Second European Congress on Operations Research
November 29 - December 1, 1976, Stockholm
North-Holland Publishing Company, 1976

LINEAR LEAST-SQUARES PREDICTION BASED ON COVARIANCE DATA FROM STATIONARY PROCESSES WITH FINITE-DIMENSIONAL REALIZATIONS*

Anders Lindquist
Department of Mathematics
University of Kentucky
Lexington, Kentucky 40506

Given a (wide sense) stationary time series $\{y_0, y_1, y_2, \ldots\}$, consider the problem to determine (for each $t = 1,2,3,\ldots$) the linear least-squares estimate $\hat{y}_t$ of y_t given the observed data $\{y_0, y_1, \ldots, y_{t-1}\}$. This is the one-step prediction problem. Suppose that $E\{y_t\} = 0$ and $E\{y_{t+i}y_t\} = c_i$, where the autocorrelation sequence $\{c_0, c_1, c_2, \ldots\}$ satisfies the recursion

$$c_{n+i} + \sum_{j=1}^{n} a_j c_{n-j+i} = 0 ; \qquad i = 1,2,3,\ldots$$

for some fixed n together with a certain positivity condition and a rank condition. The second-order properties of the process y are hence completely characterized by the $2n+1$ parameters

$$\{c_0, c_1, \ldots, c_n, a_1, a_2, \ldots, a_n\} \tag{1}$$

and the problem is to determine a predictor in terms of these.

It is shown that y can be described by a Gauss-Markov model

$$\begin{cases} x_{t+1} = Fx_t + v_t \\ y_t = hx_t + w_t \end{cases} \tag{2}$$

where x is an n-dimensional process, F is a nxn matrix, h is a row vector, and (v,w) are white noise sequences. Consequently, the required estimate can be determined by means of the Kalman filter

$$\begin{cases} \hat{x}_{t+1} = F\hat{x}_t + k_t(y_t - \hat{y}_t) ; \qquad \hat{x}_0 = 0 \\ \hat{y}_t = h\hat{x}_t \end{cases} \tag{3}$$

and hence the problem is reduced to determining the sequence $\{k_0, k_1, k_2, \ldots\}$ *from the parameters* (1).

Traditionally this problem has been solved by first determining all parameters of the model (2) from the covariance data (1) and then solving a matrix Riccati equation to obtain k. Here we shall present an alternative method, based on some previous work by the author, which does not require complete knowledge of (2). In fact, we only need (h,F) which is immediately available in terms of the covariance data (1). Moreover, the algorithm is computationally more efficient than schemes based on the Riccati equation.

1. FORMULATION OF THE PROBLEM

Let $\{y_0, y_1, y_2, \ldots\}$ be a (wide sense) stationary stochastic process with statistics

$$\begin{cases} E\{y_t\} = 0 \\ E\{y_{t+i}y_t\} = c_i . \end{cases} \tag{1.1}$$

Suppose that the covariance data $\{c_0, c_1, c_2, \ldots\}$ satisfies the following three conditions.

(i) For each $t = 1,2,3,\ldots$ the *Toeplitz matrix*

$$T_t = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{t-1} \\ c_1 & c_0 & c_1 & \cdots & c_{t-2} \\ c_2 & c_1 & c_0 & \cdots & c_{t-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{t-1} & c_{t-2} & c_{t-3} & \cdots & c_0 \end{bmatrix} \tag{1.2}$$

* This work was supported by the National Science Foundation under grant MPS 75-07028.

is positive definite.

(ii) There exists a set of finitely many real numbers $a_1, a_2, \ldots, a_n$ with the property that the polynomial

$$z^n + a_1 z^{n-1} + a_2 z^{n-2} + \cdots + a_n \tag{1.3}$$

has all its zeros inside the unit circle and such that

$$c_{n+i} + \sum_{j=1}^{n} a_j c_{n-j+1} = 0 ; \quad i = 1,2,3,\ldots \tag{1.4}$$

(iii) Let n be the integer defined in (ii). Then the *Hankel matrix*

$$H = \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_n \\ c_2 & c_3 & c_4 & \cdots & c_{n+1} \\ c_3 & c_4 & c_5 & \cdots & c_{n+2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_n & c_{n+1} & c_{n+2} & \cdots & c_{2n} \end{bmatrix} \tag{1.5}$$

has full rank.

Condition (i) ensures that the process is purely stochastic or, which is the same thing, has full rank (Masani, 1960), (ii) is a finiteness condition which provides the problem with a Markov structure, and (iii) implies that n is the smallest integer with property (ii). Note that the second-order properties of y are completely characterized by the $2n + 1$ parameters $\{c_0, c_1, \ldots, c_n, a_1, \ldots, a_n\}$.

Consider the problem to determine the linear least-squares estimate $\hat{y}_t$ of y_t given the observed data $\{y_0, y_1, y_2, \ldots, y_{t-1}\}$, i.e. find the linear combination

$$\hat{y}_t = -\varphi_{t1} y_{t-1} - \varphi_{t2} y_{t-2} - \cdots - \varphi_{tt} y_0 \tag{1.6}$$

(the minus signs are introduced to simplify notations below) which minimizes

$$r_t = E(y_t - \hat{y}_t)^2 . \tag{1.7}$$

For the moment disregarding conditions (ii) and (iii), we note that this is a classical problem in the theory of estimation, and it is well-known (Levinson, 1942) that the optimal weighing pattern $(\varphi_{t1}, \varphi_{t2}, \ldots, \varphi_{tt})$ satisfies the system of *normal equations*

$$\begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{t-1} \\ c_1 & c_0 & c_1 & \cdots & c_{t-2} \\ c_2 & c_1 & c_0 & \cdots & c_{t-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{t-1} & c_{t-2} & c_{t-3} & \cdots & c_0 \end{bmatrix} \begin{bmatrix} \varphi_{tt} \\ \varphi_{t,t-1} \\ \cdot \\ \cdot \\ \varphi_{t1} \end{bmatrix} = - \begin{bmatrix} c_t \\ c_{t-1} \\ \cdot \\ \cdot \\ c_1 \end{bmatrix} \tag{1.8}$$

and that the variance (1.7) of the estimation error is given by

$$r_t = \sum_{i=0}^{t} c_i \varphi_{ti} \tag{1.9}$$

where $\varphi_{t0} = 1$. In fact, this follows from a simple orthogonality argument.

Condition (i) ensures that eq. (1.8) has a unique solution, and therefore the required weighting pattern is immediately at hand. For large t, however, the computational burden in solving eq. (1.8) will be significant. Also, in general, we are not interested in one particular $\hat{y}_t$. Instead, we wish to determine the whole sequence $\hat{y}_0, \hat{y}_1, \hat{y}_2, \ldots$ "on line" as additional data is received. This raises the question whether there is a recursive technique by which the coefficients of $\hat{y}_{t+1}$ can be determined from those of $\hat{y}_t$. Indeed, such an algorithm has been developed by Levinson (1942), and §2 will be devoted to this topic. The same recursions can also be found in the theory of polynomials orthogonal on the unit circle (Grenander and Szegö 1958, Geronimus 1961, Akhiezer 1965) and, in a somewhat more general form, in Whittle (1963) and Wiggins & Robinson (1965).

Introducing conditions (ii) and (iii) admits considerable simplifications of the computational procedure. In fact, a (vector-valued) Gauss-Markov model for y can be constructed, as we shall demonstrate in §4. Consequently, $\hat{y}_t$ can be generated by the *Kalman filter*

$$\begin{cases} \hat{x}_{t+1} = F\hat{x}_t + k_t(y_t - h\hat{x}_t) ; & \hat{x}_0 = 0 \\ \hat{y}_t = h\hat{x}_t , \end{cases} \tag{1.10}$$

where $\{\hat{x}_0, \hat{x}_1, \hat{x}_2, \ldots\}$ is an n-vector valued process, F is the nxn-matrix

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_1 & -a_2 & -a_3 & \cdots & -a_n \end{bmatrix} , \tag{1.11}$$

h is the n-dimensional row vector

$$h = (1, 0, 0, \ldots, 0) , \tag{1.12}$$

and k is an n-vector valued sequence, called the gain, which remains to be determined.

It is the purpose of this note to present an algorithm by which the gain sequence k can be computed directly in terms of the $2n + 1$ parameters $\{c_0, c_1, \ldots, c_n, a_1, a_2, \ldots, a_n\}$. To this end we could proceed along more traditional routes by first expressing all the parameters of the Gauss-Markov model mentioned above in terms of $\{c_0, c_1, \ldots, c_n, a_1, a_2, \ldots, a_n\}$ and and then solving the corresponding nxn-matrix Riccati equation (Mehra 1971). However, we have recently developed an alternative procedure (Lindquist 1974) which is more easily adaptable to the available parameters and which is computationally more efficient. In §5 we shall present a simple derivation of this result based on the Levinson recursions. With certain modifications of our procedure we could instead use another algorithm, similar to that in Lindquist (1974), which was independently obtained by Rissanen (1973).

2. LEVINSON'S EQUATIONS

Given the weighting pattern of $\hat{y}_t$, consider the problem to determine $\hat{y}_{t+1}$. The normal equations for $\hat{y}_{t+1}$ are

$$\left[\begin{array}{c|ccccc} c_0 & c_1 & c_2 & \cdots & c_t \\ \hline c_1 & c_0 & c_1 & \cdots & c_{t-1} \\ c_2 & c_1 & c_0 & \cdots & c_{t-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_t & c_{t-1} & c_{t-2} & \cdots & c_0 \end{array}\right] \left[\begin{array}{c} \varphi_{t+1,t+1} \\ \hline \varphi_{t+1,t} \\ \cdot \\ \cdot \\ \varphi_{t+1,1} \end{array}\right] = - \left[\begin{array}{c} c_{t+1} \\ \hline c_t \\ \cdot \\ \cdot \\ c_1 \end{array}\right] \tag{2.1}$$

By partitioning the matrices as marked, eq. (2.1) can be written

$$\begin{bmatrix} c_0 & \bar{c}_t' \\ \bar{c}_t & T_t \end{bmatrix} \begin{bmatrix} -\gamma_t \\ z \end{bmatrix} = - \begin{bmatrix} c_{t+1} \\ \underline{c}_t \end{bmatrix} \tag{2.2}$$

(prime denotes transpose) or

$$\begin{cases} -\gamma_t c_0 + \bar{c}_t' z = -c_{t+1} & (2.3) \\ -\gamma_t \bar{c}_t + T_t z = -\underline{c}_t\ , & (2.4) \end{cases}$$

where the definitions of $\bar{c}$, $\underline{c}$, γ and z are clear from comparing eq. (2.1) and eq. (2.2). Equation (2.4) yields

$$z = -T_t^{-1}\underline{c}_t + \gamma_t T_t^{-1}\bar{c}_t\ , \tag{2.5}$$

which, inserted into eq. (2.3), gives

$$\gamma_t(c_0 - \bar{c}_t' T_t^{-1}\bar{c}_t) = c_{t+1} - \bar{c}_t' T_t^{-1}\underline{c}_t\ . \tag{2.6}$$

Now, observing that, in view of eq. (1.8) and eq. (1.9),

$$\begin{cases} -T_t^{-1}\underline{c}_t = (\varphi_{tt}, \varphi_{t,t-1}, \ldots, \varphi_{t1})' & (2.7) \\ -T_t^{-1}\bar{c}_t = (\varphi_{t1}, \varphi_{t2}, \ldots, \varphi_{tt})' & (2.8) \\ r_t = c_0 - \bar{c}_t' T_t^{-1}\bar{c}_t = c_0 - \underline{c}_t' T_t^{-1}\underline{c}_t\ , & (2.9) \end{cases}$$

we are in a position to state the following result.

Proposition 2.1 (Levinson): *Let* $\varphi_{t0} = 1$ *and* $\varphi_{t,t+1} = 0$. *Then, for each* t *and* $i = 0,1,2,\ldots,n$, *the weighting coefficients are generated by the recursions*

$$\varphi_{t+1,i} = \varphi_{ti} - \gamma_t \varphi_{t,t+1-i}\ ; \qquad \varphi_{00} = 1 \tag{2.10}$$

and the error variances by

$$r_{t+1} = (1 - \gamma_t^2) r_t\ ; \qquad r_0 = c_0 \tag{2.11}$$

where

$$\gamma_t = \frac{1}{r_t} \sum_{i=0}^{t} \varphi_{t,t-i} c_{i+1}\ . \tag{2.12}$$

Here eq. (2.10) follows from eq. (2.5) and the definition of γ, and eq. (2.12) is the same as eq. (2.6). To see that eq. (2.11) is true, use eq. (2.5) and the second part of eq. (2.9).

Remark 2.1: Proposition 2.1 can be generalized to the case where y is a vector-valued process (Whittle 1963, Wiggins and Robinson 1965, Rissanen 1973b, Lindquist 1974). In modifying the derivation above we must remember, however, that the parameters $c_0, c_1, c_2, \ldots$ are now matrices, and therefore we must partition two sets of normal equations, one for the problem at hand and one for a "backward problem" in which $c_0, c_1, c_2, \ldots$ have been exchanged for the transposed matrices $c_0', c_1', c_2', \ldots$. □

3. SZEGÖ'S ORTHOGONAL POLYNOMIALS

It is convenient to represent the weighting pattern of $\hat{y}_t$ by means of the generating function

$$\varphi_t(z) = z^t + \varphi_{t1} z^{t-1} + \varphi_{t2} z^{t-2} + \cdots + \varphi_{tt}\ . \tag{3.1}$$

Grenander and Szegö (1958) have shown that the polynomials $\varphi_0, \varphi_1, \varphi_2, \ldots$ are orthogonal on the unit circle, and consequently, we can take advantage of the theory of such polynomials, initiated by Szegö. Defining the reversed polynomials

$$\varphi_t^*(z) = \varphi_{tt} z^t + \varphi_{t,t-1} z^{t-1} + \varphi_{t,t-2} z^{t-2} + \cdots + 1\ , \tag{3.2}$$

equation (2.10) can be written

$$\begin{cases} \varphi_{t+1}(z) = z\varphi_t(z) - \gamma_t \varphi_t^*(z)\ ; & \varphi_0(z) = 1 & (3.3) \\ \varphi_{t+1}^*(z) = \varphi_t^*(z) - \gamma_t z \varphi_t(z)\ ; & \varphi_0^*(z) = 1\ . & (3.4) \end{cases}$$

The recursions (3.3)–(3.4) can be found in the theory of orthogonal polynomials (Geronimus 1961, Akhiezer 1965) together with the following result.

Proposition 3.1 (Akhiezer): *Condition* (i) *in* §1 *holds if and only if* $|\gamma_t| < 1$ *for* $t = 0,1,2,\ldots$.

Hence, in view of eq. (2.11), condition (i) implies that r_t is positive for all t, and hence, division by this quantity is permitted.

4. A GAUSS-MARKOV REPRESENTATION

The following result (Lemma 4.1) is based on a version of the Positive Real Lemma (Faurre 1973).

Lemma 4.1: *Let* F *and* h *be given by eqs.* (1.11) *and* (1.12) *and let*

$$c = (c_1, c_2, \ldots, c_n)' . \tag{4.1}$$

Suppose that conditions (i)–(iii) *in* §1 *hold. Then there exist two matrices* P *and* Q, *an n-dimensional vector* d *and a scalar* α *with* $\begin{pmatrix} Q & d \\ d' & \alpha \end{pmatrix}$ *and* P *nonnegative definite and such that*

$$\begin{cases} P = FPF' + Q & (4.2) \\ c = FPh' + d & (4.3) \\ c_0 = hPh' + \alpha . & (4.4) \end{cases}$$

Moreover,

$$c_i = hF^{i-1}c ; \qquad i = 1,2,3,\ldots \tag{4.5}$$

Proof. (a) From eqs. (1.11) and (1.12) we have

$$\begin{bmatrix} h \\ hF \\ \cdot \\ \cdot \\ \cdot \\ hF^{n-1} \end{bmatrix} = I \qquad \text{(identity matrix).} \tag{4.6}$$

Hence (h,F) is *observable* (Brockett 1970).
(b) Relation (1.4) implies

$$(c, Fc, F^2c, \ldots, F^{n-1}c) = H \tag{4.7}$$

where H is the Hankel matrix (1.5). Hence, by condition (iii), (F,c) is *controllable* (Brockett 1970).
(c) The matrix F is a *stability matrix*, for the polynomial (1.3) has all its zeros inside the unit circle (condition (ii)).
(d) For i = 1,2,3, ...

$$F^{i-1}c = (c_i, c_{i+1}, c_{i+2}, \ldots, c_{i+n-1})' , \tag{4.8}$$

and consequently eq. (4.5) holds.
(e) By the Positive Real Lemma (Faurre 1973), partial results (a), (b), (c), condition (i) and eq. (4.5) imply the existence of a quadruple (P,Q,d,α) with the properties required in the lemma. □

As far as the problem of §1 is concerned, we can regard stochastic processes with the same second-order properties as equivalent. With this in view, we have the following representation for the process y.

Lemma 4.2: *The stochastic process* y *defined in* §1 *can be generated by the Gauss-Markov model*

$$\begin{cases} x_{t+1} = Fx_t + v_t \\ y_t = hx_t + w_t \end{cases} \tag{4.9}$$

where (F,h) *are defined by eqs.* (1.11) *and* (1.12), v *and* w *are white noise processes with zero mean and covariances*

$$E\left\{\begin{pmatrix} v_t \\ w_t \end{pmatrix}(v_s', w_s')\right\} = \begin{cases} \begin{pmatrix} Q & d \\ d' & \alpha \end{pmatrix} & \text{for } s = t \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

and x_0 *is a zero mean stochastic vector with*

$$E\{x_0x_0'\} = P . \tag{4.11}$$

Proof. A straightforward application of eq. (4.9) yields

$$E\{x_ty_i\} = E\{x_tx_i'\}h' + E\{x_tw_i\} = F^{t-i}Ph' + F^{t-i-1}d$$

for $t > i$. Consequently, by eq. (4.3),

$$E\{x_ty_i\} = F^{t-i-1}c \quad \cdot \text{ for } t > i . \tag{4.12}$$

Hence, in view of eq. (4.5), $E\{y_{t+i}y_t\} = c_i$ for $i = 1,2,3,\ldots$. Likewise $E\{y_t^2\} = hPh' + \alpha$, which, in view of eq. (4.4), equals c_0. Therefore, (1.1) holds, for obviously $E\{y_t\} = 0$. □

We can now apply Kalman filtering techniques. The following result can be found in standard text books on systems and control.

Proposition 4.1: *Let* y *be defined by eq.* (4.9). *Then the filtered process* $\hat{y}$ *is generated by eq.* (1.10) *with* k *given by*

$$k_t = \frac{1}{r_t}(F\,\Sigma_t h' + d) , \tag{4.13}$$

where Σ_t *is the covariance matrix*

$$\Sigma_t = E\{(x_t - \hat{x}_t)(x_t - \hat{x}_t)'\} \tag{4.14}$$

and r_t *is the error variance* (1.7).

Now, given the parameters (P,Q,d,α), the error covariance Σ_t could be determined from the nxn matrix Riccati equation

$$\begin{cases} \Sigma_{t+1} = F[\Sigma_t - (\Sigma_t h' + d)(h\Sigma_t h' + \alpha)^{-1}(\Sigma_t h' + d)']F' + Q \\ \Sigma_0 = P . \end{cases} \tag{4.15}$$

However, so far we have only proved the *existence* of a model (4.9); we do not actually know (P,Q,d,α). There are ways to determine these parameters (Mehra 1971), but here we shall use a procedure which avoids this problem altogether and which, moreover, requires considerably fewer arithmetic operations.

5. THE MAIN RESULT

Theorem 5.1: *Let the process* y *be defined as in* §1, *and assume that conditions* (i)–(iii) *hold. Let* F, h *and* c *be given by eqs.* (1.11), (1.12) *and* (4.1). *Then the linear least-squares estimate* $\hat{y}_t$ *of* y_t *given* $\{y_0, y_1, y_2, \ldots, y_{t-1}\}$ *is generated by*

$$\begin{cases} \hat{x}_{t+1} = F\hat{x}_t + k_t(y_t - \hat{y}_t); \quad \hat{x}_0 = 0 \\ \hat{y}_t = h\hat{x}_t , \end{cases} \tag{5.1}$$

where k *is given by the 2n difference equations*

$$\begin{cases} k_{t+1} = [1-(hk_t^*)^2]^{-1}[k_t - (hk_t^*)Fk_t^*]; \quad k_0 = c & (5.2) \\ k_{t+1}^* = [1-(hk_t^*)^2]^{-1}[Fk_t^* - (hk_t^*)k_t]; \quad k_0^* = c & (5.3) \end{cases}$$

Remark 5.1: The algorithm (5.2)–(5.3) can be restated in a computationally more efficient form (Lindquist 1974, 1975). We have chosen the present formulation since it fits very nicely with the theory of §§2 and 3. □

Remark 5.2: There are some algebraic relations between k and k* which can be used to reduce the number of scalar first-order difference equations in the algorithm from 2n to n (Lindquist 1976). However, this reduction is bought at the expense of greater algebraic complexity and consequently we shall not go into this matter here. □

The proof of Theorem 5.1 will be based on the results of §§2 and 3 and we shall need the following lemma.

Lemma 5.1: *The gain sequence* (4.13) *is given by*

$$k_t = \frac{1}{r_t}\varphi_t^*(F)c \tag{5.4}$$

where $\varphi_t^*(F)$ *is the matrix polynomial obtained by exchanging* z *for* F *in eq.* (3.2).

Proof. Since $E\{\hat{x}_t(x_t - \hat{x}_t)'\} = 0$, eqs. (4.13) and (4.14) yield

$$k_t = \frac{1}{r_t}[FE\{x_t(x_t - \hat{x}_t)'h' + d]$$
$$= \frac{1}{r_t}[E\{Fx_t\tilde{y}_t\} + E\{v_t\tilde{y}_t\}],$$

where $\tilde{y}_t = y_t - \hat{y}_t$. Consequently,

$$k_t = \frac{1}{r_t}E\{x_{t+1}\tilde{y}_t\} . \tag{5.5}$$

Now, in view of eq. (1.6),

$$\tilde{y}_t = \sum_{i=0}^{t}\varphi_{ti}y_{t-i}$$

(for $\varphi_{t0} = 1$) which, inserted into eq. (5.5), yields

$$k_t = \frac{1}{r_t}\sum_{i=0}^{t}\varphi_{ti}E\{x_{t+1}y_{t-i}\} ,$$

which, in view of eq. (4.12), is the same as (5.4). □

Proof of Theorem 5.1. In analogy with eq. (5.4), we define

$$k_t^* = \frac{1}{r_t}\varphi_t(F)c , \tag{5.6}$$

where φ_t is given by eq. (3.1). Then, using eqs. (5.4), (3.4) and (2.11), we have

$$k_{t+1} = \frac{1}{r_{t+1}}\varphi_{t+1}^*(F)c$$
$$= (1-\gamma_t^2)^{-1}\frac{1}{r_t}[\varphi_t^*(F) - \gamma_t F\varphi_t(F)]c$$
$$= (1-\gamma_t^2)^{-1}(k_t - \gamma_t Fk_t^*) . \tag{5.7}$$

Likewise, eqs. (5.6), (3.3) and (2.11) can be used to obtain

$$k_{t+1}^* = (1-\gamma_t^2)^{-1}(Fk_t^* - \gamma_t k_t) . \tag{5.8}$$

It remains to determine γ_t. Inserting eq. (4.5) into eq. (2.12) yields $\gamma_t = \frac{1}{r_t}h\varphi_t(F)c$, which, in view of (5.6), can be written

$$\gamma_t = hk_t^* . \tag{5.9}$$

Then eqs. (5.2) and (5.3) follow from eqs. (5.7)–(5.9), for clearly $k_0 = k_0^* = c$. Note that Proposition 3.1 insures that $1 - \gamma_t^2 \neq 0$. □

6. CONCLUDING REMARKS

We have constructed a recursive prediction procedure based on covariance data from a stationary process with a finite-dimensional realization. Traditionally, this problem has been solved in two steps. The first step is the *realization problem* to construct a Gauss-Markov model from covariance data, and the second step is the *filtering problem.* Here we have combined these two steps in one. In many practical situations no covariance data is available, and consequently, the parameters of the model have to be estimated from statistical data. This is the *identification problem.* It is reasonable to assume that some covariance estimation procedure could be used together with the results of this paper to obtain a combined identification and filtering method. We shall investigate this matter further.

REFERENCES

N. I. Akhiezer (1965), **The Classical Moment Problem**, Hafner Publishing Co., New York.

R. W. Brockett (1970), **Finite-Dimensional Linear Systems**, Wiley.

P. Faurre (1973), Realisations markoviennes de processus stationaires, Research Report no. 13, Institute de Recherche d'Informatique et d'Automatique, Le Chesnay, France.

L. Ya Geronimus (1961), **Orthogonal Polynomials**, Consultant Bureau, New York.

U. Grenander and G. Szegö (1958), **Toeplitz Forms and Their Applications**, Univ. of California Press.

N. Levinson (1942), The Wiener RMS (root mean square) error in filter design and prediction, Appendix B in Wiener, **Extrapolation, Interpolation and Smoothing of Stationary Time Series**, M.I.T. Press, Cambridge, Massachusetts.

A. Lindquist (1974), A new algorithm for optimal filtering of discrete-time stationary processes, *SIAM J. Control* **12**, 736.

A. Lindquist (1975), Linear least squares estimation of discrete-time stationary processes by means of backward innovations, *Control Theory, Numerical Methods and Computer Systems Modelling,* A. Bensoussan and J. L. Lions, eds., Lecture Notes in Economics and Mathematical Systems **107**, 44.

A. Lindquist (1976), Some reduced-order non-Riccati equations for linear least squares estimation: The stationary, single-output case, *Intern. J. Control,* (to appear).

P. Masani (1960), The prediction theory of multivariate stochastic processes III, *Acta Math.* **104**, 141.

R. K. Mehra (1971), On-line identification of linear dynamic systems with applications to Kalman filtering, *IEEE Trans. Autom. Control* **AC-16**, 12.

J. Rissanen (1973a), A fast algorithm for optimum linear prediction, *IEEE Trans. Autom. Control* **AC-18**, 555.

J. Rissanen (1973b), Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with applications to factoring positive matrix polynomials, *Mathematics of Computation* **27**, 147.

P. Whittle (1963), On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix, *Biometrica* **50**, 129.

R. A. Wiggins and E. A. Robinson (1965), Recursive solutions to the multichannel filtering problem, *J. Geophys. Research* **70**, 1885.