^{i + 1} \\ U^{i + 1} \end{bmatrix}$$

(4.4)

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

$$WA^{i + 1} = \begin{bmatrix} \hat{U}^{i + 1} \\ U^{i + 1} \end{bmatrix} = Z^{i}U^{i} + \hat{U}^{i + 1}\hat{K}_{i+1}$$

$$WA^{i} = \begin{bmatrix} \hat{U}^{i + 1} \\ U^{i + 1} \end{bmatrix} = Z^{i}U^{i} + \hat{U}^{i}\hat{K}_{i+1}$$

(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 \( A' \) is equal to zero for \( i < n \), in which case \( \text{ZYA}' = WZA' \) for our definition of \( W \).

Here the shifted difference matrix \( \delta[C] \) has rank 2, and
\[
E = \begin{bmatrix} y_0 & \cdots & y_{n-1} \\ y_{n-1} & \cdots & y_0 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\]
(4.6)

The recursions are initialized as follows:
\[
q^0 \text{ is initialized column by column in } Y.
\]
(7.2)

We obtain a formula for \( CZA' \) by using (3.4):
\[
CZA' = ZH' - \begin{bmatrix} 1 & 0' \\ 0 & E \end{bmatrix} K_{i+1} \sigma_i^2.
\]
(5.3)

so that together with (3.9), we get the recursions:
\[
H^{i+1} = \tilde{H}^i + ZH' K_{i+1}^T, \quad H^{i+1} = ZH' + \tilde{H} K_{i+1},
\]
for \( i = 0, \ldots, n - 1 \).

These recursions are initialized as follows:
\[
H^0 = \begin{bmatrix} c_{0,0} & c_{1,0} & \cdots & c_{n,0} \end{bmatrix}^T, \quad \tilde{H}^0 = \begin{bmatrix} c_{0,0} & \cdots & c_{n,0} \end{bmatrix}^T
\]
(5.5)

\[
\sigma_0^2 = c_{0,0}, \quad \Delta_0 = \begin{bmatrix} 1 & 0' \\ 0 & -\sigma_0^2 \Sigma \end{bmatrix}.
\]
(4.7)

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

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 \([c_{0,0}, 0, \ldots, 0]\). The result originates from the work of Kunitz and d’Ercievile [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 \( HD^2 = [H_0] \), we take advantage of the fact that almost half of the computed variables in the algorithm in [8] are equal to zero, as \( H'(j) \) and \( H'(j + 1) \) are zero for \( j < i \), to reduce the number of computations. The algorithm to compute the matrix \( H \) from the 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
\[
N^{i+1} = N' + ZM' K_{i+1}^2, \quad \Delta^{i+1} = ZM' + N' \Delta_i^2 K_{i+1}.
\]
The vector $M'$ contains the $i$th column of the matrix $A$, $U$, or $HD^T$, 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 $M' = A'$, then inner products are required for computing $c_o$ and $K_i$;
- if $M' = U'$, then inner products are required for computing $K_i$ only;
- if $M' = H'$, then inner products are required for computing $c_o$ only.

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

References


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, Claasen 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 criterion on the Cohen kernel for determining whether that is the case for a particular distribution.

The Cohen class [4] is given by

$$C(t, f) = \frac{1}{2\pi} \int \left( e^{-jz}\int e^{-j(z-u)}\varphi(0, t) f \left( \frac{u + z}{2} \right) \right) \ast f^* \left( \frac{u - z}{2} \right) d\theta d\tau du$$

(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.