PURDUE UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

Lafayette, Indiana 47907

A NEW METHOD OF RECURSIVE ESTIMATION IN DISCRETE LINEAR SYSTEMS*

R. L. Kashyap

Technical Report No. TR-EE 68-13

June 1968

*This work was partially supported by NSF under Grant GK-1970 and by NASA under Grant ~~NGR~~ 15-005-021.
NGL

A NEW METHOD OF RECURSIVE ESTIMATION IN DISCRETE LINEAR SYSTEMS

R. L. Kashyap

ABSTRACT

Let the measurement $z(i)$ at instant i be of form $z(i) = y(i) + \eta(i)$ where $\eta(i)$ is the noise and $y(i)$ is the signal obeying a system of coupled linear difference equations of varying orders. We first derive a recursive equation for innovation or orthogonalized measurement. Using this equation, algorithms are derived for prediction, filtering and smoothing of the signal $y(i)$. The use of these schemes results in considerable reduction of computation over existing schemes. If $z(i)$ is a scalar, the relative computational efficiency is $N/2$ where N is the dimension of state vector.

I. Introduction

Recursive estimation in linear discrete dynamic systems has been treated in great detail in a number of papers. One assumes that the state vector x(i) of dimension N describing the stochastic signal at instant i obeys a known linear difference equation in state transition form. The measurement z(i) at instant i is the sum of a linear combination of x(i) and noise. One is interested in obtaining the linear least squares estimate of the state vector based on the relevant measurements. With the aid of projection methods and others, a number of recursive schemes have been developed for obtaining the prediction, filtering [Kalman[1,2]] and smoothing estimates [Bryson and Frasier[3], Rauch, Streibel and Tung[4], Rauch[5], Meditch[6]]. Even though these schemes have notational elegence, they may run into serious computational problems for large N since they involve the recursive solution of the N x N covariance matrix from a nonlinear difference equation. In obtaining smoothing estimates they may lead to serious storage problems. Further, in many problems (especially in smoothing problems) one needs only the estimates of a few components of the state vector. However, this fact does not seem to reduce the computational complexity of the problem. All these questions point to the need for alternate solutions to the estimation problem.

It has been known for a long time [Wold[7], Kolmogorov[8], Doob[9], Whittle[10]] that instead of working with the measurements z(i), i = 1,2,etc., it is very convenient to work with the new set of variables, $\tilde{z}(i) = z(i) - \hat{z}(i)$, i = 1,2,etc. known as innovations, $\hat{z}(i)$ being the predicted linear least squares estimate of z(i) based on all the past measurements. Thus the innovations are obtained by orthogonalizing the successive measurements z(i). Further all other estimates of x(i) can be expressed in terms of

the innovations [Kailath[11]]. A natural course of action for estimation is to obtain a recursive equation for the innovations without explicitly involving the state vector estimates. The estimates of the signal can be computed from the innovations. If the dimensionality of innovation equation is r (which is the dimension of measurement vector z) and that of state vector is N, then the use of these methods results in the reduction of the computation and storage by a factor of about N/2r over the schemes mentioned earlier. In the smoothing problems, the reduction in storage and computation is indeed impressive.

II. Model of the Random Process

The measurement z(i) - r-vector - is composed of the r-vector signal y(i) and noise $\eta(i)$

$$z(i) = y(i) + \eta(i) \tag{2.1}$$

$$\begin{aligned} &E(\eta(i)) = 0 \\ &E(\eta(i)\eta^T(j)) = R_\eta(i)\delta_{ij} \\ &E(y(i)\eta^T(j)) = 0 \end{aligned} \tag{2.2}$$

The signal y(i) is assumed to obey the following difference equation

$$y(t) + \sum_{j=1}^{n} A_j(t)y(t-j) = \sum_{j=1}^{n} C_{n+1-j}(t)\ \xi(t-j) \tag{2.3}$$

where the m_1-vector $\xi(i)$ has the following statistical properties:

$$\begin{aligned} &E(\xi(i)) = 0 \\ &E(\xi(i)\xi^T(j)) = R_\xi(i)\delta_{ij} \\ &E(\xi(i)\eta^T(j)) = 0 \end{aligned} \tag{2.4}$$

In equation (2.4), the r x r matrices $A_j(t)$, j = 1,...,n and r x m_1 matrices $C_j(t)$, j = 1,...,n are known for all relevant t.

For purposes of comparison we represent the random process of equations (2.1)-(2.4) in state transition form in equation (2.5) where the state vector x(i) is of dimension $N(N \geq r)$.

$$
\begin{aligned}
x(i+1) &= \mathcal{A}(i)x(i) + \mathcal{B}(i)\xi(i) \\
z(i) &= H(i)x(i) + \eta(i)
\end{aligned}
\tag{2.5}
$$

We note that in many problems, the equations (2.1) and (2.3) present a more natural starting point for the analysis of the system than the equations (2.5). In addition, the algebraic manipulations needed for the conversion of equations (2.1) and (2.3) into the form of equation (2.5) may be considerable.

III. Recursive Equations for Innovation

Let $\hat{z}(t/t-1)$ be the linear least squares predictor of z(t) based on the measurements z(t-1), z(t-2), ...

$$\hat{z}(t/t-1) = \text{Arg}\,[\min_{f_t} E\|z(t)-f_t(z(t-1), z(t-2), \ldots)\|^2]$$

where $f_t(\cdot)$ is a linear function of the arguments. Let $\tilde{z}(t) \triangleq z(t)-\hat{z}(t/t-1)$ = innovation at time t. It is easy to demonstrate the orthogonality of the successive innovations. Our intention is to find a difference equation for $\tilde{z}(i)$. In order to do this we need to represent z(t) as an autoregressive process. This will be done presently.

(A) Recursive equation for the measurement z(i)

Substituting (2.1) in (2.3) we have the following equation for z(i)

$$z(t) + \sum_{j=1}^{n} A_j(t)z(t-j) = \eta(t) + \sum_{j=1}^{n} A_j(t)\eta(t-j) + \sum_{j=1}^{n} C_{n+1-j}(t)\xi(t-j) \tag{3.1}$$

Let us denote the right hand side of (3.1) by $w_1(t)$. We want to replace $w_1(t)$ by another process $w_2(t)$ which has the form given in equation (3.2) and has the same first and second order properties and the same dimension as $w_1(t)$.

$$w_2(t) = \epsilon(t) + \sum_{j=1}^{n} B_{n+1-j}(t)\,\epsilon(t-j) \tag{3.2}$$

The r-vector stochastic process $\epsilon(t)$ is zero mean and uncorrelated

$$\begin{aligned} &E(\epsilon(i)) = 0 \\ &E(\epsilon(i)\,\epsilon^T(j)) = R_\epsilon(i)\delta_{ij} \end{aligned} \tag{3.3}$$

The r x r coefficient matrices $B_j(t)$, $j = 1,\ldots,n$ and the variance $R_\epsilon(t)$ are determined from the relation (3.4)

$$E(w_2(t)w_2^T(t+\tau)) = E[w_1(t)w_1^T(t+\tau)] \quad \forall\ t \text{ and } \tau \tag{3.4}$$

Replacing $w_1(t)$ by $w_2(t)$ in (3.1) gives us the required difference equation for $z(t)$.

$$z(t) + \sum_{j=1}^{n} A_j(t)z(t-j) = w_2(t) \triangleq \epsilon(t) + \sum_{j=1}^{n} B_{n+1-j}(t)\,\epsilon(t-j)$$

To determine the coefficients $B_j(t)$ from (3.4) note that

$$E\,[w_1(t)w_1^T(t+\tau)] \triangleq W_1(t,t+\tau) \triangleq 0 \quad \text{for } |\tau| > n$$

$$E\,[w_2(t)w_2^T(t+\tau)] \triangleq W_2(t,t+\tau) \triangleq 0 \quad \text{for } |\tau| > n$$

Further

$$\begin{bmatrix} W_2(t,t-n) \\ W_2(t,t-n+1) \\ \cdot \\ \cdot \\ \cdot \\ W_2(t,t) \end{bmatrix} \triangleq$$

$$\begin{bmatrix} B_{n+1}(t-n)R_\epsilon(t-n) & & & \\ B_n(t-n+1)R_\epsilon(t-n) & B_{n+1}(t-n+1)R_\epsilon(t-n+1) & & \\ B_{n-1}(t-n+2)R_\epsilon(t-n) & B_n(t-n+2)R_\epsilon(t-n+1) & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & B_{n+1}(t-1)R_\epsilon(t-1) & \\ B_1(t)R_\epsilon(t-n) & B_2(t)R_\epsilon(t-n+1) & B_n(t)R_\epsilon(t-1) & B_{n+1}(t)R_\epsilon(t) \end{bmatrix} \begin{bmatrix} B_1^T(t) \\ B_2^T(t) \\ B_3^T(t) \\ \cdot \\ \cdot \\ B_n^T(t) \\ B_{n+1}^T(t) \end{bmatrix} \quad \cdots\cdots (3.5)$$

where $B_{n+1}(t) = 1 \ \forall\ t$. We can solve (3.4) for $B_1(t),\ldots,B_n(t)$ and $R_\epsilon(t)$ in terms of $W_1(t,t-\tau)$, $\tau = 0,\ldots,n$ and $B_i(\tau)$, $R_\epsilon(\tau)$, $\tau < t$.

$$\left.\begin{aligned} B_1^T(t) &= R_\epsilon^{-1}(t-n)\, W_1(t,t-n) \\ B_2^T(t) &= R_\epsilon^{-1}(t-n+1)\, [W_1(t,t-n+1) - B_n(t-n+1)R_\epsilon(t-n)B_1^T(t)] \\ B_i^T(t) &= R_\epsilon^{-1}[t-n+i-1]\, \{W_1(t,t-n+i-1) - \\ &\qquad \sum_{j=1}^{i-1} B_{n-i+j+1}(t-n+i-1)R_\epsilon(t-n+j-1)B_j^T(t)\} \\ &\qquad\qquad i=2,\ldots,n \\ R_\epsilon(t) &= -\sum_{j=1}^{n} B_j(t)R_\epsilon(t-n+j-1)B_j^T(t) + W_1(t,t) \end{aligned}\right\} \quad (3.6)$$

The formula for $W_1(t,t+\tau)$ can be written down as in equation (3.5)

$$\left.\begin{aligned}
W_1(t,t-n) &= A_o(t-n)R_\eta(t-n)\,A_n^T(t)\\
W_1(t,t-n+1) &= C_n(t-n+1)\,R_\xi(t-n)\,C_1^T(t) + \\
&\qquad \sum_{k=1}^{2} A_{2-k}(t-n+1)R_\eta(t-n+k-1)A^T_{n-k+1}(t)\\
W_1(t,t-n+i-1) &= \sum_{k=1}^{i-1} C_{n-i+1+k}(t-n+i-1)R_\xi(t-n+k-1)C_k^T(t) + \\
&\qquad \sum_{k=1}^{i} A_{i-k}(t-n+i-1)R_\eta(t-n+k-1)A^T_{n-k+1}
\end{aligned}\right\} \quad (3.7)$$

$$i=2,\ldots,n+1$$

where $A_o(t) = 1 \;\forall\; t$

Thus at every instant, we have to evaluate recursively the $\{(n+1)r^2-1\}$ elements of the matrices $B_j(t)$, $j=1,\ldots,n$ and $R_\epsilon(t)$. This step roughly corresponds to the recursive evaluation of the $\frac{N(N+1)}{2}$ elements of the covariance matrix in the Kalman filtering. We can summarize the result in the form of propositions.

<u>Proposition 1</u>: The process $z(t)$ defined in equations (2.1)-(2.4) obeys the following difference equation (3.8) in a wide sense [Doob[9]]

$$z(t) + \sum_{j=1}^{n} A_j(t)z(t-j) = \epsilon(t) + \sum_{j=1}^{n} B_{n+1-j}(t)\;\epsilon(t-j) \qquad (3.8)$$

$\epsilon(t)$ is a zero mean uncorrelated stochastic process with variance $R_\epsilon(t)$. The coefficient matrices $B_i(t)$, $i=1,\ldots,n$ and $R_\epsilon(t)$, are computed recursively by equations (3.6) and (3.7).

Later we shall need to compute the correlation matrix between $\eta(t)$ and $\epsilon(j)$. These results are expressed in Proposition 2.

Proposition 2: Let $D(i,k) = E\,[\eta(i)\;\epsilon^T(j)]$. Then

$$D(i,k) = 0 \qquad k < i \tag{3.9a}$$

$$D(i,i) = R_\eta(i) \tag{3.9b}$$

$$D(i,k) + \sum_{j=1}^{k-i} D(i,k-j)\;B^T_{n+1-j}(k) = R_\eta(i)\;A^T_{k-i}(k)\;, \qquad i < k < n+i \tag{3.9c}$$

$$D(i,k) + \sum_{j=1}^{n} D(i,k-j)\;B^T_{n+1-j}(k) = 0\;,\quad k > n+i \tag{3.9d}$$

Equation (3.9)a follows directly from causality. Equations (3.9)b, c, and d can be derived by multiplying (3.8) with $\eta^T(i)$, taking expectations on either side and using equations (2.1), (2.2) and (3.9)a.

(B) Recursive Relation for Innovation

Let us rewrite the equation for $z(t)$

$$z(t) = -\sum_{j=1}^{n} A_j(t)z(t-j) + \sum_{j=1}^{n} B_{n+1-j}(t)\epsilon(t-j)+\epsilon(t) \tag{3.10}$$

By definition, $\hat{z}(t/t-1)$ equals the sum of the terms in equation (3.10) projected onto the space of measurements $z(t-1)$ $z(t-2)$, etc. Of these, $\epsilon(t)$ is clearly orthogonal to the subspace spanned by $z(t-1)$, $z(t-2)$, etc. (Two zero mean random vectors, x and y, are said to be orthogonal if $E(xy^T) = 0$.) Hence

$$\hat{z}(t/t-1) = -\sum_{j=1}^{n} A_j(t)z(t-j) + \sum_{j=1}^{n} B_{n+1-j}(t)\;\epsilon(t-j) \tag{3.11}$$

We can evaluate $\epsilon(t)$ by subtracting equation (3.11) from (3.10)

$$\tilde{z}(t) \triangleq z(t) - \hat{z}(t/t-1) = \epsilon(t) \tag{3.12}$$

Substituting (3.12) in (3.11) gives us the required recursive equation for (3.13) for $\hat{z}(t/t-1)$ which is expressed as Proposition 3. This proposition is the foundation for the result of our results.

Proposition 3: Consider the process z(t) obeying equation (3.8). The one step optimal predictor $\hat{z}(t/t-1)$ obeys the following equation (3.13).

$$\hat{z}(t/t-1) = -\sum_{j=1}^{n} A_j(t)z(t-j) + \sum_{j=1}^{n} B_{n+1-j}\,\tilde{z}(t-j) \qquad (3.13)$$

where $\tilde{z}(i) \triangleq z(i) - \hat{z}(i/i-1)$. Alternately, the equation (3.13) can be rewritten completely in terms of z and $\tilde{z}$.

$$z(t) + \sum_{j=1}^{n} A_j(t)z(t-j) = \tilde{z}(t) + \sum_{j=1}^{n} B_{n+1-j}(t)\,\tilde{z}(t-j) \qquad (3.14)$$

The error covariance matrix is given in equation (3.15)

$$E\,[\tilde{z}(i)\,\tilde{z}^T(j)] = R_\epsilon(i)\delta_{ij} \qquad (3.15)$$

IV. Prediction and Filtering Estimates of the Signal y(i)

Let $\hat{y}(t/\tau)$ = Linear least squares estimate of y(t) based on the measurements $z(\tau), z(\tau-1)$ etc.

$\triangleq$ Arg $[\min E\|y(t)-f(z(\tau),z(\tau-1)\ \ldots)\|^2]$

where $f(\cdot)$ is a linear function.

The starting point in our investigations is the recursive equation (3.13) for the innovations.

(A) Prediction

Here we are interested in computing $\hat{y}(t/\tau)$, $t > \tau$.

$$\hat{y}(t+m/t) = \hat{z}(t+m/t)\ , \quad m > 0 \ \forall\ t \qquad (4.1)$$

We have already evaluated $\hat{z}(t+1/t)$. To evaluate $\hat{z}(t+m/t)$ with $m > 1$, we

start with the equation for z(t)

$$z(t+m) + - \sum_{j=1}^{n} A_j(t+m)z(t+m-j) + \sum_{j=1}^{n} B_{n+1-j}(t+m)\epsilon(t+m-j) \tag{4.2}$$

$\hat{z}(t+m/t)$ is obtained by projecting the right hand side of (4.2) onto the space spanned by z(t), z(t-1), etc. We note immediately that

$$\begin{aligned} \hat{\epsilon}(i/j) &= \text{Projection of } \epsilon(i) \text{ on the space spanned by } z(j), z(j-1), \ldots \\ &= 0 \quad \text{if} \quad j < i \\ &= \tilde{z}(i) \text{ if } j \geq i \end{aligned} \tag{4.3}$$

The above mentioned manipulations lead us to the predictor equation (4.4).

<u>Proposition 4</u>: Consider the signal process y(i) and the related observed process z(i) described in equations (2.1)-(2.4). Then their m-step optimal predictors are identical.

$$\hat{y}(t+m/t) = \hat{z}(t+m/t) \ , \ M = 1,2,\ldots \tag{4.1}$$

$\hat{z}(t+1/t)$ is given by the innovation equation (3.15). Then $\hat{z}(t+m/t)$, $m > 1$ can be computed recursively using $\hat{z}(t+1/t)\ldots,\hat{z}(t+m-1/t)$ and by measurements from the equation (4.4).

$$\begin{aligned} \hat{z}(t+m/t) &= - \sum_{j=1}^{m-1} A_j(t+m)\ \hat{z}(t+m-j/t) - \sum_{j=m}^{n} A_j(t+m)z(t+m-j) + \\ &\quad \sum_{j=m}^{n} B_{n+1-j}(t+m)\ \tilde{z}(t+m-j) \ , \ m \leq n \\ &= - \sum_{j=1}^{n} A_j(t+m)\ \hat{z}(t+m-j/t) \ , \ m > n \end{aligned} \tag{4.4}$$

The covariance matrices of the estimates are given below

$$\begin{aligned} &\text{Cov } [y(t+m)-\hat{y}(t+m/t)/z(\tau), \tau \leq t] = R_\epsilon(t+1)-R_\eta(t+1), \text{if } m = 1 \\ &= R_\epsilon(t+m)-R_\eta(t+m) + \sum_{j=1}^{m-1} B_{n+1-j}(t+m)R_\epsilon(t+m-j)B^T_{n+1-j}(t+m), \text{if } m > 1 \end{aligned} \tag{4.5}$$

[illegible]

[illegible] Then [illegible]

[illegible]

(5.3) [illegible]

Equation (4.5) can be derived by subtracting equation (4.4) from (4.2), squaring up either side and taking expectation.

(B) Filtering Estimate

We will compute $\hat{y}(t/t)$ starting from $\hat{z}(t/t-1)$. On account of the orthogonality of the innovation, we can represent $\hat{y}(t/t)$ as

$$\hat{y}(t/t) = z(t) - K_F(t)\,\tilde{z}(t) \tag{4.6}$$

or

$$y(t) - \hat{y}(t/t) = -\,\eta(t) + K_F(t)\,\tilde{z}(t) \tag{4.7}$$

To evaluate the gain $K_F(t)$, recall that $\{y(t) - \hat{y}(t/t)\} \perp \tilde{z}(t)$. Thus multiplying (4.7) on either side by $\tilde{z}^T(t)$ and taking expectations, we get

$$E\,[\eta(t)\epsilon^T(t)] = K_F(t)\,E\,[\epsilon(t)\epsilon^T(t)] \tag{4.8}$$

with the aid of (4.8) and (3.9) we get

$$K_F(t) = R_\eta(t)R_\epsilon^{-1}(t)$$

We can similarly derive a formula for covariance from equations (4.7) and (4.9). The results are expressed in proposition 5.

Proposition 5: The filtered estimate $\hat{y}(t/t)$ obeys equation (4.9)

$$\hat{y}(t/t) = z(t) - R_\eta(t)\,R_\epsilon^{-1}(t)\,\tilde{z}(t) \tag{4.9}$$

Let $$E\,[(y(t) - \hat{y}(t/t))(y(t) - \hat{y}(t/t)^T|z(\tau), \tau < t] \triangleq F(t/t)$$

$$F(t/t) = R_\eta(t) - R_\eta(t)R_\epsilon^{-1}(t)R_\eta(t) \tag{4.10}$$

V. Smoothing Estimates

In computing the smoothing estimate $\hat{y}(t/\tau)$, $\tau > t$, we use two different recursive schemes depending on whether t is fixed or τ is fixed.

Fixed point smoothing: compute $\hat{y}(t/\tau)$ with t fixed and τ varying.

Fixed interval smoothing: compute $\hat{y}(t/\tau)$ with t varying and τ fixed.

In certain situations, we need fixed lag smoothing; i.e., computing $\hat{y}(t/t+\Delta)$

for different t and fixed $\Delta > 0$. In this paper we shall obtain the fixed lag smoothing estimates from the fixed interval smoothing schemes.

(A) Fixed Point Smoothing

We would like to express $\hat{y}(t/\tau)$ in terms of $\hat{y}(t/\tau-1)$. By definition

$$y(t) - \hat{y}(t/\tau-1) \perp z(i) \quad i \leq \tau-1 \tag{5.1}$$

Hence we can write

$$\hat{y}(t/\tau) = \hat{y}(t/\tau-1) + K_{1s}(\tau)\,\tilde{z}(\tau) \tag{5.2}$$

where $K_{1s}(\tau)$ is an undetermined gain. Rewrite (5.2) as

$$y(t) - \hat{y}(t/\tau) = y(t) - \hat{y}(t/\tau-1) - K_{1s}(\tau)\,\tilde{z}(\tau) \tag{5.3}$$

To evaluate the gain $K_{1s}(\tau)$, use the fact

$$y(t) - \hat{y}(t/\tau) \perp \tilde{z}(\tau) \tag{5.4}$$

Equations (5.3) and (5.4) imply

$$K_{1s}(\tau) = E\,[(y(t)-\hat{y}(t/\tau-1))\,\tilde{z}^T(\tau)]\;\_E(\tilde{z}(\tau)\tilde{z}^T(\tau))]^{-1} \tag{5.5}$$

But $\quad E\,[(y(t) - \hat{y}(t/\tau-1))\,\tilde{z}^T(\tau)]$

$$= E\,[(z(t) - \hat{y}(t/\tau-1))\,\tilde{z}^T(\tau) - \eta(t)\,\tilde{z}^T(\tau)]$$

$$= -\,D(t,\tau) \tag{5.6}$$

Thus $\quad K_{1s}(\tau) = -\,D(t,\tau)\,R_{\epsilon}^{-1}(\tau) \quad (5.7)$

Similarly we can obtain an equation for the covariance of the estimate using equations (5.3) (5.6) and (5.7). The results are summarized in proposition 6.

Proposition 6: The smoothing estimate $\hat{y}(t/\tau)$ is recursively computed from (5.8) where the gains $D(t,\tau)$ are obtained from equation (3.9)

$$y(t/\tau) = y(t/\tau-1) - D(t,\tau)R_{\epsilon}^{-1}(\tau)\,\tilde{z}(\tau) \qquad \tau \geq 1 + t \tag{5.8}$$

Let $\quad \text{Cov}\,[\hat{y}(t/\tau)/z(i),\ i \leq \tau] \triangleq F(t/\tau)$

$$F(t/\tau) = F(t/\tau-1) - D(t,\tau)R_{\epsilon}^{-1}(\tau)D^T(t,\tau) \tag{5.9}$$

(B) Fixed Interval Smoothing

On many occasions, one wishes to compute the estimates $\hat{y}(\tau/\tau)$, $\hat{y}(\tau-1/\tau),\ldots,\hat{y}(1/\tau)$. Then the formula given in (5.8) is not very convenient if $(\tau-t) > n$. In that case, we manipulate the formula (5.8) to obtain the following proposition.

Proposition 7: If $t < \tau \leq t + n$

$$\hat{y}(t/\tau) = \hat{y}(t/t) - \sum_{i=1}^{(\tau-t)} D(t,t+1)\, R_{\epsilon}^{-1}(i)\, \tilde{z}(i) \tag{5.10}$$

If $\tau > t+n$

$$\hat{y}(t/\tau) = \hat{y}(t/t) - \sum_{i=1}^{n} D(t,t+1)\, \lambda_i(t) \tag{5.11}$$

where $\lambda_1(t),\ldots,\lambda_n(t)$ obey the backward difference equation (5.12)

$$\begin{aligned}
\lambda_1(t) &= R_{\epsilon}^{-1}(t+1)\, \tilde{z}(t+1) - B_1^T(t+n+1)\, \lambda_n(t+1) \\
\lambda_2(t) &= \lambda_1(t+1) - B_2^T(t+n+1)\, \lambda_n(t+1) \\
\lambda_i(t) &= \lambda_{i-1}(t+1) - B_i^T(t+n+1)\, \lambda_n(t+1) \\
\lambda_n(t) &= \lambda_{n-1}(t+1) - B_n^T(t+n+1)\, \lambda_n(t+1)
\end{aligned} \tag{5.12)a$$

with the final conditions

$$\lambda_i(\tau-n) = R_{\epsilon}^{-1}(\tau-n+i)\, \tilde{z}(\tau-n+i)\,, \quad i=1,.,n$$

The proposition is proved by induction in appendix 1.

It is possible to rewrite the backward difference equation (5.12)a as follows:

$$\begin{aligned}
&\lambda_n(t) + B_n^T(t+n+1)\, \lambda_n(t+1) + B_{n-1}^T(t+n+2)\, \lambda_n(t+2) \\
&+\ldots+ B_1^T(t+2n)\, \lambda_n(t+n) = R_{\epsilon}^{-1}(t+n)\, \tilde{z}(t+n)
\end{aligned} \tag{5.12)b$$

Equation (5.12)b can be regarded as the adjoint of the fundamental equation for innovation.

Proposition 7 is used in 2 different ways. In the first instance, τ is fixed and we want to compute $y(t-1/\tau)$, $y(\tau-2/\tau)$...recursively. This can be done as demonstrated diagrammatically.

$$\begin{array}{ccccccccc} & & y(t-2/\tau) & & y(t-1/\tau) & & y(t/\tau) & & y(\tau-n/\tau) \\ & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\ \ldots & \leftarrow & \lambda(t-2) & \leftarrow & \lambda(t-1) & \leftarrow & \lambda(t) & \leftarrow\ldots\leftarrow & \lambda(\tau-n) \\ & & & & & & & & \text{(given)} \end{array}$$

In the second case $\tau = t + \Delta$ where Δ is fixed. In this case we have to compute $y(1/1 + \Delta)$, $y(2/2 + \Delta)$,..., separately using proposition 7. The computations used in obtaining $\hat{y}(t/t + \Delta)$ will not be explicitly useful in computing $\hat{y}(t + 1/t + 1 + \Delta)$, in contrast to the earlier case.

VI. Comparison of Computational Aspects

As mentioned in the introduction, the recursive estimation schemes of this paper are useful in problems in which one is interested only in the estimates of the signal $y(i)$ and not that of the entire state vector $x(i)$ of dimension N. Indeed, in smoothing problems, it is hard to justify the computation of those components of the state vector other than $\hat{y}(t/\tau)$. The comparison depends on N, r (the dimension of measurement vector) and n (the maximal order of the difference equation). Note that $nr \geq N$. If $r = 1$, $N = n$. It is needless to say that the larger the ratio N/r, the larger will be the utility of our scheme over the existing ones.

In the linear estimation schemes of the literature, the basic equations are the Kalman filtering equations [1,2]. Using the state transition model of equation (2.5), the recursive equation for the filtered estimate $\hat{x}(i/i)$ (N-vector) and the N x N conditional covariance matrix P(i) are given below:

$$\hat{x}(i+1/i+1) = A(i)\hat{x}(i/i) + P(i+1)H^T(i+1)R_\eta^{-1}(i+1)\,[z(i+1) - H(i+1)A\hat{x}(i/i)] \tag{6.1}$$

$$P(i+1) = M(i+1) - M(i+1)H^T(i+1)\,[H(i+1)M(i+1)H^T(i+1) + R_\eta(i+1)]\,H(i+1)\,M(i+1) \tag{6.2a}$$

$$M(i+1) = A(i)P(i)A^T(i) + \beta(i)R_\varepsilon(i)\,\beta^T(i) \tag{6.2b}$$

The recursive computation of the covariance matrix P(i) in Kalman filter corresponds to the recursive evaluation of the $\{(n+1)r^2-1\}$ quantities represented by $B_1(t)\ldots,B_n(t)$ and $R_\epsilon(t)$ in our scheme. The amount of computation per iteration of the covariance equation (6.2) and the gains is

$$= \left\{\frac{3N^3}{2} + \frac{5}{2}N^2r + \frac{N^2}{2} + 3Nr^2 + \frac{Nr}{2}\right\}$$

The amount of computation per iteration of the B equation (3.6) and the gains

$$= \left(\frac{7}{2}n^2 + \frac{9}{2}n + 2\right)r^3 + \left(\frac{n(n+1)}{2} + 2n\right)r^2 + \frac{n(n+1)}{2}\,r$$

Thus the use of our estimation scheme results in a reduction in the <u>auxiliary computation</u> by a factor which roughly equals $\frac{3N^3}{7n^2r^3}$. Let us consider the filtering problem.

The ratio of the amount of computation per iteration of the Kalman filter (6.1) to our filter of equations (3.14) and (4.9) is $((N^2 + Nr)/2nr^2)$.

Let us consider the smoothing problems. In these problems, the storage aspects become important. Suppose we are concerned with a smoothing problem involving t_f time steps. In the smoothing schemes in the literature one has to store the measurement vectors z(t), the filtered estimates $\hat{x}(t)$ and the covariance matrices $P(t)$ for all t_f steps whereas here we have to store the measurements $z(t)$, innovations $\tilde{z}(t)$ and the coefficients $B_i(t)$. Thus the reduction in storage occurs by a factor of $t_f\left(\frac{N^2 + r + n}{(n+1)\ r^2+2r}\right)$. For comparing the relative amounts of computations, we have to treat fixed point smoothing and fixed interval smoothing separately. For fixed point smoothing the equations of Bryson and Frasier [3] are given below.

$$\hat{x}(t/\tau) = \hat{x}(t/\tau-1) + P(t/\tau)H^T(\tau)R_\eta^{-1}(\tau)[z(\tau)-H(\tau)\hat{x}(\tau/\tau-1)]$$
$$P(t/\tau) = P(t/\tau-1)\ [I-P(\tau)H^T(\tau)R_\eta^{-1}(\tau)H(\tau)]^T \qquad (6.3)$$

where $P(t/t) = P(t)$.

The ratio of multiplications per iteration of equation (6.3) over that of (5.8) is $\left(\frac{4N^2r + 2Nr^2 + N^2 + 2Nr + r^2}{nr^3 + 2r^2}\right)$.

Similarly, for fixed interval smoothing, the equations of Rauch[5] are

$$\hat{x}(t/\tau) = \hat{x}(t/t) + P(\tau)\,A^T(\tau)M^{-1}(\tau+1)[\hat{x}(t+1/\tau) - \hat{x}(t+1/t)] \qquad (6.4)$$

Comparing (6.4) with our scheme of equations (5.11) and (5.12), we see that (6.4) is highly inconvenient since it has to invert a N x N matrix at every instant in addition to doing other comparable jobs.

VII. Conclusions

We have developed estimation schemes for prediction, filtering and smoothing for linear discrete systems. The foundation of the entire paper rests on the recursive equation for the innovation or orthogonalized measurement. The greater the ratio N/r; i.e., the ratio of the number of state variables to the number of measurements, the smaller will be the relative amount of computation and storage of our schemes over the traditional ones. Moreover, in the schemes of this paper there is no need to express the dynamics of the signal in the state variable form.

It seems possible that these schemes could be modified to allow for the uncertainties regarding the noise covariances[12]. These aspects will be discussed in a later paper.

REFERENCES

1. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems", J. of Basic Engrg., Vol. 82, pp. 34-45, March 1960.

2. Kalman, R. E., "New Methods in Wiener Filtering Theory", Proc. of the First Symposium on Engineering Applications of Random Function Theory and Probability", (eds.) J. L. Bogdanoff and F. Kozin, John Wiley and Sons, 1963.

3. Bryson, A. E., and Frasier, "Smoothing for Linear and Nonlinear Dynamic System", presented at Optimum Synthesis Conf., Wright-Patterson AFB, Ohio, Sept. 1962.

4. Rauch, H. E., F. Tung and C. T. Streibel, "Maximum Likelihood Estimates of Linear Dynamic Systems", AIAA J., Vol. 3, No. 8, pp. 1445-1450, August 1965.

5. Rauch, H. E., "Solutions to the Linear Smoothing Problem", IEEE Trans. on Automatic Control, Vol. AC-8, pp. 371-372, October 1963.

6. Meditch, J., "Orthogonal Projection and Discrete Optimal Linear Smoothing", SIAM, J. on Control, Vol. 5, pp. 76-89, February 1967.

7. Wold, H., <u>A Study in the Analysis of Stationary Time Series</u>, Almquist and Wiksell, Upsala, 1938.

8. Kolmogorov, A. N., "Interpolation and Extrapolation of Stationary Random Sequences", (Trans.) Rand memorandum, RM-3090-PR.

9. Doob, J. L., <u>Stochastic Processes</u>, John Wiley and Sons, New York, 1963.

10. Whittle, P., <u>Prediction and Regulation</u>, D. Van Norstrand Co., New Jersey, 1963.

11. Kailath, T., "A Wold-Kolmogorov Approach to Linear Least-Squares Estimation", Parts I and II, Tech. Report No. 7050-13, Stanford University, Stanford, 1968.

12. Kashyap, R. L., "Optimal Identification and Prediction in Linear Systems", (to be published).

13. Whittle, P., "Recursive Relations for Predictors of Non-stationary Processes", J. Royal Statistical Soc., Series B, Vol. 27, No. 3, 1965, pp. 525-532.

APPENDIX 1

Proof of the Proposition 7

There is no need to prove (5.10) since it follows directly from (5.11); we shall prove (5.11) by induction.

Let $\bar{z}(i) \triangleq R_{\epsilon}^{-1}(i)\, \tilde{z}(i)$

From (5.8)

$$\hat{y}(t/\tau) = \hat{y}(t/t) - \sum_{i=t+1}^{\tau} D(t,i)\, \bar{z}(i) \qquad \text{(A-1)}$$

Let

$$h(\tau) = \sum_{i=t+1}^{\tau} D(t,i)h(i) \qquad \text{(A-2)}$$

Express $D(t,\tau)$ in (A-2) in terms of $D(t,\tau-1)$, $D(t,\tau-2)$ etc. using (3.9).

$$h(\tau) = \sum_{i=1}^{\tau-n-1} D(t,t+i)\bar{z}(t+i) + \sum_{k=1}^{n} D(t,\tau-n-1+k)[\bar{z}(\tau-n-1+k) - B_k^T(\tau)\bar{z}(\tau)] \qquad \text{(A-3)}$$

Recall that

$$\lambda_i(\tau-n) = \bar{z}(\tau-n+i) \quad , \quad i=1,\dots,n \qquad \text{(A-4)}$$

If we substitute (A-4) and (5.12) in (A-3) we get

$$h(\tau) = \sum_{i=1}^{\tau-n-1} D(t,t+i)\bar{z}(t+i) + \sum_{k=1}^{n} D(t,\tau-n-1+k)\lambda_k(\tau-n-1)$$

To establish induction, we assume the validity of (A-5) and demonstrate the validity of (A-6).

$$h(\tau) = \sum_{i=1}^{\tau-n-a} D(t,t+i)\bar{z}(t+i) + \sum_{j=1}^{n} D(t,\tau-n-a-j)\lambda_j(t-n-a) \qquad \text{(A-5)}$$

$$h(\tau) = \sum_{i=1}^{\tau-n-(a+1)} D(t, t+i)\bar{z}(t+i) + \sum_{j=1}^{n} D(t, \tau-n-j-\overline{a+1})\lambda_j(\tau-n-\overline{a+1}) \tag{A-6}$$

To prove this, consider (A-5)

$$\sum_{i=1}^{\tau-n-a} D(t, t+i)\ \bar{z}(t+i) + \sum_{j=1}^{n} D(t, \tau-n-a+j)\ \lambda_j(\tau-n-a)$$

$$= \sum_{i=1}^{\tau-n-a} D(t, t+i)\ \bar{z}(t+i) + \sum_{j=1}^{n-1} D(t, \tau-n-a+j)\ \lambda_j(\tau-n-a)$$

$$- \sum_{\ell=1}^{n} D(t, \tau-a+\ell n-1)\ B_\ell^T(\tau-a)\ \lambda_n(\tau-n-a)$$

$$= \sum_{i=1}^{\tau-n-(a+1)} D(t, t+i)\ \bar{z}(t+i) + \sum_{j=2}^{n} D(t, \tau-n-\overline{a+1}+j)\ \Big\{\lambda_j(\tau-n-a)$$

$$- B_j^T(\tau-a)\lambda_n(\tau-n-a)\Big\} + D(t, \tau-n-\overline{a+1}+1)\ \Big\{\bar{z}(\tau-n-a)-B_1^T(\tau-a)\lambda_n(\tau-n-a)\Big\}$$

$$= \sum_{i=1}^{\tau-n-(a+1)} D(t, t+i)\ \bar{z}(t+i) + \sum_{j=1}^{n} D(t, \tau-n-(a+1)+j)\ \lambda_j(\tau-n-\overline{a+1})$$

which is the desired result.

Repeated use of the induction directly yields us the required expression (5.11).