Annals of Economic and Social Measurement, 6/3, 1977

RECENT RESULTS IN LEAST-SQUARES ESTIMATION THEORY†

BY M. MORF and T. KAILATH

Dynamic models have become of increasing interest in economics, with consequent attention to state-space models, quadratic control and least-squares Kalman filters. We present a survey of results in two new trends in the study of dynamic systems. One is the observation that while the Riccati-equation based Kalman filter has the advantage of applying equally to models with constant or time-variant parameters, substantial computational benefits can be obtained for constant models by using "fast" Chandrasekhar-type equations or square-root algorithms. These algorithms enable order-of-magnitude reductions in storage and computation, from $O(N^2)$ and $O(N^3)$ to $O(N)$ and $O(N^2)$ respectively. As an illustration we derive new fast algorithms for the well-known polynomial regression problem. The second group of results deals with the trend back to external input-output and transfer-function descriptions as a counter to the almost total concentration on state-space models in the recent literature. We have generalized the work of Levinson (1947) on efficient recursive methods for solving Toeplitz equations, by introducing the concept of the "distance from Toeplitz" of an arbitrary matrix and thereby obtaining recursive algorithms for general nonstationary processes. For state-space models, our new recursive algorithms can be reduced to the previously known Chandrasekhar and Riccati equations.

I. INTRODUCTION

Economists have become increasingly interested in dynamic models and consequently are paying more attention to state-space models, quadratic control and least-squares Kalman filtering and prediction. The study of dynamic systems and their use in control and estimation has been a very active field in recent years. The applications of these results have ranged from classical industrial process control, space-applications, and air pollution estimation, to identification and estimation in econometric systems, see, e.g., the special issue on identification and time series analysis [1], and the survey on linear filtering [2].

The emergence of state-space models in control and subsequently in linear estimation theory in the 1960's led to a voluminous literature on the so-called Riccati equation of the state-space-based Kalman filter [3], [4] and optimal (quadratic) control solutions, see, e.g., [5].

Although these solutions are elegant and widely used, with some hindsight we can quote several reasons for looking for alternatives:

—First, the computational complexity of these solutions might be prohibitive for large systems since the number of equations required per

†This work was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command under Contract AF44-620-74-C-0068, by the National Science Foundation under Contract NSF-Eng75-18952, and by the Joint Services Electronics Program under Contract N00014-75-C-0601.

time step is proportional to the *cube* of the number of states, the number of variables that are to be controlled or estimated. Economists are now studying systems with hundreds or thousands of variables; for such models a single observation would require millions of operations using standard optimal estimation or control solutions.

—Second, *time-invariant* models lead to *no* significant *simplification* of these solutions, since the solutions are generally valid for time-variant models.

—Third, these solutions require state-space models. The procedures to find them are *not* always trivial and conversion of an input-output model (that might be more readily available in economics) may not be desirable.

—Fourth, extensions of these optimal estimation and control solutions have not been too successful and alternative models and methods might be more appropriate.

Alternatives to Riccati-equation based solutions have actually been in existence in the past, but they have perhaps not been fully recognized or developed. For instance, "nice" recursive solutions to estimation problems were found, at least for stationary time series, by Levinson [7] in 1947, rediscovered in 1960 by Durbin [8] and extended by Whittle [9], [10] and Robinson [11]. Interestingly, roughly at the same time as Levinson found his results, Ambartsumian [12] and Chandrasekhar [13] developed what we now recognize as the equivalent results for continuous-time processes.

This led to the discovery of Chandrasekhar-type equations for constant-parameter state-space models [14], [15]. These equations have the property that the number of parameters that have to be computed is reduced from being proportional to the *cube* to proportional to the *square*, or in some cases directly proportional to the number of states, in a large class of problems of practical interest. This reduction of the computational complexity for state-space models is the first of two new trends in the study of dynamic systems that we shall report on here.

The second set of alternatives to the current state-space methods is exemplified by a trend back to external or input-output and transfer-function descriptions. These types of models have actually always been more popular in econometrics and related fields. The work of Levinson and Chandrasekhar resulted in efficient recursive methods for filtering of stationary processes not explicitly requiring a model. Mathematically these methods are efficient means for solving linear Toeplitz or displacement kernel equations. By introducing the notion of "shift low rank", of an arbitrary matrix, in [16], or the related concept of "distance from Toeplitz", in [17], we have obtained recursive algorithms for general nonstationary processes, without requiring an apriori model. It is interesting and important however that the state-space assumption can be imbedded in the input-output descriptions and our new algorithms can be specialized to the previously known Chandrasekhar and Riccati equations.

II. Dynamic Models

For the sake of definiteness let us first define the two most important types of dynamic models. Define

m—input variables $\quad u_t \quad$ (an $m \times 1$ vector)

p—output variables $\quad y_t \quad (p \times 1)$

n—state variables $\quad x_t \quad (n \times 1)$.

The discrete-time linear *state-space model* is then given by

$$\begin{aligned} x_{t+1} &= \Phi_t x_t + \Gamma_t u_t \\ y_t &= H_t x_t + \mathbf{v}_t, \end{aligned} \tag{2.1}$$

with

p—output (observation) noise variables $\mathbf{v}_t$, *a p* by 1 vector.

where

(Φ_t, Γ_t, H_t) are compatible matrices.

The linear input/output or autoregressive moving-average *(ARMA)-type model* (or ARMAX) can be defined by

$$\begin{aligned} A_{0,t} y_t + A_{1,t} y_{t-1} + \cdots + A_{q,t} y_{t-q} \\ = B_{0,t} u_t + B_{1,t} u_{t-1} + \cdots + B_{q,t} u_{t-q} \\ + C_{0,t} e_t + C_{1,t} e_{t-1} + \cdots + C_{q,t} e_{t-q}, \end{aligned} \tag{2.2}$$

with $|A_{0,t}| \neq 0$ and

p—(random) driving noise variables e_t, *a p* by 1 vector.

Here the input variables u_t are assumed to be known.

In econometrics the variables $\{y_t\}$ are sometimes labled the "dependent" or "endogenous" variables and $\{u_t\}$ sometimes "exogenous" or "control" variables.†

The Riccati-equation-based Kalman filter can now be summarized as

†Depending on the precise application, these labels are not necessarily fixed. For instance in certain time-series modeling or identification procedures the roles of the input, output and state variables and the model parameters are switched, see, e.g., [1], [18].

follows. If u_t and v_t are white noise processes (E denotes expectation)

$$E\begin{bmatrix} u_t \\ v_t \end{bmatrix} [u'_\tau, v'_\tau] = \begin{bmatrix} Q_t & 0 \\ 0 & R_t \end{bmatrix} \delta_{t,\tau}, \delta_{t,\tau} \triangleq \begin{matrix} 1, \\ 0, \end{matrix} \begin{cases} \text{if } t = \tau \\ \text{else} \end{cases}$$

and the initial conditions are *random* with

$$E\ x_0 = 0, \quad Ex_0x_0' = \Pi_0, \quad Eu_tx_0' \equiv 0 \equiv Ev_tx_0'$$

for simplicity, then the Kalman-Bucy equations, [3], [4], give the n estimated state variables recursively by

$$\hat{x}_{t+1|t} = \Phi_t\hat{x}_{t|t-1} + K_t(R_t^\epsilon)^{-1}\epsilon_t, \ \hat{x}_{0|-1} = 0, \tag{2.3}$$

where the p *output prediction errors* or *innovations* are equal to

$$\epsilon_t = y_t - H_t\hat{x}_{t|t-1},$$

and have variance

$$E\epsilon_t\epsilon_t' = R_t^\epsilon = H_tP_{t|t-1}H_t' + R_t.$$

Here $P_{t+1|t}$ is the variance of the state prediction error,

$$\tilde{x}_{t+1|t} = x_{t+1} - \hat{x}_{t+1|t}$$

and

$$E\tilde{x}_{t+1|t}\tilde{x}'_{t+1|t} = P_{t+1|t} = \Phi_tP_{t|t-1}\Phi_t' + \Gamma_tQ_t\Gamma_t' - K_t(R_t^\epsilon)^{-1}K_t', \tag{2.4}$$

where

$$P_{0|-1} = \Pi_0 = Ex_0x_0',$$

(2.4) is the $n \times n$ matrix Riccati Equation (RE). K_t in (2.3) is given by

$$K_t = \Phi_tP_{t|t-1}H_t'.$$

The number of operations for the RE is of order $O(n^3)$ per time step.

We may note that there exist many related forms.† For instance for *high initial uncertainty* ($\Pi_0 = \infty$) we can use the Bayes or information filter form; it leads to a Riccati equation for (P_t^{-1}), see, e.g., [19].

Alternatives to the RE are the Chandrasekhar-type equations [14], [15] based on the fact that even if P_t if *full rank*, the rank (α) of $\delta P_{t+1} \triangleq P_{t+1} - P_t$ *need not be full.*

†For notational convenience we shall from now on drop the conditioning of P on past data, as is customary, i.e., $P_t = P_{t|t-1}$.

Examples

For low initial uncertainty, i.e. certainty: $\Pi_0 = 0$

$\delta P_1 = P_1 - P_0 = \Gamma Q \Gamma'$, rank: $\rho[\delta P_1] = \alpha \leq m \Lambda n \blacktriangle \min(\min)$

For stationary process: $\overline{\Pi} = \Phi \overline{\Pi} \Phi' + \Gamma Q \Gamma'$ = steady state covariance

$$\delta P_1 = P_1 - P_0 = -\Phi \overline{\Pi} H'(R + H \overline{\Pi} \overline{H})^{-1} H \overline{\Pi} \Phi', \ \alpha \leq p \Lambda n$$

For high initial uncertainty, i.e. complete ignorance: $\Pi_0 = \infty$

$$\delta(P_1^{-1}) = P_1^{-1} - P_0^{-1} = H'R^{-1}H, \rho[\delta(P_1^{-1})] = \bar{\alpha} \leq p \Lambda n.$$

It can be proven, see [14], [15], that for constant parameter models $\rho[\delta P_1]$ is an upper bound for $\rho[\delta P_t]$. This fact can be exploited via the Chandrasekhar-type equations [15]. Let

$$\delta P_t = Y_t M_t^{-1} Y_t',$$

a low rank factorization (nonunique) where Y is $n \times \alpha$, M is $\alpha \times \alpha$, then

$$\text{(2.5)††} \qquad \left.\begin{matrix} K_{t+1} \\ R^{\epsilon}_{t+1} \end{matrix}\right] = \left.\begin{matrix} K_t \\ R^{\epsilon}_t \end{matrix}\right] - \left.\begin{matrix} \Phi \\ H \end{matrix}\right] Y_t M_t^{-1} Y_t' H'$$

$((n + p) \times p)$ variables

$$\text{(2.6)} \qquad \left.\begin{matrix} Y_{t+1} \\ M_{t+1} \end{matrix}\right] = \left.\begin{matrix} \Phi Y_t \\ M_t \end{matrix}\right] - \left.\begin{matrix} K_t \\ Y_t' H' \end{matrix}\right] (R^{\epsilon}_t)^{-1} H Y_t$$

$((n + \alpha) \times \alpha)$ variables,

with initial conditions found from $\delta P_0 = Y_0 M_0^{-1} Y_0'$.†

Equations (2.5) and (2.6) can replace the $n \times n$ matrix equation (2.4) for constant models (Φ, Γ, H, R, Q). The *most interesting feature* of these new equations is their reduced *computation complexity:* (for $p \leq \alpha \ll n$), they require *$O(n^2\alpha)$ operations* per time step (or $O(n\alpha)$ for canonical forms), and *$O(n\alpha)$ in storage* (plus $O(n^2)$ for noncanonical forms).

For example, in an air-pollution study in [1] the order of the model was 500 and a reduction of roughly a thousand was achieved using (2.5), (2.6).

Another alternative to (2.4) is the *square-root filter* [19]. We can roughly describe this method by triangularly factoring the matrix P_t into $P_t^{1/2} P_t^{T/2}$ and defining the "pre-array" (containing *a-priori* information) and the "post-array" (with *a-posteriori* information) as

††A right bracket denotes a (block) vector.

†This decomposition is only unique modulo orthogonal transformations, i.e., $\tilde{Y}_0 = Y_0 T$ with $TT' = I$ is another factorization, where $\tilde{M}_0 = T' M_0 T$ could be a signature matrix, see [15], [19].

$$
\begin{array}{ccc}
& \textit{pre-array} & & \textit{post-array} \\
(2.7) & \begin{bmatrix} R_t^{1/2} & H_tP_t^{1/2} & 0 \\ 0 & \Phi_tP_t^{1/2} & \Gamma_tQ_t^{1/2} \end{bmatrix} \cdot \tau & = & \begin{bmatrix} x_{11} & 0 & 0 \\ x_{21} & x_{22} & 0 \end{bmatrix},
\end{array}
$$

where τ is an orthogonal matrix (e.g., product of Householder transformations [19])

$$\tau\tau' = I,$$

that triangularizes the pre-array into the post-array with

$$x_{11} = (R_t^\epsilon)^{1/2} \quad \text{(triangular)}$$

$$x_{21} = K_t(R_t^\epsilon)^{-T/2} = \tilde{K}_t, \quad x_{22} = P_{t+1}^{1/2} \quad \text{(triangular)},$$

the required "a posteriori" information for the filter.

The proof can be seen from

$$\text{"Post-Array"} = \text{"Pre-Array"} \cdot \tau$$

or with $[M]^2 \triangleq MM'$

$$(2.8) \qquad [\text{"Post-Array"}]^2 = [\text{"Pre-Array"}]^2 =$$

$$
\begin{bmatrix}
\underbrace{R_t + H_tP_tH_t'}_{R_t^\epsilon = x_{11}x_{11}'} & \underbrace{H_tP_t\Phi_t'}_{K_t' = x_{11}x_{21}'} \\
\underbrace{\Phi_tP_tH_t'}_{K_t = x_{21}x_{11}'} & \underbrace{\Phi_tP_t\Phi_t' + \Gamma_tQ_t\Gamma_t'}_{P_{t+1} + K_t(R_t^\epsilon)^{-1}K_t' = x_{22}x_{22}' + x_{21}x_{21}'}
\end{bmatrix}
\begin{array}{l}
\leftarrow \text{[Gain Equation]} \\
\leftarrow \text{[Riccati Equation 2.4]}
\end{array}
$$

Similarly we can find square-root Chandrasekhar forms [19] from

$$
\begin{array}{c}
\begin{array}{c} p \\ n \end{array}
\begin{bmatrix} R^{1/2} & HP_t^{1/2} & 0 \\ 0 & \Phi P_t^{1/2} & \Gamma Q^{1/2} \end{bmatrix}
\tau_1 =
\begin{bmatrix} (R_t^\epsilon)^{1/2} & 0 & 0 \\ \tilde{K}_t & P_{t+1}^{1/2} & 0 \end{bmatrix}
\begin{array}{c} p \\ n \end{array} \\
\qquad p \qquad n \qquad m \qquad\qquad\qquad\qquad p \qquad n \qquad m
\end{array}
$$

Let $\delta P_t = \tilde{L}_t\tilde{L}_t'$ or $\tilde{L}_t = Y_t(M_t^{T/2})^{-1}$ (with possibly imaginary columns), then

$$P_{t+1} = P_t + \tilde{L}_t\tilde{L}_t' \Leftrightarrow [P_{t+1}^{1/2}, 0] = [P_t^{1/2}, \tilde{L}_t]\tau,$$

$$
\begin{array}{c}
\begin{array}{c} p \\ n \end{array}
\begin{bmatrix} (R_{t-1}^\epsilon)^{1/2} & H\tilde{L}_{t-1} \\ \tilde{K}_{t-1} & \Phi\tilde{L}_{t-1} \end{bmatrix}
\tau_2 =
\begin{bmatrix} (R_t^\epsilon)^{1/2} & 0 \\ \tilde{K}_t & \tilde{L}_t \end{bmatrix}
\begin{array}{c} p \\ n \end{array} \\
\qquad p \qquad \alpha \qquad\qquad\qquad\qquad p \qquad \alpha
\end{array}
$$

This leads to a reduction of the arrays from $p + n + m$ (the equivalent of

(2.4)) to $p + \alpha$ $(\alpha \leq n)$ columns in the arrays (the equivalent of (2.5) and (2.6)). The initial conditions are given by

$$P_1 - P_0 = \tilde{L}_0 \tilde{L}_0' = \Phi\ \Pi_0\ \Phi' + \Gamma Q \Gamma' - \tilde{K}_0 \tilde{K}_0'$$

$$\tilde{K}_0 = K_0 (E_0^\epsilon)^{-T/2},\ K_0 = \Phi\ \Pi_0\ H' \qquad R_0^\epsilon = H\ \Pi_0\ H' + R.$$

The estimator (2.3) can then be written as

$$\hat{x}_{t+1|t} = \Phi \hat{x}_{t|t-1} + \tilde{K}_t \nu_t \tag{2.9}$$

$$\nu_t \triangleq (R_t^\epsilon)^{-1/2} \epsilon_t = (R_t^\epsilon)^{-1/2} (y_t - H\hat{x}_{t|t-1}),$$

where ν_t are the orthonormalized innovations, since their covariance is given by

$$E\nu_t \nu_\tau' = (R_t^\epsilon)^{-1/2} E\epsilon_t \epsilon_\tau' (R_\tau^\epsilon)^{-T/2} \delta_{t,\tau}$$
$$= I\delta_{t,\tau}.$$

By examining these alternative algorithms to the RE it becomes evident that they have considerable advantages:

i) Because of the fact that they work with (matrix) square roots of P (or δP) the numerical conditioning is much better, since the eigenvalues of the square root matrix are the square roots of the eigenvalues of P (or δP); e.g., if the eigenvalues of P are 10^6 and 10^{-6}, those of $P^{1/2}$ are 10^3 and 10^{-3}, a considerable reduction that allows, for example, the use of single-precision instead of double-precision computations in a computer.

ii) In addition the nonnegativeness of P can be better insured, since P is obtained by "squaring" (i.e., $[``P^{1/2}"]^2 = P^{1/2} P^{T/2}$).

iii) If the value of α = rank (δP) is small compared to n, and if the system parameters are constant (for simplicity) considerable savings can be obtained in computation and storage requirements, typically a reduction by a factor of n.

III. Significance of Dimensionality Reduction

For state-space models with constant H and Φ matrices and with α = rank $(P_{t+1} - P_t)$, the order of computation and storage required by the fast algorithms is equal to $O(n\alpha)$, or in other words proportional to α, see [14], [15], [19]. This reduction in computation is not necessarily restricted to state space models.

The general results can be stated as follows [16], [17]. The order of computation and storage required to invert various types of matrices of size $N \times N$ are

		operations	storage
for a general matrices:	R^{-1}	$O(N^3)$	$O(N^2)$
for a Toeplitz† matrices:	$R^{-1} = \bar{L}_1\bar{L}_1' - \bar{L}_2\bar{L}_2'$	$O(N^2)$	$O(N)$
for a "α-distant" matrices:	$R^{-1} = \sum_{i=1}^{\alpha} \bar{L}_i\bar{U}_i$	$O(N^2\alpha)$	$O(N\alpha)$
where	$R = \sum_{i=1}^{\alpha} U_iL_i = \sum_{i=1}^{\alpha+2} \tilde{L}_i\tilde{U}_i$	$O(N^2\alpha)$	$O(N\alpha)$

and

$$\alpha = \operatorname{rank}(R - Z'RZ).$$

If Z *is the "delay matrix"*

$$Z = \begin{bmatrix} 0 & & & \bigcirc \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{bmatrix}, \tag{2.10}$$

the α is the so-called "shift rank" or tensor rank in [16] or the "distance from Toeplitz" [17]. The matrices L_i, U_i', $\tilde{L}_i$, $\tilde{U}_i'$, $\bar{L}_i$, $\bar{U}_i'$ are lower-triangular Toeplitz matrices that are found by decomposing a matrix R into sums of α products of the type U_iL_i or $\tilde{L}_i\tilde{U}_i$. The significant point is that products and inverses of Toeplitz matrices are not Toeplitz in general, while inverses of sums of products of matrices of the type above have the same form and therefore a nice closure property. The $\bar{L}_i\bar{U}_i$ matrices can be computed, for instance, via Levinson-type algorithms [16], [17] of $O(N^2)$ operation. Examples of matrices with low "shift rank α':

$\alpha = 1 : L$ – lower triangular Toeplitz matrix
$\alpha = 2 : T$ – full Toeplitz matrix $= I \cdot T_+ + T_- \cdot I$
$\alpha = 3 : LU$ = lower times upper triangular Toeplitz
$\qquad = U_1L_1 + U_2L_2 - U_3L_3$
$\alpha = 4 : T_1 \cdot T_2$ – product of two full Toeplitz matrices
$\alpha \leq n$: covariances of state space models with constant parameters e.g.,
$\qquad R = LL' + \tilde{\mathcal{O}}\tilde{\mathcal{O}}$ ($\tilde{\mathcal{O}}$ is $N \times n$).

For these types of matrices, Levinson-type equations can be used [16], [17] (and they can be specialized to the Chandrasekhar-type equations (2.5), (2.6)). In another set of important applications $\alpha = 3{,}4 : Y'Y$, where Y is a $(t \times n)$ Toeplitz matrix, found in least squares problems and identification of AR, ARMA models, see, e.g., [16], [20] and equations (2.19) to (2.24).

†R is Toeplitz if $[R]_{i,j}$ = function of $(i - j)$ only, e.g., covariance matrices of stationary processes.

The *matrices Z and Z'* do not have to be "delay matrices" in order to get a closure property, but many other interesting choices can be made. For instance if Z is a particular *circulant matrix*

$$(2.11) \qquad Z = Z_c \triangleq \begin{bmatrix} 0 & & & 1 \\ 1 & \ddots & \bigcirc & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{bmatrix}, Z' = Z^{-1},$$

we can define the "distance from circularity"

$$\alpha_c = \text{rank}[R - Z_c' R Z_c].$$

This concept is useful in inverting so-called banded Toeplitz matrices

$$[B]_{ij} = \begin{cases} b_{i-j}, & |i - j| \leq n = \text{constant} \ll N \\ 0 \quad , & \text{else, where } 1 \leq i, j \leq N \end{cases}.$$

Here $\alpha_c(B) = 2n$, therefore we can find a decomposition of B

$$(2.12) \qquad B = C + PMP', \quad \text{(full) rank}[M] \leq 2n,$$

where $B - C$ is of rank at most $2n \ll N$, therefore $PMP' = B - C$ is a low rank factorization (of this type in Eq. (2.6)). C is a circulant matrix, defined by

$$(2.13) \qquad C = [\mathbf{c}_1, \ldots, \mathbf{c}_N] \triangleq \sum_{i=-n}^{n} Z_c^i c_i = WDW^*,$$

where $c_i = b_i$ for instance. D is a diagonal matrix of the eigenvalues of C obtained from the first column $\mathbf{c}_1$ of C by

$$(2.14) \qquad D = \text{diag}(W^* \mathbf{c}_1)$$

and W is the orthogonal matrix of the discrete Fourier transform ($W^* W = I$). The eigenvalues in (2.14) can therefore be computed very efficiently, $O(N \log n)$, using the Fast Fourier Transform (FET) as is well-known in numerical analysis. Thus a representation for the inverse of the matrix B in (2.12) (or the solution to a linear equation with B) can be computed with $O(N \log n) + O(n^2)$ operations using another well-known fact, the matrix identity

$$(2.15) \qquad [A + BCD]^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})DA^{-1}$$

in

$$B^{-1} = [C + PMP']^{-1},$$

with L and P,M given by

$$B - C = \begin{bmatrix} & & L' \\ & 0 & \\ L & & \end{bmatrix}, \quad P \triangleq \left[\begin{array}{c|c} 0 & I_n \\ 0 & 0 \\ I_n & 0 \end{array}\right], \quad M \triangleq \begin{bmatrix} 0 & L' \\ L & 0 \end{bmatrix}.$$

Now

$$B^{-1} = C^{-1} - C^{-1}PT^{-1}P'C^{-1}, \tag{2.16}$$

where T turns out to be a Toeplitz matrix in this case. The inverse of a circulant is particularly simple because of the eigenvalue decomposition in (2.13)

$$C^{-1} = WD^{-1}W^{*}. \tag{2.17}$$

We may note as an example that W itself has "weighted shift-low rank" (see [16]) since ($D = \operatorname{diag}(W_{ii}/W_{i+1,i+1})$)

$$\alpha_w = \operatorname{rank}[W - Z'DWD^{*}Z] = 2, \tag{2.18}$$

a fact that is related to the so-called Chirp-Z transform.

A Least-Squares Prediction Problem

Let $\{y\}$ be generated by an autoregressive model of order n, then define the one-step predicted estimate by

$$\hat{y}_t = -\sum_{k=1}^{n} a_k y_{t-k}, \quad 0 \le t \le T \tag{2.19a}$$

and

$$e_t = y_t - \hat{y}_t$$

as the prediction error. Then we can distinguish four cases of interest, the set of equations with braces 1 to 4 ($= i$)

$$\begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_n \\ \vdots \\ e_T \\ e_{T+1} \\ \vdots \\ e_{T+n} \end{bmatrix} = \begin{bmatrix} y_0 & & & \bigcirc \\ y_1 & y_0 & & \\ \vdots & & \ddots & \\ y_n & & & y_0 \\ \vdots & & & \vdots \\ y_{T-n} & & \ddots & \vdots \\ y_T & & & y_{T-n} \\ & \ddots & & \vdots \\ \bigcirc & & & y_T \end{bmatrix} \begin{bmatrix} 1 \\ -a_1 \\ \vdots \\ -a_n \end{bmatrix} = \begin{bmatrix} L \\ X \\ U \end{bmatrix} \mathbf{a} \tag{2.19b}$$

Note that L, X, U are Toeplitz matrices. Define the vectors

$$\mathbf{e}_i = Y_i \mathbf{a}_i, \quad i = 1,2,3,4.$$

The object is to minimize a squared error criteria say

$$\| \mathbf{e}_i \|^2 = \mathbf{e}_i' \mathbf{e}_i = P_i.$$

The solution is given by the normal equations

$$[Y_i' Y_i] \mathbf{a}_i = [Y_i' \mathbf{e}_i]$$

or

(2.20)
$$\mathcal{R}_i \mathbf{a}_i = \left. \begin{matrix} P_i \\ \theta \end{matrix} \right], i = 1,2,3,4.$$

Case 1

If

(2.21)
$$\mathbf{e} \triangleq [e_0, \ldots, e_{T+n}]'$$
$$\mathcal{R}_T \triangleq [Y_1' Y_1]_0^{T+n} \text{ is Toeplitz,}$$

and we can use Levinson's algorithm. It can be shown that a_i is stable, (i.e., $a(z) = [z^n, \ldots, 1]\mathbf{a}$ has its root inside the unit circle); Gohberg (see reference in [17]) showed that $\mathcal{R}_T^{-1} = AA' - BB'$, where A, B are (lower) triangular Toeplitz matrices.

Case 2

If

$$\mathbf{e} \Delta [e_n, \ldots, e_T]',$$

we get the most commonly used least-squares solution, where

(2.22)
$$\mathcal{R}_c \Delta [Y_2' Y_2]_n^T \text{ is not Toeplitz}$$
$$= \mathcal{R}_T - L'L - U'U,$$

however, in this case one can show that

$$\mathcal{R}_c^{-1} = AA' - BB' - CC' + DD',$$

where C, D are of similar type as A, B.

Case 3

If

(2.23)
$$\mathbf{e} \triangleq [e_0, \ldots, e_T]' \quad (\sim \text{I.R.})$$
$$\mathcal{R} \triangleq [Y_3' Y_3]_0^T$$

is not Toeplitz, but equals

$$= \mathcal{R}_T - U'U$$

and

$$\mathcal{R}^{-1} = AA' - BB' - CC' \tag{2.24}$$

By considering block Toeplitz matrices of the type

$$\begin{bmatrix} \mathcal{R}_T & U \\ U' & I \end{bmatrix}$$

this result can be verified.

Case 4 is not commonly used; however, it turns out that $\mathbf{a}_4$ is also stable.

Outline of a Least-Squares Polynomial Regression Application

Often the problem of fitting polynomial trends to observations $\{y\}$ arises, i.e., we want to find a vector x such that

$$\begin{aligned} y_t &= [1, t, t^2, \ldots, t^n]x + e_t \\ &= H_t x + e_t \end{aligned}$$

with $\{e_t\}$ = error sequence. Note that H_t is time-varying. The usual solution is givenby defining

$$\begin{aligned} \mathcal{Y}_t &\triangleq [y_0, \ldots, y_t] \\ [\mathcal{O}_t]_{k,i} &= k^i, \quad i = 0, \ldots, n \\ &\qquad\quad\; k = 0, \ldots, t \end{aligned}$$

then the least squares regression coefficients are given by

$$\hat{x}_{LS} = [\mathcal{O}_t'\mathcal{O}_t]^{-1}\mathcal{O}_t'\mathcal{Y}_t,$$

which requires $O(n^2t) + O(n^3)$ operations in general.

However, notice the fact that powers of t are solutions to *constant parameter* difference equations! For example, equivalently consider

$$\mathcal{O}_t = \begin{bmatrix} 1 & & & & & \bigcirc & & \\ 1 & 1 & & & & & & \\ 1 & 2 & 1 & & & & & \\ 1 & 3 & 3 & 1 & & & & \\ 1 & 4 & 6 & 4 & 1 & & & \\ 1 & 5 & 10 & 10 & 5 & 1 & & \\ 1 & 6 & 15 & 20 & 15 & 6 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} = \begin{bmatrix} H \\ H\Phi \\ H\Phi^2 \\ \vdots \\ H\Phi^t \end{bmatrix},$$

"Pascal Triangle"

We can now get a constant parameter state-model of the type

$$x_{t+1} = \Phi x_t,$$

$$y_t = Hx_t + v_t,$$

where (c.f. (2.10)) $$\Phi = [I_n + Z'] = \begin{bmatrix} 1 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 1 \end{bmatrix}$$

$$H = [1, 0, \ldots, 0]$$

Let $R = I$, $\Pi_0^{-1} = 0$ in the Kalman filter, then the state estimates will contain the least-squares solution to the problem of fitting a polynomial to the observations $\{y\}$. Since we have now a constant-parameter model, we can use the Chandrasekhar-type equations for prediction purposes or trend removal (the prediction errors $\{\epsilon_t\}$ will be the desired reduced process). The least-squares estimate of x in the usual solution has to be found actually via smoothing versions of these estimators, see, e.g., [19]. Notice that since $\Pi_0^{-1} = 0$, we have a high initial uncertainty and the Bayes or information filter forms [19] should be used.

IV. Conclusions

Due to space limitations, this paper is only a rough and somewhat biased outline of some new results in time-series analysis and estimation. The control aspects have essentially been ignored, partially because for the linear model, quadratic cost case, many of the mathematical results have their duals in linear estimation problems and can therefore with some skill be translated from one area into another.

As a conclusion we would like to encourage mathematically oriented workers in econometrics to try to follow the field of estimation and control, for instance through the journals of [1] and [2], since many of the methods developed in these fields have interesting potential applications in econometrics and related areas.

References

[1] Special Issue on System Identification and Time-Series Analysis, *IEEE Trans. on Automatic Control*, vol. AC-19, no. 6, pp 637–951, December 1974.

[2] T. Kailath, "A View of Three Decades of Linear Filtering Theory," *IEEE Trans. on Information Theory*, vol. IT-20, no. 2, pp. 145–181, March 1974 (invited paper).

[3a] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., vol. 82, pp. 34–45, March 1960.

[3b] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Trans. ASME*, Ser. D, J. Basic Eng., vol. 83, pp. 95–107, December 1961.

[4] R. L. Stratonovich, "Application of the Theory of Markov Processes for Optimum

Filtration of Signals," *Radio Eng. Electron. Phys. (USSR)*, vol. 1, pp. 1–19, November 1960.

[5] Special Issue on the Linear-Quadratic-Gaussian Estimation and Control Problem, *IEEE Trans. on Automatic Control*, vol. AC-16, no. 6, pp. 527–869, December 1971.

[6] B. Dickinson, T. Kailath and M. Morf, "Canonical Matrix Fraction and State Space Description for Deterministic and Stochastic Linear Systems," *IEEE Trans. on Automatic Control*, Special Issue on System Identification and Time-Series Analysis, vol. AC-19, no. 6, pp. 656–667, December 1974.

[7] N. Levinson, "The Weiner RMS (Root-Mean-Square) Error Criterion in Filter Design and Prediction," *J. Math Phys.*, vol. 25, pp. 261–278, January 1947.

[8] J. Durbin, "The Fitting of Time-Series Models," *Rev. Intern. Statist. Inst.*, vol. 28, pp. 233–244, 1960.

[9] P. Whittle, "On the Fitting of Multivariate Autoregressions and the Approximate Canonical Factorization of a Spectral Density Matrix," *Biometrika*, vol. 50, pp. 129–134, 1963.

[10] P. Whittle, *Prediction and Regulation*, New York: Van Nostrand Reinhold, 1963.

[11] E. A. Robinson, *Multichannel Time-Series Analysis with Digital Computer Programs*, San Francisco, Ca.: Holden-Day, 1967.

[12] V. A. Ambartsumian, "Diffuse Reflection of Light by a Foggy Medium," *Dokl. Akad. Sci. SSSR*, vol. 38, pp. 229–322, 1943.

[13] S. Chandrasekhar, "On the Radiative Equilibrium of a Stellar Atmosphere, Pt. XXI," *Astrophys. J.*, vol. 106, pp. 152–216, 1947; Pt. XXII, *ibid*, vol. 107, pp. 48–72, 1948.

[14] T. Kailath, "Some New Algorithms for Recursive Estimation in Constant Linear Systems," *IEEE Trans. on Information Theory*, vol. IT-19, pp. 750–760, November 1973.

[15] M. Morf, G. S. Sidhu and T. Kailath, "Some New Algorithms for Recursive Estimation in Constant Linear Discrete-Time Systems," *IEEE Trans. on Automatic Control*, vol. AC-19, no. 4, pp. 315–323, August 1974.

[16] M. Morf, "Fast Algorithms for Multivariable Systems," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, Stanford, Ca., 1974.

[17] B. Friedlander, M. Morf, T. Kailath and L. Ljung, "New Inversion Formulas for Matrices Classified in Terms of their Distance from Toeplitz Matrices," submitted to SIAM Journal of Numerical Analysis.

[18] M. Morf, L. Ljung and T. Kailath, "Fast Algorithms for Recursive Identification," submitted to the 1976 IEEE Conference on Decision and Control; also presented at the IEEE International Symposium on Information Theory, Ronneby, Sweden, June 21–24, 1976.

[19] M. Morf and T. Kailath, "Square-Root Algorithms for Least-Squares Estimation," *IEEE Trans. on Automatic Control*. vol. AC-20, no. 4, pp. 487–497, August 1975.

[20] M. Morf, B. Dickinson, T. Kailath and A. Vieira, "Efficient Solutions of Covariance Equations for Linear Prediction," *IEEE Trans. on Acoustics, Speech and Signal Processing*, to appear 1977.

Stanford University