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. A method is given for computing the gains of the predictor and filter for the signal $y(i)$ and the corresponding state $x(i)$. The gains are computed recursively from the previous gains without involving the covariance matrix of the state. The computational advantages of the scheme are also discussed.

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 [1], [2], and smoothing estimates [3]–[6].

Even though these schemes have notational elegance, they seem to involve many redundant variables. Specifically, even though the schemes involve the evaluation of an $N \times N$ covariance matrix at every instant, the optimal gains for the filter involves only N elements of the covariance matrix. 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 [13].

It has been known for a long time [7]–[10] that instead of working with the measurements $z(i)$, $i = 1,2,\cdots$, it is very convenient to work with the new set of variables, $\epsilon(i) = (z(i) - \hat{z}(i))$, $i = 1,2,\cdots$, 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 $z(i)$. Furthermore, all other estimates of $x(i)$ can be expressed in terms of innovations [12]. A natural course of action is to obtain a recursive equation for the innovations involving only the measurements $z(i)$. Moreover, the gains of this equation should be computed in such a way that they depend only on the previous values of the gains, the system parameters, and nothing else. Finally a comparison will be made between this scheme and the traditional methods.

II. MODEL OF THE RANDOM PROCESS

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

$$z(i) = y(i) + \eta(i) \tag{1}$$

$$\begin{aligned} E(\eta(i)) &= 0 \\ E(\eta(i)\eta^T(j)) &= R_\eta(i)\delta_{ij} \\ E(y(i)\eta^T(j)) &= 0. \end{aligned} \tag{2}$$

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

$$y(t) + \sum_{j=1}^{n} A_{n+1-j}(t-j)y(t-j) = \sum_{j=1}^{n} C_{n+1-j}(t-j)\xi(t-j) \tag{3}$$

where the m_1 vector $\xi(i)$ has the statistical properties[1]

$$\begin{aligned} E(\xi(i)) &= 0 \\ E(\xi(i)\xi^T(j)) &= R_\xi(i)\delta_{ij} \\ E(\xi(i)\eta^T(j)) &= 0. \end{aligned} \tag{4}$$

In (4), the $r \times r$ matrices $A_j(t)$, $j = 1,\cdots,n$ and $r \times m_1$ matrices $C_j(t)$, $j = 1,\cdots,n$ are known for all relevant t.

For purposes of comparison we represent the random process of (1)–(4) in state-transition form in (5) where the state vector $x(i)$ is of dimension $N = nr$

$$\begin{aligned} x(i+1) &= \mathcal{A}(i)x(i) + C(i)\xi(i) \\ y(i) &= Hx(i) \\ z(i) &= Hx(i) + \eta(i). \end{aligned} \tag{5}$$

The $nr \times nr$ matrix $\mathcal{A}(i)$, $nr \times m_1$ matrix $C(i)$, and $r \times nr$ matrix H have the representation

$$\mathcal{A}(i) = \begin{bmatrix} O_r,O_r,\cdots,O_r, & -A_1(i) \\ I_r,O_r,\cdots,O_r, & -A_2(i) \\ O_r,I_r,\cdots,O_r, & -A_3(i) \\ O_r,\cdots,I_r, & -A_n(i) \end{bmatrix} \tag{6}$$

Manuscript received July 1, 1968; revised February 10, 1969, and September 18, 1969. This work was partially supported by the National Science Foundation under Grant GK-1970 and by NASA under Grant NGR 15-005-021.

The author is with the Department of Electrical Engineering, Purdue University, Lafayette, Ind.

[1] The assumption of uncorrelation between the $\eta(i)$ and $\xi(j)$ can be relaxed.

and

$$C^T(i) = [C_1^T(i), C_2^T(i), \cdots, C_n^T(i)]$$

$$H = [O_r, O_r, \cdots, O_r, I_r]$$

where O_r and I_r are the null and identity matrices of dimension $r \times r$, respectively.

The state vector $x(i)$ can also be represented in the partitioned form to bring out its relation to $y(i)$

$$x^T(i) = [x_1^T(i), x_2^T(i), \cdots, x_n^T(i)]$$

$$y(i) = x_n(i) \tag{7}$$

where $x_k(i)$ is an r vector for $k = 1, \cdots, n$.

The system of (1)–(3) presents a good starting point in the construction of stochastic input–output models in diverse fields, such as, hydrology, tide prediction where the precise relation between output and input may be unknown, and hence it may have to be approximated by linear relations such as (3). Moreover, the model in (1)–(3) has the appealing feature of involving far fewer parameters than the corresponding state variable representation. This feature is of crucial importance if we have to estimate the parameters from actual input–output data, as is the case in many problems of practical interest. However, there are some dynamical systems whose equations may be represented in the state-transition form, but not in the form of (1)–(3). For such systems, the results of the paper are not valid.

III. RECURSIVE EQUATION FOR INNOVATION

Let $\hat{z}(t \mid t-1)$ be the linear least squares (LLS) predictor of $z(t)$ based on the measurements $z(t-j), j \geq 1$, i.e., if $f_t(z(t-1), z(t-2), \cdots)$ is any linear function of the measurements $z(t-j), j \geq 1$ then $\hat{z}(t \mid t-1)$ is the function which minimizes the criterion function

$$E \,\| z(t) - f_t(z(t-1), z(t-2), \cdots) \|^2.$$

Let $\epsilon(t) \triangleq z(t) - \hat{z}(t \mid t-1)$ be innovation at time t. We are interested in developing an equation for the innovation when the $z(\cdot)$ process obeys (1)–(4). Moreover, we are particularly interested in developing an equation for the innovation process which involves only the innovations $\epsilon(\cdot)$ and the measurements $z(\cdot)$, and none of the extraneous quantities like the estimates of state vectors, etc. Such a form of the innovation equation is not available in the literature and, hence, the result is presented in Proposition 1. This equation forms the basis for all the computational methods of this paper.

Proposition 1

Consider the process $z(\cdot)$ defined in (1)–(4). The corresponding innovation process $\epsilon(\cdot)$ obeys

$$\epsilon(t) + \sum_{j=1}^{n} B_{n+1-j}(t-j)\epsilon(t-j) = z(t) + \sum_{j=1}^{n} A_{n+1-j}(t-j) z(t-j). \tag{8}$$

The process $\epsilon(\cdot)$ has the following second-order properties

$$E(\epsilon(t)) = 0, \quad E(\epsilon(i)\epsilon^T(j)) = R_\epsilon(i)\delta_{ij}. \tag{9}$$

The coefficients $B_1(t), \cdots, B_n(t)$ and $R_\epsilon(t)$ are defined recursively in terms of the known quantities like $A_i(t)$, $R_\eta(t)$, $R_\xi(t)$, etc.

$$\begin{aligned} R_\epsilon(t) = R_\eta(t) &+ \sum_{j=1}^{n} [A_{n+1-j}(t-j) R_\eta(t-j) A_{n+1-j}^T(t-j) \\ &+ C_{n+1-j}(t-j) R_\xi(t-j) C_{n+1-j}^T(t-j) \\ &- B_{n+1-j}(t-j) R_\epsilon(t-j) B_{n+1-j}^T(t-j)] \\ B_1^T(t) &= R_\epsilon^{-1}(t) R_\eta(t) A_1^T(t) \\ B_k^T(t) &= R_\epsilon^{-1}(t)[R_\eta(t) A_k^T(t) \\ &+ \sum_{j=1}^{k-1} (A_{n+1-j}(t-j) R_\eta(t-j) A_{k-j}^T(t-j) \\ &+ C_{n+1-j}(t-j) R_\xi(t-j) C_{k-j}(t-j) \\ &- B_{n+1-j}(t-j) R_\epsilon(t-j) B_{k-j}^T(t-j))], \\ &\qquad k = 2, \cdots, n. \end{aligned} \tag{10}$$

The initial conditions for the recursions in (10) can be expressed in terms of the initial value of the covariance matrix of the state vector $x(i)$ defined in (5). Let

$$M_{kn}(t) = E\{x_k(t) x_n^T(t) \mid z(t-j), j \leq 1\} \tag{11}$$

$$\begin{aligned} R_\epsilon(t) &= R_\eta(t) + M_{nn}(t) \\ B_1(t) &= A_1(t) - A_1(t) M_{nn}(t) R_\epsilon^{-1}(t) \\ B_k(t) &= (A_k(t) R_\eta(t) + M_{kn}(t)) R_\epsilon^{-1}(t), \\ &\qquad t = 0, 1, \cdots, n-1. \end{aligned} \tag{12}$$

Remark 1: The recursive computation of the gains $B_i(t)$ in (10) roughly corresponds to the computation of the $nr \times nr$-dimensional covariance matrix in the Kalman–Bucy filter. The precise relation between (10) and the Kalman–Bucy variance equation has been discussed in the Appendix.

Proof of Proposition 1: We define two vectors x and y to be orthogonal to each other if $E(xy^T) = 0$. We will begin with the well-known fact [7], [8], [10], [12] that the innovations $\epsilon(i)$ are orthogonal to one another obeying (9). In view of this fact, the innovations $\epsilon(1), \cdots, \epsilon(i)$ span the same space as the corresponding measurements $z(1), \cdots, z(i)$. Hence, without any loss of generality, we can represent the optimal estimate $\hat{z}(t \mid t-1)$ by the infinite series

$$\hat{z}(t \mid t-1) = -\sum_{j=1}^{n} A_{n+1-j}(t-j) z(t-j) + \sum_{j=1}^{\infty} B_{n+1-j}(t-j)\epsilon(t-j) \tag{13}$$

where the undetermined gains $B_k(\cdot), k = n, n-1, \cdots, -\infty$ have to be chosen to minimize $E \,\| z(t) - \hat{z}(t \mid t-1) \|^2$. We will show that the optimum value of the coefficients

$B_{n+1-j}(t)$ is zero for all $j \geq n + 1$ and the rest of coefficients $B_1(\cdot),\cdots,B_n(\cdot)$ obey (11).

By substituting (1) in (3) we get the following equation for $z(t)$

$$z(t) = \sum_{j=1}^{n} A_{n+1-j}(t-j)z(t-j) + \omega_1(t) \qquad (14)$$

where

$$\omega_1(t) = \eta(t) + \sum_{j=1}^{n} (A_{n+1-j}(t-j)\eta(t-j) + C_{n+1-j}(t-j)\xi(t-j). \qquad (15)$$

Subtracting (13) from (14) we get

$$\epsilon(t) = \omega_1(t) - \sum_{j=1}^{n} B_{n+1-j}(t-j)\epsilon(t-j). \qquad (16)$$

Minimizing $E\,||\,\epsilon(t)\,||^2$ with respect to the coefficients $B_{n+1-j}(t-j), j = 1,\cdots,\infty$, we get

$$B_{n+1-j}(t-j) = (E[\epsilon(t-j)\epsilon^T(t-j)])^{-1}E(\omega_1(t)\epsilon^T(t-j)).$$

For every value of $j \geq n+1$, every term of $\omega_1(t)$ is orthogonal to $\epsilon(t-j)$ by causality. Hence

$$E(\omega_1(t)\epsilon^T(t-j)) = 0, \quad \forall j \geq n+1$$

and

$$B_{n+1-j}(t-j) = 0, \quad j \geq n+1 \text{ and } \forall t. \qquad (17)$$

To determine the remaining coefficients $B_{n+1-j}(t-j)$, $j = 1,\cdots,n$ we adopt the following method. Substituting (17) in (16) we get

$$\epsilon(t) + \sum_{j=1}^{n} B_{n+1-j}(t-j)\epsilon(t-j) = \omega_1(t). \qquad (18)$$

Let us denote the left-hand side of (18) by $\omega_2(t)$. Equation (18) implies

$$E(\omega_2(t)\omega_2(t+\tau)) = E(\omega_1(t)\omega_1^T(t+\tau)), \quad \tau = 0,1,\cdots,n. \qquad (19)$$

Let

$$\mathcal{W}_i(t,t+\tau) = E(\omega_i(t)\omega_i^T(t+\tau)), \quad i = 1,2.$$

By definition of $\omega_2(t)$, we get

$$\mathcal{W}_2(t,t+1) = R_\epsilon(t)B_{n+1-i}^T(t) + \sum_{j=1}^{n-i} B_{n+1-n}(t-j) \cdot R_\epsilon(t-j)B_{n+1-j-i}^T(t-j), \quad i = 0,1,\cdots,n-1 \qquad (20)$$

and

$$\mathcal{W}_2(t,t+n) = R_\epsilon(t)B_1^T(t).$$

Similarly, by the definition of $\omega_1(t)$, we get

$$\mathcal{W}_1(t,t+i) = R_\eta(t)A_{n+1-i}^T(t) + \sum_{j=1}^{n-1} (B_{n+1-j}(t-j) \cdot R_\epsilon(t-j)\, B_{n+1-j-i}^T(t-j) + C_{n+1-j}(t-j)R_\xi(t-j)C_{n+1-j-i}^T(t-j)), \quad i = 0,1,\cdots,n-1 \qquad (21)$$

$$\mathcal{W}_1(t,t+n) = R_\eta(t)A_1^T(t)$$

where

$$B_{n+1}(t) \triangleq A_{n+1}(t) \triangleq 1, \quad \forall t.$$

Using (20) and (21) we can solve (19) or (22) for $B_1(t)\cdots B_n(t)$ and $R_\epsilon(t)$

$$\mathcal{W}_1(t,t+\tau) = \mathcal{W}_2(t,t+\tau), \quad \tau = 0,1,\cdots,n. \qquad (22)$$

Equations (22) are triangular in $B_1(t),\cdots,B_n(t)$ and, hence, we can solve for them. The solutions are displayed in (10).

The relation between the initial values for $B(t)$ and the covariance matrix of $x(t)$ is proved in the Appendix.

IV. PREDICTED AND FILTERED ESTIMATES OF SIGNAL $y(t)$ AND STATE $x(t)$

We will show that recursive equations can be given for the various estimators of the state $x(t)$ and signal $y(t)$ in terms of the gains $B_i(t)$, $i = 1,\cdots,n$ defined earlier.

A. One-Step Predictor of Signal $y(t)$

Let $\bar{y}(t)$ = LLS estimate of $y(t)$ based on $\{z(j), j \leq (t-1)\}$. Since $y(t) = z(t) - \eta(t)$, and $\eta(t)$ is an uncorrelated noise obeying (2)

$$\bar{y}(t) = \hat{z}(t \mid t-1) \qquad (23)$$

but

$$z(t) = \epsilon(t) + \hat{z}(t \mid t-1) = \epsilon(t) + \bar{y}(t). \qquad (24)$$

Substitute for $z(t)$ in (8) from (24) and obtain the required recursive equation (25) for the estimate $\bar{y}(i)$

$$\bar{y}(t) + \sum_{j=1}^{n} A_{n+1-j}(t-j)\bar{y}(t-j) = \sum_{j=1}^{n} K_{n+1-j}(t-j)(z(t-j) - \bar{y}(t-j)) \qquad (25)$$

where

$$K_i(t) = B_i(t) - A_i(t), \quad i = 1,\cdots,n. \qquad (26)$$

B. One-Step Predictor of State $x(t)$

Consider the system of (5) with the matrices $\mathcal{A}(i)$, $C(i)$,H defined in (6). Let $\bar{x}(t)$ = LLS estimate of $x(t)$ based on $\{z(j), j \leq t-1\}$. Using a technique similar to the one used in the proof of Proposition 1, we can show that $\bar{x}(i)$ obeys the recursive equation (27) where the gains $K_i(t)$ are given in (26)

$$\bar{x}(t+1) = \mathcal{A}(t)\bar{x}(t) + K(t)(z(t) - H\bar{x}(t))$$

$$K^T(t) = [K_1^T(t),K_2^T(t),\cdots,K_n^T(t)]. \qquad (27)$$

As a check on our result, we note that (27) is the familiar Kalman predictor. Hence the gain $K(t)$ must be identical with the Kalman predictor gain $K(t)$ defined recursively in (28). This statement is proved in the Appendix.

$$K(t) = \mathcal{A}(t)M(t)H^T[(HM(t)H^T + R_\eta(t))^{-1}]$$

$$M(t+1) = \mathcal{A}(t)M(t)\mathcal{A}^T(t) - K(t)(HM(t)H^T + R_\eta(t))^{-1}K^T(t) + C(t)R_\xi(t)C^T(t). \qquad (28)$$

C. *Multistep Predictor of Signal* $y(t)$

Let $\hat{y}(t \mid \tau)$ = LLS estimate of $y(t)$ based on $z(j)$, $j \leq \tau$, and $\hat{z}(t \mid \tau)$ = LLS estimate of $z(t)$ based on $z(j)$, $j \leq \tau$.

Here we are interested in computing $\hat{y}(t \mid \tau)$, $t > \tau$. We write from (1) that

$$\hat{y}(t + m \mid t) = \hat{z}(t + m \mid t), \quad m > 0 \ \forall t \tag{29}$$

We have already evaluated $\hat{z}(t + 1 \mid t)$. To evaluate $\hat{z}(t + m \mid t)$ with $m > 1$, we start with (8) for $z(t)$

$$z(t + m) = \sum_{j=1}^{n} \{-A_{n+1-j}(t + m - j)z(t + m - j) + B_{n+1-j}(t + m - j)\epsilon(t + m - j)\}. \tag{30}$$

$\hat{z}(t + m \mid t)$ is obtained by projecting the right-hand side of (30) onto the space spanned by $z(t)$,$z(t - 1)$, etc. We note immediately that

$$\begin{aligned} \hat{\epsilon}(i \mid j) &= \text{projection of } \epsilon(i) \text{ on the space spanned by } z(j), \\ &\quad z(j-1),\cdots, \\ &= 0, \quad \text{if } j < i, \\ &= \epsilon(i), \quad \text{if } j \geq i. \end{aligned} \tag{31}$$

Similarly, $\hat{z}(t \mid \tau) = z(t)$, if $\tau \geq t$.

Projecting every term of (30) onto the space spanned by $z(t)$, $z(t - 1)$, etc., we get

$$\hat{z}(t + m \mid t) = \{-\sum_{j=1}^{n} A_{n+1-j}(t + m - j)\hat{z}(t + m - j \mid t) + \sum_{j=1}^{n} B_{n+1-j}(t + m - j)\hat{\epsilon}(t + m - j \mid t)\}.$$

Substituting (31) in the preceding equation, we get

$$\begin{aligned} \hat{z}(t + m \mid t) = \{&-\sum_{j=1}^{n} A_{n+1-j}(t + M - j)\hat{z}(t + m - j \mid t) \\ &- \sum_{j=m}^{n} A_{n+1-j}(t + m - j)z(t + m - j) \\ &+ \sum_{j=m}^{n} B_{n+1-j}(t + m - j)\epsilon(t + m - j)\}, \\ &\qquad\qquad \text{if } m \leq n \\ = &-\sum_{j=1}^{n} A_{n+1-j}(t + m - j)\hat{z}(t + m - j \mid t), \\ &\qquad\qquad \text{if } m > n. \end{aligned} \tag{32}$$

The estimate $\hat{z}(t + m \mid t)$ can be computed recursively starting from $\hat{z}(t + 1 \mid t)$ using (32). The covariance matrices of the estimates are

$$\begin{aligned} &\operatorname{cov}\,[y(t + m) - \hat{y}(t + m \mid t) \mid z(\tau \leq t)] \\ &\quad = R_\epsilon(t + 1) + R_\eta(t + 1), \quad \text{if } m = 1 \\ &\quad = R_\eta(t + m) + R_\epsilon(t + m) \\ &\qquad + \sum_{j=1}^{m-1} B_{n+1-j}(t + m - j)R_\epsilon(t + m - j) \\ &\qquad \cdot B_{n+1-j}{}^{T}(t + m - j), \quad \text{if } m > 1. \end{aligned} \tag{33}$$

Equation (33) can be derived by subtracting (32) from (30), scalar multiplying either side by itself, and taking expectation.

D. *Filtering Estimate of Signal* $y(t)$

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

$$\hat{y}(t \mid t) = \bar{y}(t) + K_A(t)\epsilon(t).$$

Adding and subtracting $z(t)$, and using (23) we get

$$\hat{y}(t \mid t) = z(t) - K_F(t)\epsilon(t)$$

where $K_F(t) \triangleq (-K_A(t) + I)$ is an undetermined gain.

$$y(t) - \hat{y}(t \mid t) = -\eta(t) + K_F(t)\tilde{z}(t). \tag{34}$$

To evaluate the gain $K_F(t)$, recall that $\{y(t) - \hat{y}(t \mid t)\}$ is orthogonal to $\epsilon(t)$. Thus, multiplying (34) on either side by $\epsilon^T(t)$ and taking expectations, we get

$$E[\eta(t)\epsilon^T(t)] = K_F(t)E[\epsilon(t)\epsilon^T(t)]. \tag{35}$$

It is easy to show from (18) that $E[\eta(t)\epsilon^T(t)] = R_\eta(t)$. Hence

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

Thus the filtered estimate

$$\begin{aligned} \hat{y}(t \mid t) &= z(t) - R_\eta(t)R_\epsilon^{-1}(t)\epsilon(t) \\ &= \bar{y}(t) + (I - R_\eta(t)R_\epsilon^{-1}(t))\epsilon(t). \end{aligned} \tag{36}$$

We can also show that

$$\begin{aligned} \operatorname{cov}\,[y(t) \mid z(j), j \leq t] &\triangleq F(t \mid t) \\ &= R_\eta(t) - R_\eta(t)R_\epsilon^{-1}(t)R_\eta(t). \end{aligned} \tag{37}$$

E. *Filtered Estimate of State* $x(t)$

Let $\hat{x}(t \mid t)$ = LLS estimate of $x(t)$ based on $z(j)$, $j \leq t$. Using the orthogonality of innovations, we can write

$$\hat{x}(t \mid t) = \bar{x}(t) + G(t)z(t) \tag{38}$$

where

$$\begin{aligned} G^T(t) &= [G_1{}^T(t), G_2{}^T(t), \cdots, G_n{}^T(t)] \\ G_j(t) &= B_{j+1}(t) - A_{j+1}(t)R_\eta(t)R_\epsilon^{-1}(t), \quad j = 1,\cdots,n. \end{aligned} \tag{39}$$

Combining (38) and (27), the following recursive equation can be written for $\hat{x}(t \mid t)$

$$\begin{aligned} \hat{x}(t + 1 \mid t + 1) &= \mathcal{A}(t)\hat{x}(t \mid t) \\ &\quad + G(t + 1)(z(t + 1) - H\mathcal{A}(t)\hat{x}(t \mid t)) \\ \hat{y}(t \mid t) &= H\hat{x}(t \mid t). \end{aligned} \tag{40}$$

As a check on the results, note that (40) is the familiar Kalman filter and, hence, the gain $G(t)$ in it must be identical with the Kalman filter gain in

$$G(t) = M(t)H^T(HM(t)H^T + R_\eta(t))^{-1}. \tag{41}$$

We will establish the equivalence of (39) and (41) in the Appendix. In a similar manner, we can compute smoothing estimates. For details, see [13].

V. Comparison of Computational Needs

There are two aspects to be considered, namely the amount of computations involved and the amount of storage and bookkeeping involved.

Let us consider these aspects for the problem in which the entire state vector has to be estimated. Then the only point of difference between our scheme and others lies in the computation of gains $B_i(t)$. In other words, we want to compare the computations needed to compute the predictor gains from (10) with that obtained from (28). We will only consider the multiplications since they constitute the principle part of the computations. Since different programs for the same set of equations lead to different amounts of computation depending on the complexity of the program, we will only consider the program which leads to the minimum number of computations in each case, exploiting special features like symmetric matrices, matrices with many zero entries like $\mathcal{A}(i)$, etc.

The minimum number of multiplications per iteration by Kalman's method (28) is[2]

$$\begin{aligned}
&\{\text{multiplication for } K(t)\}\\
&\quad+\{\text{multiplication for I term of } M(t+1)\}\\
&\quad+\{\text{multiplication for II term of } M(t+1)\}\\
&\quad+\{\text{multiplication for III term of } M(t+1)\}\\
&=\{2nr^3\}+\{n^2r^3+\tfrac{1}{2}n(r^3+r^2)+(n^2-n)r^3\}\\
&\quad+\{n^2r^3+\tfrac{1}{2}n(r^3+r^2)\}+\{\tfrac{1}{2}(n^2-n)(rm_1^2+r^2m_1)\\
&\quad+\tfrac{1}{2}nm_1(r^2+r)+n\min(m_1r^2,m_1^2r)\}\\
&=[r^3(3n^2+2n)]+[nr^2]+[\tfrac{1}{2}(n^2-n)(rm_1^2+r^2m_1)\\
&\quad+\tfrac{1}{2}nm_1(r^2+r)+n\min(m_1r^2,m_1^2r)]. \qquad (42)
\end{aligned}$$

The minimum number of multiplications per iteration of (10) is

$$\begin{aligned}
&\{\text{multiplication for terms like } A_iR_\eta A_j\}\\
&\quad+\{\text{multiplication for terms like } B_iR_EB_j\}\\
&\quad+\{\text{multiplication for terms like } C_iR_\xi C_j\}\\
&\quad+\{\text{miscellaneous multiplication}\}\\
&=\{n^2r^3+\tfrac{1}{2}n(r^3+r^2)\}+\{n^2r^3+\tfrac{1}{2}n(r^3+r^2)\}\\
&\quad+\{\tfrac{1}{2}(n^2-n)(rm_1^2+r^2m_1)+\tfrac{1}{2}nm_1(r^2+r)\\
&\quad+n(\min(m_1r^2,m_1^2r))\}+\{2nr^3\}\\
&=[r^3(2n^2+3n)]+[nr^2]+[\tfrac{1}{2}(n^2-n)(rm_1^2+r^2m_1)\\
&\quad+\tfrac{1}{2}nm_1(r^2+r)+n(\min(m_1r^2,m_1^2r))]. \qquad (43)
\end{aligned}$$

Comparing (42) and (43), we see that their last two terms in square brackets are identical. Their first terms indicate that the use of (10) results in a reduction of computation by about 33 percent provided the time interval of estimation is large. Moreover, if we decrease the complexity of program, say by using standard matrix multiplication subroutines, then the corresponding computation for (10) may be at most double that given in (43), whereas for (28) it will be many times the amount in (42). This is because (28) involves $nr \times nr$ matrices, whereas (10) involves only $r \times r$ matrices. Secondly, the bookkeeping and storage requirements of (10) is far less than with (28) since the former does not involve the storage of $nr \times nr$ matrix $M(i)$ at each instant.

Similarly, if we are interested in $\bar{y}(t)$ alone, (25) gives the required answer in a less complex manner than (27), since the latter involves far fewer variables than the former.

VI. Conclusions

The paper gives a new method of computing the gains for the optimal filter and predictor. The gains at any instant are expressed entirely in terms of the previous gains and the system parameters. This scheme of computing the gains is of special importance in identification problems in which it is imperative to keep the number of extraneous parameters to a minimum. Moreover, this method results in a decrease in computation of about 33 percent over the Kalman scheme, in estimating the state from noisy measurements, when the time interval of estimation is large.

Appendix

Relation of Kalman's Variance Equation to the Equation for $B_i(t)$

Define the matrices $\mathcal{A}(t), C(t), H,$ as in (6). Consider the equations for the gain $K(t)$ of the Kalman predictor in (27)

$$K(t) = \mathcal{A}(t)M(t)H^TR_\epsilon^{-1}(t) \qquad (44)$$

$$\begin{aligned}M(t+1) &= \mathcal{A}(t)M(t)\mathcal{A}^T(t) - K(t)R_\epsilon(t)K^T(t)\\ &\quad + C(t)R_\xi(t)C^T(t) \qquad (45)\end{aligned}$$

$$R_\epsilon(t) = HM(t)H^T + R_\eta(t). \qquad (46)$$

Define a set of gains $B(t)$ so that

$$B(t) \triangleq K(t) + A(t) \qquad (47)$$

where

$$B(t) = \begin{bmatrix} B_1(t) \\ \vdots \\ B_n(t) \end{bmatrix}, \quad A(t) = \begin{bmatrix} A_1(t) \\ \vdots \\ A_n(t) \end{bmatrix}.$$

1) To show that the gains $B_i(t)$ and $R_\epsilon(t)$ defined in (44) and (45) obey the recursive equation (10).

Proof: Let

$$M(t) = \left|\begin{array}{c:c} \mathfrak{M}(t) & \bar{m}(t) \\ \hdashline \underbrace{\bar{m}^T(t)}_{(n-1)r} & \underbrace{M_{nn}(t)}_{r} \end{array}\right| \begin{array}{l} \}(n-1)r \\ \}r \end{array}$$

---

[2] $\min(a,b) = a, \quad \text{if } a \leq b$
$\qquad\qquad = b, \quad \text{if } a \geq b.$

and

$$\bar{m}(t) = \begin{bmatrix} M_{1n}(t) \\ \vdots \\ M_{n-1,n}(t) \end{bmatrix}, \quad \bar{A}(t) = \begin{bmatrix} A_2(t) \\ \vdots \\ A_n(t) \end{bmatrix},$$

$$\bar{B}(t) = \begin{bmatrix} B_2(t) \\ \vdots \\ B_n(t) \end{bmatrix}, \quad (n-1)r \text{ rows.} \tag{48}$$

By definition of H, $HM(t)H^T = M_{nn}(t)$ so that

$$M_{nn}(t) = R_\epsilon(t) - R_\eta(t). \tag{49}$$

Let us simplify (44) with the aid of (49)

$$B_1(t) - A_1(t) = -A_1(t)M_{nn}(t)R_\epsilon^{-1}(t) \tag{50}$$

$$\bar{B}(t) - \bar{A}(t) = \bar{m}(t)R_\epsilon^{-1}(t) - \bar{A}(t)M_{nn}(t)R_\epsilon^{-1}(t) \tag{51}$$

From (48)–(51) we get

$$B_1(t) = A_1(t)R_\eta(t)R_\epsilon^{-1}(t) \tag{52}$$

$$\bar{m}(t) = \bar{B}(t)R_\epsilon(t) - \bar{A}(t)R_\eta(t). \tag{53}$$

Combining (52) and (53) we get

$$\begin{bmatrix} 0 \\ \\ \bar{m}(t) \end{bmatrix} = B(t)R_\epsilon(t) - A(t)R_\eta(t). \tag{54}$$

Let us consider the terms in (45)

$$\begin{aligned} \mathcal{A}(t)M(t)\mathcal{A}^T(t) &= \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \mathcal{M}(t) \end{array}\right] + A(t)M_{nn}(t)A^T(t) \\ &\quad - A(t)(0,\bar{m}^T(t)) - \begin{bmatrix} 0 \\ \\ \bar{m}(t) \end{bmatrix} A^T(t). \end{aligned} \tag{55}$$

Define a matrix

$$F = \begin{array}{c} \left[\begin{array}{c:c} 0 & 0 \\ \hdashline I & 0 \end{array}\right] \begin{array}{l} \}r \\ \}(n-1)r \end{array} \\ \underbrace{\qquad\qquad}_{(n-1)r} \;\; \underbrace{\;\;}_{r} \qquad\qquad \end{array}$$

$$\left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & \mathcal{M}(t) \end{array}\right] = FM(t)F^T. \tag{56}$$

Substitute (54)–(56) in the right-hand side of (45) and simplify

$$\begin{aligned} M(t+1) &= FM(t)F^T + A(t)R_\eta(t)A^T(t) \\ &\quad + C(t)R_\xi(t)C^T(t) - B(t)R_\epsilon(t)B^T(t). \end{aligned} \tag{57}$$

Repeated use of (59) gives us

$$\begin{aligned} M(t+1) = \sum_{i=0}^{n-1} \{&F^iA(t-i)R_\eta(t-i)A^T(t-i)(F^i)^T \\ &+ F^iC(t-i)R_\xi(t-i)C^T(t-i)(F^i)^T \\ &- F^iB(t-i)R_\epsilon(t-i)B^T(t-i)(F^i)^T \\ &+ F^nM(t-n)(E^n)^T. \end{aligned} \tag{58}$$

The last term in (58) is zero since $F^n = 0$. Simplify (58) by noting that

$$FA(t) = \begin{bmatrix} 0 \\ \\ A_2(t) \\ \vdots \\ A_n(t) \end{bmatrix}, \quad F^2A(t) = \begin{bmatrix} 0 \\ \\ 0 \\ \\ A_3(t) \\ \vdots \\ A_n(t) \end{bmatrix},$$

$$F^{n-1}A(t) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \\ A_n(t) \end{bmatrix}$$

similarly for $F^iB(t)$ and $F^iC(t)$. From (58) the elements (matrices) in the last column of $M(t+1)$ equal

$$\begin{aligned} M_{k,n}^T(t+1) = \sum_{j=1}^{k} \{&A_{n+1-j}(t_j)R_\eta(t_j)A_{k-j}^T(t_j) \\ &+ C_{n+1-j}(t_j)R_\xi(t_j) \\ &- B_{n+1-j}(t_j)R_\epsilon(t_j)B_{k-j}^T(t_j)\} \\ &t_j = t+1-j, \quad k = 1,\cdots,n. \end{aligned} \tag{59}$$

But from (53)

$$M_{k,n}^T(t) = R_\epsilon(t)B_{k+1}^T(t) - R_\eta(t)A_{k+1}^T(t). \tag{60}$$

Substitute (60) in (61). This result, in conjunction with (59), constitutes the required recursive equation (10). Note that it is also possible to solve for all the elements of $M_{i,j}(t)$ only in terms $B_i(j)$, $A_i(j)$, and $C_i(j)$.

2) Consider the Kalman filter equation in (40) where the gain $G(t)$ is defined by

$$G(t) = M(t)H^TR_\epsilon^{-1}$$

where $M(t)$ has been defined in (45). We want to establish the relation (61) between $G(t)$ and $B(t)$, where $B(t)$ has been defined in (47)

$$G_i(t) = B_{i+1}(t) - A_{i+1}(t)R_\eta(t)R_\epsilon^{-1}(t), \quad i = 1,2,\cdots,n. \tag{61}$$

Proof: Let

$$\bar{G}^T(t) = [G_1^T(t),\cdots,G_{n-1}^T(t)]$$

by definition of Kalman filter gain

$$G(t) = M(t)H^TR_\epsilon^{-1}(t)$$

or

$$\begin{bmatrix} \bar{G}(t) \\ \\ G_n(t) \end{bmatrix} = \begin{bmatrix} \bar{m}(t) \\ \\ M_{nn}(t) \end{bmatrix} R_\epsilon^{-1}(t).$$

Hence

$$\begin{aligned} \bar{G}(t) &= \bar{m}(t)R_\epsilon^{-1}(t) \\ &= (\bar{B}(t)R_\epsilon(t) - \bar{A}(t)R_\eta(t))R_\epsilon^{-1}(t) \\ &= \bar{B}(t) - \bar{A}(t)R_\eta(t)R_\epsilon(t) \end{aligned} \tag{62}$$

$$\begin{aligned} G_n(t) &= M_{nn}(t)R_\epsilon^{-1}(t) \\ &= I - R_\eta(t)R_\epsilon^{-1}(t). \end{aligned} \tag{63}$$

Equations (62) and (63) yield the required result (61) recalling

$$B_{n+1}(t) \triangleq A_{n+1}(t) \triangleq I.$$

3) The relation between $B_i(t)$, $R_\epsilon(t)$, and $M_{kj}(t)$ is now considered. From (49) we have

$$R_\epsilon(t) = R_\eta(t) + M_{nn}(t) \tag{64}$$

from (49) and (50) we have

$$B_1(t) = A_1(t) - A_1(t)M_{nn}(t)R_\epsilon^{-1}(t) \tag{65}$$

and from (48) and (53) we have

$$B_k(t) = (A_k(t)R_\eta(t) + M_{k,n}(t))R_\epsilon^{-1}(t). \tag{66}$$

Equations (64)–(66) have been rewritten as (12) to yield the initial conditions $R_\epsilon(t)$, $B_k(t)$, $k = 1,\cdots,n$, $t = 0,1,\cdots,n-1$, for the recursive equation (10).

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Engrg.*, ser. D, vol. 82, pp. 35–45, March 1960.

[2] R. E. Kalman, "New methods in Wiener filtering theory," *Proc. 1st Symp. Engineering Applications of Random Function Theory and Probability*, J. L. Bogdanoff and F. Kozin, Eds. New York: Wiley, 1963.

[3] A. E. Bryson and M. Frasier, "Smoothing for linear and nonlinear dynamic system," presented at Optimum Synthesis Conf., Wright-Patterson Air Force Base, Ohio, September 1962.

[4] H. E. Rauch, F. Tung, and C. T. Streibel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, pp. 1445–1450, August 1965.

[5] H. E. Rauch, "Solutions to the linear smoothing problem," *IEEE Trans. Automatic Control* (Short Papers), vol. AC-8, pp. 371–372, October 1963.

[6] J. Meditch, "Orthogonal projection and discrete optimal linear smoothing," *SIAM J. Control*, vol. 5, pp. 76–89, February 1967.

[7] H. Wold, *A Study in the Analysis of Stationary Time Series.* Upsala, Sweden: Almquist and Wiksell, 1938.

[8] A. N. Kolmogorov, "Interpolation and extrapolation of stationary random sequences" (transl.), RAND Corp., Tech. Memo. RM-3090-PR.

[9] J. L. Doob, *Stochastic Process.* New York: Wiley, 1963.

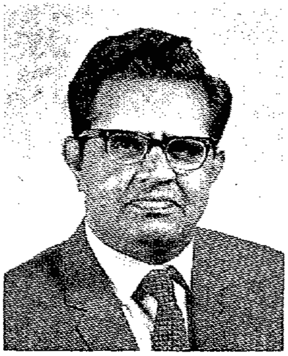
[10] P. Whittle, *Prediction and Regulation.* Princeton, N. J.: Van Nostrand, 1963.

[11] P. Whittle, "Recursive relations for predictors of non-stationary processes," *J. Roy. Statist. Soc.*, ser. B, vol. 27, no. 3, pp. 525–532, 1965.

[12] T. Kailath, "A Wold–Kolmogorov approach to linear least-squares estimation, pts. I and II," Stanford University, Stanford, Calif., Tech. Rept. 7050-13, 1968.

[13] R. L. Kashyap, "A new method of recursive estimation in discrete linear systems," Dept. of Elec. Engrg., Purdue University, Lafayette, Ind., Tech. Rept. TR-EE 68-13, June 1968.

[14] K. J. Astrom, "Computer control of paper machine—an application of linear stochastic control theory," *IBM J.*, pp. 389–405, July 1967.



R. L. Kashyap was born in Mysore, India, on March 28, 1938. He received the D.I.I.Sc. and M.E. degrees from the Indian Institute of Science, Bangalore, India, in 1960 and 1962, respectively, and the Ph.D. degree in engineering from Harvard University, Cambridge, Mass., in 1965.

During 1965–1966 he was a postdoctoral fellow in applied mathematics at Harvard University. In 1966 he was appointed an Assistant Professor of Electrical Engineering at Purdue University, Lafayette, Ind., where he currently holds the rank of Associate Professor. His current interests include statistical inference and prior probabilities, estimation theory, and control of plants with uncertainty.

Dr. Kashyap is a member of Sigma Xi. In 1966 he won the National Electronics Conference annual award for his paper "Optimization of Stochastic Finite-State Machines," which was judged to be the best original paper presented.