LINEAR LEAST-SQUARES ESTIMATION OF DISCRETE-TIME STATIONARY PROCESSES BY MEANS OF BACKWARD INNOVATIONS

Anders Lindquist

Reprinted from

Control Theory, Numerical Methods
and Computer Systems Modelling.

Proceedings of the International Symposium,
Rocquencourt, France, June 17-21, 1974,
edited by A. Bensoussan and J. L. Lions.

Lecture Notes in Economics and
Mathematical Systems, Vol. 107, pp. 44—63,
Springer Verlag, Berlin, Heidelberg and New York, 1975.

Linear Least-Squares Estimation of Discrete-Time Stationary Processes by Means of Backward Innovations

Anders Lindquist†

1. Introduction

There has lately been a considerable interest in fast algorithms for recursive linear least squares estimation. This is clearly witnessed by a series of recent papers by Casti, Kalaba and Murthy [1], Rissanen [2], Casti and Tse [3], Kailath [4,5,6], and Lindquist [7,8,9], to just mention some contributions related to the work presented in this paper. Among these [1,3,4,5,8,9] concern stochastic processes in continuous time, while [2,6,7,9] deal with discrete-time processes. For an account on the relation between these papers, among which [4,5,6,7,8] and to a certain extent [3] are concerned with Kalman-Bucy filtering, we refer the reader to [9] where we also try to clarify the connections between these recent results and some classical results in filtering [10,11,12], the theory of polynomials orthogonal on the unit circle [13,14,15] and the theory of Fredholm integral equations [16,17,18,19].

In this paper we shall consider the algorithm for the discrete-time Kalman-Bucy gain first presented in [7]. In the important case when the number of outputs are much fewer that the dimension n of the system, this algorithm, which holds for stationary systems, requires a number of scalar equations of order n rather than n^2 as with the conventional method based on the Riccati equation.

The continuous-time counterpart of this algorithm was first derived by Kailath [4] by means of a decomposition of the Riccati equation due to Bucy [20]. Since this decomposition holds for all constant Kalman-Bucy models, Kailath has been able to obtain similar equations [5] also for certain nonstationary systems, although the computional advantage of these algorithms rely heavily on the possible low rank of a certain $n \times n$ - matrix—a condition which is automatically fulfilled in the stationary case. However, in the sequel only stationary models will be considered.

Our discrete-time result [7] was obtained independently by a method based on the work [10,11,12,13, 14,15], and somewhat suprisingly these equations are more complicated than their continuous-time counterparts. In a subsequent paper [6] Kailath et al have demonstrated that the algorithm [7] can also be derived from the discrete-time Riccati-equation by a decomposition akin to that of [4], which of course is only to be expected. However, now the decomposition is no longer unique so that several versions of the algorithm emerge. Unfortunately, the Riccati - approach [6] gives very little insight into the relation between these versions.

This paper will be devoted to a more thorough study of the discrete - time algorithm [7]. By proceeding from basic principles we shall be able to present a number of different versions *and* explain the relation between them. This will be done in Sections 3 and 4 by an essentially *deterministic* technique, the

†Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506, U.S.A.

basic feature of which is a certain reversed-time operation previously used in [9]. In Section 5 we provide a *stochastic* interpretation of this approach in terms of the *forward and backward innovation processes.* This method, which suggested the investigation of Sections 3 and 4, was first used in [8] and is based on the simple observation that the available data can be orthogonalized either in the forward or the backward direction, thereby providing two different innovation processes.

As we did in [7,8,9], we shall find it convenient to develop our results in a smoothing context, although our primary interest is the one-step predictor and the pure filter. This is of course no coincidence, because forward and backward recursions do play a central role in the theory of linear smoothing. In particular our approach reminds one of the interpretation, due to Mayne and Fraser, of the optimal smoother as a combination of a forward and backward filter [21,22,23,24]. However, in contrast to the basic idea of the Mayne-Fraser technique we use the backward innovation for the filtering part and the forward for the smoothing part. In fact this is the key idea of our method.

Although our main interest is in Kalman-Bucy filtering, we shall deliberately introduce the Gauss-Markov condition as late in our development as possible. The reason for this is to pinpoint what properties of the algorithms depend on this assumption and what properties hold for wide sense stationary processes in general.

Finally, to explain why our discrete-time results are considerably more complex than the corresponding results in continuous time, in Section 6 we shall briefly discuss the analogous continuous-time approach.

2. Preliminaries

Let $\{x(t); t = \ldots -2, -1, 0, 1, 2, \ldots\}$ be an n-dimensional vector process with zero mean and covariance function

$$\Gamma_{ts} = E\{x(t)\, x(s)'\} \tag{2.1}$$

and consider the linear least-squares estimate $\hat{x}(t|r)$ of $x(t)$ given $z(0), z(1), \ldots, z(r)$, where the m-dimensional process z is defined by

$$z(t) = H\, x(t) + w(t). \tag{2.2}$$

Here H is an m x n - matrix and w is a zero mean white noise sequence with covariance

$$E\{w(t)\, w(s)'\} = I\, \delta_{ts}. \tag{2.3}$$

We assume that x and w are uncorrelated.

Now, denoting the estimation error

$$\tilde{x}(t|r) = x(t) - \hat{x}(t|r), \tag{2.4}$$

we define the error covariance function

$$P_r(t,s) = E\{\tilde{x}(t|r)\,\tilde{x}(s|r)'\} \tag{2.5}$$

which provides us with the weightingfunction of the estimate:

Proposition 2.1: $$\hat{x}(t|r) = \sum_{s=0}^{r} P_r(t,s)\,H'z(s) \tag{2.6}$$

Proof: The projection theorem implies that

$$E\{\tilde{x}(t|r)\,z(s)'\} = E\{\tilde{x}(t|r)\,x(s)'\}H' - E\{\hat{x}(t|r)\,w(s)'\} \tag{2.7}$$

is zero for all $s = 0,1,\ldots,r$. Therefore inserting

$$\hat{x}(t|r) = \sum_{s=0}^{r} N_s z(s)$$

in (2.7), we have

$$N_s = E\{\tilde{x}(t|r)\,x(s)'\}H'$$

from which our assertion follows.■

However note that (2.5) defines $P_r(t,s)$ for arguments ($s < 0$, $s > r$) which are not needed in the representation (2.6). The function P has the following properties:

Proposition 2.1: *The function* P *satisfies*

$$P_r(s,t) = P_r(t,s)', \tag{2.8}$$

and is the unique solution of the system of linear equations

$$P_r(t,s) + \sum_{i=0}^{r} \Gamma_{ti}\,H'H\,P_r(i,s) = \Gamma_{ts}. \tag{2.9}$$

Moreover it satisfies the relation

$$P_{r+1}(t,s) = P_r(t,s) - P_{r+1}(t,r+1)H'HP_r(r+1,s) \tag{2.10}$$

Proof: Relation (2.8) follows directly from the definition. To obtain (2.9) insert (2.6) into

$$P_r(t,s) = E\{x(t)\,\tilde{x}(s|r)'\}\,.$$

Since the matrix T formed by the blocks

$$T_{ij} = I\,\delta_{ij} + \Gamma_{ij}H'H$$

is nonsingular (it is the sum of an identity matrix and the product of two non-negative matrices), (2.9) has a unique solution. Finally, to prove (2.10) first note that for $t = 0,1,\ldots,r$ we can write (2.9) as

$$\sum_{i=0}^{r} T_{ti}\,P_r(i,s) = \Gamma_{ts}\,,$$

and for the same values of t we also have

$$\sum_{i=0}^{r} T_{ti}\, P_{r+1}(i,s) + \Gamma_{t,r+1} H'HP_{r+1}(r+1,s) = \Gamma_{ts}.$$

Upon eliminating Γ between these two equations we obtain

$$\sum_{i=0}^{r} T_{ti}\, [P_{r+1}(i,s) - P_r(i,s) + P_r(i,r+1)H'HP_{r+1}(r+1,s)] = 0 \tag{2.11}$$

and since T is nonsingular the quantities within the square brackets must be zero. Hence we can exchange T_{ti} for $\Gamma_{ti}H'H$ in (2.11), where t is now arbitrary, and apply (2.9) to cancel all sums. Then take the transpose and invoke (2.8) to obtain (2.10). ∎

We may regard (2.9) as the discrete-time analog of the Fredholm integral equation occurring in the continuous-time theory and (2.10) as the corresponding *Bellman-Krein equation* [17,18]. (Also see [9].)

In Section 5 we shall need the innovation process

$$\nu(t) = z(t) - H\,\hat{x}\,(t|t-1). \tag{2.12}$$

It is well-known and easy to show (see e.g. [25]) that ν is a white noise process

$$E\{\nu(s)\nu(t)'\} = R_t\, \delta_{ts}, \tag{2.13}$$

where

$$R_t = I + HP_{t-1}(t,t)H'. \tag{2.14}$$

In the same way it can be shown that

$$\mu(t) = z(t) - H\,\hat{x}(t|t) \tag{2.15}$$

is also a white noise process:

$$E\{\mu(s)\,\mu(t)'\} = \bar{R}_t\;\; \delta_{ts}\,, \tag{2.16}$$

where

$$\bar{R}_t = I - HP_t(t,t)H'\,, \tag{2.17}$$

but we shall postpone the discussion of this until Section 5, for the moment defining R_t and $\bar{R}_t$ by (2.14) and (2.17) respectively.

We shall be interested in the *gain functions*

$$Q_t = P_{t-1}(t,t)H' \tag{2.18}$$

$$K_t = P_t(t,t)H' \tag{2.19}$$

in terms of which we can write

$$R_t = I + HQ_t \tag{2.20}$$

$$\overline{R}_t = I - HK_t . \tag{2.21}$$

Then defining the *feedback function*

$$\widetilde{F}_t = I - K_t H, \tag{2.22}$$

we can list the following useful relations:

Proposition 2.2: *The functions defined above are related in the following way:*

$$P_t(t,s) = \widetilde{F}_t P_{t-1}(t,s) \tag{2.23}$$

$$H\widetilde{F}_t = \overline{R}_t H \tag{2.24}$$

$$K_t = Q_t \overline{R}_t \tag{2.25}$$

$$\overline{R}_t = R_t^{-1} \tag{2.26}$$

$$\widetilde{F}_t^{-1} = I + Q_t H \tag{2.27}$$

Proof: Putting $r = t - 1$ in (2.10) we have

$$P_t(t,s) = [I - P_t(t,t)H'H]\; P_{t-1}(t,s)$$

which is precisely (2.23). Relation (2.24) follows immediately from (2.21) and (2.22), and (2.25) is then a consequence of (2.23) and (2.24):

$$\begin{aligned} P_t(t,t)H' &= P_{t-1}(t,t)\, \widetilde{F}_t'\, H' \\ &= P_{t-1}(t,t)H'\overline{R}_t \end{aligned}$$

where we have also used (2.8). To see that (2.26) holds, note that by (2.21) and (2.25)

$$\overline{R}_t = I - HQ_t\overline{R}_t$$

or

$$(I + HQ_t)\,\overline{R}_t = I$$

which together with (2.20) yields the desired result. Finally, to prove (2.27) first observe that $(I + Q_t H)$ is nonsingular, $Q_t H$ being the product of the two nonnegative matrices $P_{t-1}(t,t)$ and $H'H$. Then apply the "matrix inversion lemma" to obtain

$$(I + Q_t H)^{-1} = I - Q_t(I + HQ_t)^{-1}H.$$

Therefore, by successively applying (2.20), (2.26) and (2.25), we have

$$(I + Q_tH)^{-1} = I - K_tH,$$

which in view of (2.22) is the same as (2.27). ∎

The reason for our interest in the functions K_t and $\tilde{F}_t$ will be made clear presently upon applying (2.23) to the representation (2.6) which yields

$$\hat{x}(t|t) = \tilde{F}_t\hat{x}(t|t-1) + K_tz(t) \tag{2.28}$$

This formula, which of course holds without any special assumptions on the x-process, constitutes the *measurement update* of the Kalman filter. (See e.g. [26].) In order to obtain the *time update* part of the filter we need to impose a Markov structure on the covariance functions of x, i.e.

$$\Gamma_{t+1,s} = F\Gamma_{ts} \quad \text{for } t > s, \tag{2.29}$$

where F is an n x n matrix. Indeed, by applying this condition to (2.9) we have

$$P_r(t+1,s) = FP_r(t,s) \tag{2.30}$$

whenever $t > s,r$, which together with (2.6) gives the time update formula

$$\hat{x}(t+1|t) = F\hat{x}(t|t). \tag{2.31}$$

Hence we can combine (2.28) and (2.31) to obtain

$$\hat{x}(t+1|t) = F\tilde{F}_t\hat{x}(t|t-1) + FK_tz(t) \tag{2.32}$$

which is the Kalman filter formula for the one-step predictor. Of course we may instead want the pure filtering formula

$$\hat{x}(t|t) = \tilde{F}_tF\,\hat{x}(t-1|t-1) + K_tz(t). \tag{2.33}$$

In any case, the application of these recursive filtering formulas requires determining the gain function K_t, which is usually done by solving a matrix Riccati equation. In this paper, however, we shall take a different course and, for the case when x is stationary, develop a different set of equations for K_t from basic principles. This will be done in the following sections.

Finally, we should point out that of course the results of this section do not require that the matrices H and F be constant as our notations suggest. However we have deliberately left out the time arguments since the main result of this paper concerns stationary processes.

3. Reversed-time estimation of stationary processes

In the sequel we shall assume that the covariance function (2.1) is given by

$$\Gamma_{ts} = C_{t-s}, \tag{3.1}$$

i.e. the process x is wide sense stationary. Clearly the sequence C_t must satisfy

$$C_{-t} = C_t' \, . \tag{3.2}$$

The error covariance function P is uniquely determined by C, being the unique solution of (2.9) with Γ given by (3.1). To remind ourselves of this fact, we may write P[C] although we shall refrain from this whenever there is no reason for misunderstanding.

Much of what follows will heavily rely on the simple observation that P^* defined by

$$P_r^*(t,s) = P_r(r-t, r-s) \tag{3.3}$$

is also an error covariance function of type (2.5) and that consequently we can define "starred" versions of the quantities defined in Section 2 by merely exchanging P for P^* everywhere. In fact, we have

Proposition 3.1: $$P_r^*[C] = P_r[C'] \tag{3.4}$$

Proof: Insert (3.1) into (2.9), make a simple change in the order of summation, and observe (3.2) to see that P^* is the unique solution of (2.9) with $\Gamma_{ts} = C_{t-s}'$. (This new Γ is clearly a covariance function.) ■

Hence the new functions K^*, Q^*, R^*, $\bar{R}^*$, $\tilde{F}^*$ etc. have precise meanings. We note that the starred version of the important relation (2.10) is unchanged since it does not depend on C, and that the star operation applied twice gives us the original quantity back (for $P^{**} = P$). Also note that the star operation degenerates for m = 1, so that the starred quantities are equal to the unstarred.

As explained in Section 2 we shall be interested in obtaining equations for the gain function K_t. Since (2.25), (2.26) and (2.20) provide us with the relation

$$K_t = Q_t(I + HQ_t)^{-1}, \tag{3.5}$$

equations in Q_t will also serve our purpose. Hence we shall develop several sets of equations in both K_t and Q_t and explain the relation between them.

Our basic tool is the Bellman-Krein type equation (2.10) presented in the previous section, i.e.

$$P_{r+1}(t,s) = P_r(t,s) - P_{r+1}(t,r+1)H'HP_r(r+1,s) \tag{3.6}$$

In order to determine K_t and Q_t, we need $P_t(t,t)$ and $P_{t-1}(t,t)$ respectively. Equ. (3.6) only provides us with a recursion in the first (index) argument, but by introducing the Markov structure (2.29) we can readily derive a recursion updating all three arguments. This will lead to the *Riccati equation*, the non-linear term of which is supplied by (3.6). We shall however proceed in a different direction:

First note that

$$P_{t-1}(t,t) = P_{t-1}^*(-1,-1)$$

and that

$$P_t(t,t) = P_t^*(0,0) \, .$$

Therefore the *starred version* of (3.6) will immediately provide us with a recursion of desired type, for only the index argument need to be updated. Indeed,

$$P_t^*(-1,-1) = P_{t-1}^*(-1,-1) - P_t^*(-1,t)H'HP_{t-1}^*(t,-1) \tag{3.7}$$

and

$$P_{t+1}^*(0,0) = P_t^*(0,0) - P_{t+1}^*(0,t+1)H'HP_t^*(t+1,0) \tag{3.8}$$

Since, in view of (3.3) and (2.8), the last terms of (3.7) and (3.8) can be written as

$$P_t(t+1,0)H'HP_{t-1}(t,-1)'$$

and

$$P_{t+1}(t+1,0)H'HP_t(t,-1)'$$

respectively, this suggests introducing the following *auxiliary* functions formed in analogy with Q_t and K_t,

$$U_t = P_{t-1}(t,-1)H' \tag{3.9}$$

$$V_t = P_t(t+1,0)H' \tag{3.10}$$

$$X_t = P_t(t,-1)H' \tag{3.11}$$

$$Y_t = P_{t+1}(t+1,0)H', \tag{3.12}$$

the physical interpretation of which will be made clear in Section 5 upon introducing the *backward innovation processes.* Therefore, postmultiplying (3.7) and (3.8) by H′, we obtain

$$Q_{t+1} = Q_t - V_tU_t'H' \tag{3.13}$$

$$K_{t+1} = K_t - Y_tX_t'H' \tag{3.14}$$

and, in view of (2.20) and (2.21),

$$R_{t+1} = R_t - HV_tU_t'H' \tag{3.15}$$

$$\bar{R}_{t+1} = \bar{R}_t + HY_tX_t'H'. \tag{3.16}$$

Consequently, it remains to determine the functions U_t, V_t, X_t and Y_t, and to this end we shall first investigate the relation between them:

Lemma 3.2: *The functions* U_t, V_t, X_t *and* Y_t *are related in the following way:*

$$X_t = \tilde{F}_tU_t \tag{3.17}$$

$$Y_t = \tilde{F}_{t+1}V_t \tag{3.18}$$

$$V_t = U_t\bar{R}_t^* \tag{3.19}$$

$$Y_t = X_t\bar{R}_{t+1}^* \;. \tag{3.20}$$

Proof: Relations (3.17) and (3.18) follow directly from (2.23). To show that (3.19) holds, observe that V_t can be written

$$P_t^*(-1,t)H'$$

and therefore we can invoke the starred version of (2.23) and (2.24) together with (2.8) to obtain

$$V_t = P_{t-1}^*(-1,t)\widetilde{F}_t^{*\prime}H'$$

$$= P_{t-1}^*(-1,t)H'\overline{R}_t^*$$

which is equal to the right member of (3.19). In the same way we prove (3.20). ∎

Lemma 3.2 provides us with a means to obtain any one of the quantities U_t, V_t, X_t and Y_t in terms of any of the others by a linear transformation, for both $\widetilde{F}_t$ and $\overline{R}_t^*$ are clearly invertible, $\widetilde{F}_t^{-1}$ being given by (2.27) and $\overline{R}_t^*$ being equal to $(R_t^*)^{-1}$ by † (2.26)*. These transformations can now be used to reformulate (3.13)–(3.16) so that only one auxiliary function is needed. This can be done in several obvious ways and we shall return to this in the next sections. In this context we may also note that (3.17) and (3.18) together with (2.24) give us

$$HX_t = \overline{R}_t HU_t \tag{3.21}$$

$$HY_t = \overline{R}_{t+1} HV_t, \tag{3.22}$$

where we should remember that, by (2.26), $\overline{R}_t$ equals R_t^{-1}. However, while $\widetilde{F}_t$ and R_t together with their inverses can be determined via the recursions developed sofar, R_t^* is as yet an unknown quantity.

To determine R_t^* and $\overline{R}_t^*$ we need starred versions of (3.15) and (3.16). We can obtain these by the simple observation that

$$HU_t^* = (HU_t)' \tag{3.23}$$

$$HV_t^* = (HX_t)' \tag{3.24}$$

$$HX_t^* = (HV_t)' \tag{3.25}$$

$$HY_t^* = (HY_t)', \tag{3.26}$$

which is an immediate consequence of the definitions and (3.3). Therefore, the appropriate modifications of (3.15) and (3.16) yield

$$R_{t+1}^* = R_t^* - X_t'H'HU_t \tag{3.27}$$

$$\overline{R}_{t+1}^* = \overline{R}_t^* + Y_t'H'HV_t \tag{3.28}$$

We should however point out that, in view of (2.14)* and (2.17)*, we have

$$R_t^* = I + HP_{t-1}(-1,-1)H'$$

and

$$\overline{R}_t^* = I - HP_t(0,0)H',$$

† (a)* means "the starred version of (a)"

so that in fact (3.31) and (3.32) can be obtained directly from the *unstarred* Bellman-Krein type equation (3.6), the stationarity assumption being unnecessary in this case. Equations (3.27) and (3.28) can also be reformulated using the transformations of Lemma 3.2, but we shall postpone the discussion of this to the next section.

The above results leave us with the problem to find recursions for the auxiliary functions U_t, V_t, X_t and Y_t. The transformations of Lemma 3.2 only provide *static* relations between these functions in that they relate quantities with the same time index. To obtain *dynamic* relations we introduce the Markov condition (2.29) which in our present (stationary) setup reads

$$C_t = F^t C_0 \quad \text{for } t > 0, \tag{3.29}$$

where F and C_0 are constant n x n - matrices. (For $t > 0$, C_t is defined through (3.2).) Then relation (2.30) holds, and, in view of the definitions (3.9)-(3.12), we have:

$$U_{t+1} = FX_t \tag{3.30}$$

$$V_{t+1} = FY_t, \tag{3.31}$$

which together with (3.17) and (3.18) provide us with recursions for the auxiliary functions. We shall return to this in the next section.

However, let us first summarize the results of this section in the following theorem:

Theorem 3.3: *Assume that x is wide sense stationary. Then the gain functions* K_t *and* Q_t *defined by* (2.18) *and* (2.19) *and related through* (3.5), *satisfy recursion* (3.13) *and* (3.14) *respectively. The auxiliary functions* U_t, V_t, X_t *and* Y_t *are related by the linear* static *transformations* (3.17) - (3.22), *where* $\tilde{F}_t$ *and* $\tilde{F}_t^{-1}$ *are given by* (2.22) *and* (2.27), R_t^* *and its inverse* $\bar{R}_t^*$ *satisfy the recursions* (3.27) *and* (3.28), *and* R_t *and its inverse* $\bar{R}_t$ *can be determined either from* (2.20) *and* (2.21) *or by the recursions* (3.15) *and* (3.16). *Also given the Markov condition* (3.29), *we have the* dynamic *relations* (3.30) *and* (3.31) *for the auxiliary functions, which can then also be determined recursively.*

4. Algorithms for the Kalman-Bucy gain

Equipped with the results of the previous section we are now in a position to formulate algorithms for the gain function K_t of the optimal filters (2.32) and (2.33). Of course we assume that x is (wide sense) stationary with covariance function (3.29). Since K_t can be determined from Q_t by means of (3.5) we shall be interested in equations in Q_t also.

The form of the algorithm will primarily depend on our choice of auxiliary functions. So, for example, using U_t we have the following algorithm for Q_t:

$$Q_{t+1} = Q_t - U_t \bar{R}_t^* U_t' H' \tag{4.1}$$

$$U_{t+1} = F\tilde{F}_t U_t \tag{4.2}$$

$$\bar{R}_{t+1}^* = \bar{R}_t^* + \bar{R}_t^* U_t' H' \bar{R}_{t+1}^{-1} H U_t \bar{R}_t^*, \tag{4.3}$$

where $\tilde{F}_t$ and R_t stand for

$$\tilde{F}_t = I - Q_t(I + HQ_t)^{-1} \tag{4.4}$$

and

$$R_t = I + HQ_t. \tag{4.5}$$

To obtain (4.1) and (4.2) we have substituted (3.19) into (3.13) and (3.17) into (3.30) respectively. To see that (4.3) holds, first insert (3.22) into (3.28) which gives us

$$\bar{R}^*_{t+1} = \bar{R}^*_t + V'_t H' \bar{R}^{-1}_{t+1} HV_t. \tag{4.6}$$

Then use transformation (3.19) to obtain (4.3). Relations (4.4) and (4.5) follow immediately from (2.22), (3.5) and (2.20). Instead of (4.3) we could use

$$R^*_{t+1} = R^*_t - U'_t H' R_t^{-1} HU_t, \tag{4.7}$$

obtained by plugging (3.21) into (3.27), but then we would have to invert R^*_t in each step to get the inverse $\bar{R}^*_t$. (In fact, (4.3) can be determined from (4.7) by using the matrix inversion lemma.) Since $P_{-1}(0,0)$ and $P_{-1}(-1,-1)$ both equal C_0 we have the following initial conditions:

$$Q_0 = C_0 H' \tag{4.8}$$

$$U_0 = FC_0 H' \tag{4.9}$$

$$R^*_0 = I + HC_0 H', \tag{4.10}$$

$\bar{R}^*_0$ being simply the inverse of R^*_0. In determining (4.9) we have also used (2.30).

Now, with (4.4) and (4.5) properly inserted, (4.1)–(4.3) provide us with $2mn + \frac{1}{2}m(m+1)$ scalar equations to determine Q_t, for Q_t and U_t are $m \times n$ - matrices and $\bar{R}^*_t$ is a symmetric $m \times m$ - matrix. When $m = 1$ the star operation (3.3) degenerates so that R^*_t is given by (4.5). Then the $\bar{R}^*_t$ - equation (4.3) becomes superfluous and only $2n$ scalar equations are needed. This should be compared with the $\frac{1}{2}n(n+1)$ scalar equations of the Riccati equation ; a much larger number whenever, as often is the case, $m \ll n$.

Other versions of the above algorithm can now be constructed by instead using a different auxiliary function. From (3.13), (3.19), (3.31) and (3.18) we have

$$Q_{t+1} = Q_t - V_t R^*_t V'_t H' \tag{4.11}$$

$$V_{t+1} = F\tilde{F}_{t+1} V_t \tag{4.12}$$

$$R^*_{t+1} = R^*_t - R^*_t V'_t H' R_t^{-1} HV_t R^*_t, \tag{4.13}$$

the last equation of which can be obtained by applying (3.19) to (4.7). Again we have used the short-hand notation (4.4) and (4.5), and the initial condition V_0 is provided by (3.19)

$$V_0 = FC_0 H'(I + HC_0 H)^{-1}. \tag{4.14}$$

We could use (4.6) instead of (4.13) but then again we would have to take the inverse in each step.

Yet another version is provided by

$$Q_{t+1} = Q_t - FX_{t-1}(R_t^*)^{-1}X_{t-1}'F'H' \tag{4.15}$$

$$X_t = \widetilde{F}_t FX_{t-1} \tag{4.16}$$

$$R_{t+1}^* = R_t^* - X_t'H'HFX_{t-1}. \tag{4.17}$$

These equations are simply obtained by applying the transformation (3.30) to (4.1), (3.17) and (3.27) respectively. Equations (3.17) and (4.9) give us the initial condition

$$X_0 = \widetilde{F}_0 FC_0H'. \tag{4.18}$$

Note that the essential difference between this algorithm and (4.1) - (4.3) is the equation for R_t^*, and we can get still other versions by instead using the symmetric form

$$R_{t+1}^* = R_t^* - X_t'H'R_tHX_t \tag{4.19}$$

derived from (3.21) and (3.27), or the inverse recursion

$$\overline{R}_{t+1}^* = \overline{R}_t^* + \overline{R}_t^*X_t'H'R_tR_{t+1}^{-1}R_tHX_t\overline{R}_t^*, \tag{4.20}$$

obtained by applying transformation (3.21) to (4.3).

Proceeding exactly analogously we also have

$$Q_{t+1} = Q_t - FY_{t-1}(\overline{R}_t^*)^{-1}Y_{t-1}'F'H' \tag{4.21}$$

$$Y_t = \widetilde{F}_{t+1} FY_{t-1} \tag{4.22}$$

$$\overline{R}_{t+1}^* = \overline{R}_t^* + Y_t'H'HFY_{t-1}, \tag{4.23}$$

with the alternative of using

$$\overline{R}_{t+1}^* = \overline{R}_t^* + Y_t'H'R_{t+1}HY_t \tag{4.24}$$

or

$$R_{t+1}^* = R_t^* - R_t^*Y_t'H'R_{t+1}R_t^{-1}R_{t+1}HY_tR_t^* \tag{4.25}$$

in place of (4.23). The initial condition of (4.17) is

$$Y_0 = \widetilde{F}_1 FC_0H'(I + HC_0H')^{-1}. \tag{4.26}$$

We can also formulate algorithms directly in terms of the gain K_t. We have, for example,

$$K_{t+1} = K_t - X_t(R_{t+1}^*)^{-1}X_t'H' \tag{4.27}$$

$$X_{t+1} = \widetilde{F}_{t+1} FX_t \tag{4.28}$$

$$R_{t+1}^* = R_t^* - X_t'H'(\overline{R}_t)^{-1}HX_t, \tag{4.29}$$

into which we should substitute $\tilde{F}_t$ and $\bar{R}_t$ as originally defined by (2.22) and (2.21):

$$\tilde{F}_t = I - K_t H \tag{4.30}$$

$$\bar{R}_t = I - HK_t. \tag{4.31}$$

Equation (4.27) is immediately obtained from (3.14) and (3.20), and (4.28) and (4.29) are identical to (4.16) and (4.19) respectively. Instead of (4.29) we could use (4.17), thereby avoiding one inversion. Equ. (4.20) on the other hand could *not* be used, since we need K_{t+1} to determine R_{t+1}^* which in turn is needed to determine K_{t+1}. The initial condition

$$K_0 = C_0 H'(I + HC_0H')^{-1} \tag{4.32}$$

is obtained from (3.5) and (4.8).

Similarily, we can also express K_t in terms of Y_t:

$$K_{t+1} = K_t - Y_t\, R_{t+1}^* Y_t' H', \tag{4.33}$$

but this equation is unfortunately unusable since the corresponding auxiliary equation (4.22) requires knowledge about K_{t+1}.

We should point that from a computional point of view a separate equation for R_t or $\bar{R}_t$ is usually to be preferred. However, in view of (4.5) and (4.31), such a recursion is immediately obtained by premultiplying the appropriate Q_t or K_t equation by H.

The various sets of equations presented above are of course essentially different versions the same algorithm first presented in [7], where we gave the first version (4.1)-(4.3). In the subsequent paper [6] by Kailath et. al., both (4.1)-(4.3) and (4.11) – (4.13) were obtained by two different decompositions of the Riccati equation, but the relation between them was not clearly explained. The other versions, among which the ones in K_t are of particular interest, seem to appear here for the first time.

The purpose of this section has been to demonstrate that the continuous-time algorithm [4,5,8], to which we shall briefly return in Section 6, have *many* counterparts in discrete time (*some* of which we have presented in this section) and to determine the relation between them. We are not at this time prepared to comment on the computional properties of the different versions.

5. Forward and backward innovations

We shall now give a stochastic interpretation of the results presented above. To this end let us first summarize some useful facts about the innovation processes (2.12) and (2.15):

Proposition 5.1: *The processes ν and μ, defined by* (2.12) *and* (2.15) *respectively, satisfy* (2.13), (2.16) *and*

$$E\{\mu(t)\, \nu(s)'\} = I\, \delta_{ts}. \tag{5.1}$$

Moreover we have the innovations representations

$$\hat{x}(t|r) = \sum_{s=0}^{r} P_s(t,s)H'\nu(s) \tag{5.2}$$

and

$$\hat{x}(t|r) = \sum_{s=0}^{r} P_{s-1}(t,s)H'\mu(s), \tag{5.3}$$

where P *is defined by* (2.5).

Proof: The proof of (2.13) is straight-forward using orthogonality arguments and we refer the reader to [25] for it. We can prove (2.16) and (5.1) along the same lines, but instead we shall use relation (2.28) to write

$$\mu(t) = z(t) - H\tilde{F}_t\hat{x}(t|t-1) - HK_t\, z(t).$$

Therefore we can invoke (2.24) and (2.21) to obtain

$$\mu(t) = \bar{R}_t\nu(t), \tag{5.4}$$

which, in view of (2.13) and (2.26), gives us (2.16) and (5.1). It is easy to see [25] that z(t) can be expressed as a linear function of $\nu(0), \nu(1), \ldots, \nu(t)$ so that

$$\hat{x}(t|r) = \sum_{i=0}^{r} \Lambda_i\, \nu(i)$$

for some weighting function Λ. Since

$$E\{\tilde{x}(t|r)\mu(s)'\} = 0 \qquad (s = 0, 1, \ldots, r),$$

using (5.1), we have

$$E\{x(t)\, \tilde{x}(s|s)'\}H' - \Lambda_s = 0,$$

and therefore (5.2) holds. To prove (5.3) we proceed analogously, first noting that, in view of (5.4), z(t) is also a linear function of $\mu(0), \mu(1), \ldots, \mu(t)$. ∎

Remark 5:2: We are now in a position to give an alternative (innovations) proof of the Bellman-Krein type equation (2.10). In fact, using representation (5.2) for $\hat{x}(t|r)$ and (5.3) for $\hat{x}(s|r)$, we have

$$E\{\hat{x}(t|r)\, \hat{x}(s|r)'\} = \sum_{i=0}^{r} P_i(t,i)H'HP_{i-1}(i,s) \tag{5.5}$$

where we have also used (5.1) and (2.8). Then (2.10) is obtained by inserting (5.5) into

$$P_r(t,s) = E\{x(t)x(s)'\} - E\{\hat{x}(t|r)\, \hat{x}(s|r)'\} \tag{5.6}$$

and making the appropriate reformulation. Likewise by using *one* of the representations (5.2) and (5.3) for *both* $\hat{x}(t|r)$ and $\hat{x}(s|r)$, we can also derive two symmetric versions of (2.10), both of which can also be obtained directly from (2.10) by using (2.23) and (2.24). ∎

We shall now proceed to the main result of this section: Consider the string of data

$$\{z(0), z(1), z(2), \ldots, z(T)\} \tag{5.7}$$

where z, of course, is defined by (2.2). The innovation approach amounts to orthogonalizing the data (5.7) and to express the estimate in terms of the so constructed innovation process. This orthogonalization can be performed in two obvious ways: either start with z(0) and proceed in the forward direction up to z(T) or begin with z(T) and go backwards. The former procedure will provide us with a (forward) innovation process such as (2.12) or (2.15), while the latter will give us a *backward innovation process.*

To formalize this idea we define the processes $x_T(t)$ and $w_T(t)$ to be equal to $x(T-t)$ and $w(T-t)$ respectively. Then, defining z_T analogously, we have the following counterpart of (2.2)

$$z_T(t) = Hx_T(t) + w_T(t). \tag{5.8}$$

Hence we shall consider the linear least squares estimate $\hat{x}_T(t|r)$ of $x_T(t)$ given the data

$$\{z_T(0), z_T(1), z_T(2), \dots, z_T(r)\} \tag{5.9}$$

and the corresponding estimation error

$$\tilde{x}_T(t|r) = x_T(t) - \hat{x}_T(t|r). \tag{5.10}$$

We can now define the backward innovation processes

$$\nu_T(t) = z_T(t) - H\hat{x}_T(t|t-1) \tag{5.11}$$

and

$$\mu_T(t) = z_T(t) - H\hat{x}_T(t|t), \tag{5.12}$$

for which we have the following result:

Theorem 5.3: *Assume that* x *is wide sense stationary. Then, for all* T *, the error covariance*

$$E\{\tilde{x}_T(t|r)\tilde{x}_T(s|r)'\} = P_r^*(t,s), \tag{5.13}$$

where P^* *is defined by* (3.3) *and* (2.5). *The innovation processes* (5.11) *and* (5.12) *satisfy*

$$E\{\nu_T(t)\,\nu_T(s)'\} = R_t^*\,\delta_{ts}, \tag{5.14}$$

$$E\{\mu_T(t)\,\mu_T(s)'\} = \bar{R}_t^*\,\delta_{ts}, \tag{5.15}$$

and

$$E\{\mu_T(t)\,\nu_T(s)'\} = I\,\delta_{ts}, \tag{5.16}$$

where R_t^* *and its inverse* $\bar{R}_t^*$ *are defined as in Section* 3. *Moreover, we have the following representations for* $\hat{x}(t|r)$:

$$\hat{x}(t|r) = \sum_{s=0}^{r} P_s(s+t-r,0)\,H'\nu_r(s) \tag{5.17}$$

and

$$\hat{x}(t|r) = \sum_{s=0}^{r} P_{s-1}(s+t-r-1,-1)H'\mu_r(s), \tag{5.18}$$

where P *is defined by* (2.5).

Proof: Since w_T has the same covariance function (2.3) as w, the only difference between the problem to determine $\hat{x}_T(t|r)$ and the estimation problem considered earlier in this paper is that we now have x_T where we previously had x. However, given that x has the covariance function (3.1), we can use (3.2) to see that

$$E\{x_T(t)x_T(s)'\} = C'_{t-s},$$

which does not depend on T. Therefore (5.13) immediately follows from Proposition 3.1. Consequently, (5.14), (5.15) and (5.16) are merely starred versions of (2.13), (2.16) and (5.1). Also,

$$\hat{x}_r(t|r) = \sum_{s=0}^{r} P_s^*(t,s)H'\nu_r(s) \qquad (5.19)$$

is the starred version of (5.2) taking T to be r. Since the data (5.7) and (5.9) coincide for T = r, we have

$$\hat{x}(t|r) = \hat{x}_r(r-t|r), \qquad (5.20)$$

and therefore, in view of (3.3), (5.19) is the same as (5.17). The proof of (5.17) is analogous. ■

Remark 5.4: If we wish to make the above proof independent of any result in Sections 2 and 3, we may simply note that the stationarity of x implies that the left member of (5.13) is independent of T so that we can put T = r. Then we can use (5.20) to see that this error covariance is $P_r(r-t, r-s)$, which is precisely equal to the right member of (5.13). ■

However, our prime interest is in the one-step predictor and the pure filter. Therefore we shall now invoke the representations (5.17) and (5.18) to obtain

$$\hat{x}(t+1|t) = \sum_{s=0}^{t} U_s\,\mu_t(s) \qquad (5.21)$$

$$= \sum_{s=0}^{t} V_s\nu_t(s) \qquad (5.22)$$

and

$$\hat{x}(t|t) = \sum_{s=0}^{t} X_{s-1}\mu_t(s) \qquad (5.23)$$

$$= \sum_{s=0}^{t} Y_{s-1}\nu_t(s), \qquad (5.24)$$

where U,V,X and Y are defined by (3.9) - (3.12), hence providing the previously mentioned stochastic interpretation of the auxiliary functions introduced in Section 3. We can now use these representations to derive the equations for Q_t and K_t presented in Sections 3 and 4. In fact, inserting (5.21) and (5.22) in appropriate combinations into

$$Q_t = C_0H' - E\{\hat{x}(t|t-1)\,\hat{x}(t|t-1)'\}H'$$

(obtained from (5.6)), and observing (5.14), (5.15) and (5.16), we obtain (4.1), (4.11) and (3.13) respectively. Likewise we derive equations (4.27), (4.33) and (3.14) by substituting (5.23) and (5.24) into

$$K_t = C_0H' - E\{\hat{x}(t|t)\,\hat{x}(t|t)'\}H'.$$

To derive equations for R_t^* and its inverse $\bar{R}_t^*$, we shall need *forward* innovation representations for $\hat{x}_t(t+1|t)$ and $\hat{x}_t(t|t)$. Therefore, invoking (5.2), (5.3), (2.8) and (5.20), we have

$$H\hat{x}_t(t+1|t) = \sum_{s=0}^{t} U_s'H'\mu(s) \tag{5.25}$$

$$= \sum_{s=0}^{t} X_s'H'\nu(s) \tag{5.26}$$

and

$$H\hat{x}_t(t|t) = \sum_{s=0}^{t} Y_{s-1}'H'\nu(s) \tag{5.27}$$

$$= \sum_{s=0}^{t} V_{s-1}'H'\mu(s), \tag{5.28}$$

which plugged into

$$R_t^* = I + HE\{\hat{x}_{t-1}(t|t-1)\hat{x}_{t-1}(t|t-1)'\}H' \tag{5.29}$$

and

$$\bar{R}_t^* = I - HE\{\hat{x}_t(t|t)\hat{x}_t(t|t)'\}H' \tag{5.30}$$

in different combinations, yield (3.27), (3.28), (4.6), (4.7), (4.19) and (4.24). Here (5.29) and (5.30) have been obtained by inserting (5.13) with appropriate choice of T and r into the starred versions of (2.14) and (2.17) respectively.

In this section we have so far made no use of the Markov condition (3.29), the need of which enters upon deriving the auxiliary equations (4.2), (4.12), (4.16) and (4.22). In our present stochastic setting the most suitable way to introduce this condition is to assume (as usual) that x is generated by the stochastic difference equation

$$x(t+1) = Fx(t) + v(t)$$

where v is a white sequence. It is then well-known that $\hat{x}(t+1|t)$ is given by (2.32) and consequently

$$\tilde{x}(t+1|t) = F\tilde{F}_t\tilde{x}(t|t-1) + v(t) - FK_tw(t)$$

which inserted into

$$U_t = E\{\tilde{x}(t|t-1)\,x(-1)'\}H'$$

gives us (4.2) (for $x(-1)$ is uncorrelated with v and w). Similar arguments give us the other auxiliary equations (4.12), (4.16) and (4.22). However for X_t and Y_t we must use the pure filtering formula (2.33) instead.

The main purpose of this section has been to provide a stochastic framework for the previous, essentially nonrandom, development. However, despite frequent references to results in Sections 2–4, the presentation has been essentially self-contained.

6. A remark on the continuous-time result

For the sake of comparison we shall *briefly* outline the continuous-time analog of the development in the previous sections. The forward-backward innovation method has already been described in continuous time in [8], so we shall only have to consider the analog of the method of Sections 3 and 4.

The two gain functions K and Q defined in Section 2 have only one counterpart in continuous time, namely

$$K(t) = P_t(t,t)H'. \tag{6.1}$$

Likewise, there is only one auxiliary function

$$Y(t) = P_t(t,0)H', \tag{6.2}$$

corresponding to the functions U,V,X and Y defined by (3.9)-(3.12).

We have the following *Bellman-Krein* equation:

$$\frac{\partial P_r}{\partial r}(t,s) = -P_r(t,r)H'HP_r(r,s) \tag{6.3}$$

to replace (3.10), the starred version of which gives us

$$\frac{\partial P_t^*}{\partial t}(0,0) = -P_t^*(0,t)H'HP_t^*(t,0),$$

which, in view of (3.3) and "(2.8)", provides us with the counterpart of (3.13) and (3.14), namely

$$\dot{K}(t) = -Y(t)\,Y(t)'H'. \tag{6.4}$$

To determine an equation for Y, first note that (6.3) implies

$$\left[\frac{\partial P_t}{\partial t}(s,0)H'\right]_{s=t} = -K(t)HY(t). \tag{6.5}$$

Then, to proceed we must introduce the counterpart of (2.30): the Markov condition

$$\frac{\partial P_r}{\partial t}(t,s) = FP_r(t,s) \quad \text{for } t > r,s, \tag{6.6}$$

which with a simple limit argument gives us

$$\left[\frac{\partial P_t}{\partial s}(s,0)H'\right]_{s=t} = FY(t) \tag{6.7}$$

Hence (6.5) and (6.7) provide us with the counterpart of (4.2), (4.12), (4.16) and (4.22), namely

$$\dot{Y}(t) = [F - K(t)H]\,Y(t)\;. \tag{6.8}$$

Equations (6.4) and (6.8) together with their initial conditions $K(0) = Y(0) = C_0H'$ constitute the continuous-time analog of *all* algorithms of Section 4. These matrix differential equations, which contain 2n scalar equations, were first obtained (independently of our own work [7]) by Kailath [4] who derived them from the Riccati equation.

References

1. J. Casti, R. Kalaba and V.K. Murthy, A new initial value method for on–line filtering and estimation, IEEE Trans. on Information Theory, July 1972, pp. 515–517.

2. J. Rissanen, A fast algorithm for optimum linear prediction, IEEE Transactions Automatic Control, vol. AC–18, p.555, Oct. 1973.

3. J. Casti and E. Tse, Optimal linear filtering theory and radiative transfer: Comparisons and inter-connections, J. Math. Anal. and Appl., vol. 40, October 1972, pp. 45–54.

4. T. Kailath, Some Chandrasekhar–type algorithms for quadratic regulators, Proc. IEEE Decision and Control Conference, Dec. 1972.

5. T. Kailath, Some new algorithms for recursive estimation in constant linear systems, IEEE Trans. Information Theory, Vol. IT–19, November 1973, pp. 750–760.

6. T. Kailath, M. Morf and S. Sidhu, Some new algorithms for recursive estimation in constant discrete–time linear systems, Proc. Seventh Princeton Symp. on Information and System Science, March 1973.

7. A. Lindquist, A new algorithm for optimal filtering of discrete–time stationary processes, SIAM J. Control (to appear).

8. A Lindquist, Optimal filtering of continuous–time stationary processes by means of the backward innovation process, SIAM J. Control (to appear).

9. A. Lindquist, On Fredholm Integral Equations, Toeplitz Equations and Kalman–Bucy Filtering, International J. Appl. Math. and Optimization (to appear).

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

11. P. Whittle, On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix, Biometrica, vol. 50, 1963, pp. 129–134.

12. R.A. Wiggins and E.A. Robinson, Recursive solutions to the multichannel filtering problem, J. Geophys. Research 70, pp. 1885–1891 (1965).

13. L. Ya. Geronimus, Orthogonal Polynomials, Consultant Bureau, New York, 1961.

14. N.I. Akiezer, The Classical Moment Problem, Hafner Publishing Company, New York 1965.

15. U. Grenander and G. Szego, Toeplitz Forms and their Applications, University of California Press, Berkeley and Los Angeles, 1958.

16. S. Chandrasekhar, Radiative Transfer, Oxford University Press, 1950.

17. R. Bellman, Functional equations in the theory of dynamic programming–VII: A partial differential equation for the Fredholm resolvent, Proc. Amer. Math. Soc., 8, 1957, pp. 435–440.

18. M. Krein, On a new method for solving linear integral equations of the first and second kinds, Dokl. Akad. Nauk.SSSR, 100, 1955, pp. 413–416.

19. M. Krein, Continuous analogs of propositions concerning polynomials orthogonal on the unit circle, Dokl. Akad. Nauk. SSSR, 105, 1955, pp.637–640.

20. R.S. Bucy and P.D. Joseph, Filtering for Stochastic Processes with Applications to Guidance, Inter-science Publishers, New York 1968.

21. D.Q. Mayne, A solution of the smoothing problem for linear dynamic systems, Automatica 4, 1966, pp. 73–92.

22. D.C. Fraser, On the application of optimal linear smoothing techniques to linear and nonlinear dynamic systems, PhD Thesis, M.I.T., Cambridge, Mass., 1967.

23. D.C. Fraser and J.E. Potter, The optimum linear smoother as a combination of two optimum linear filters, IEEE Trans. AC, August 1969.

24. R.K. Mehra, On optimal and suboptimal linear smoothing, Proc. Nat. Electr. Conf. XXIV, 1968.

25. T. Kailath, An innovations approach to linear least squares estimation, Part I: Linear filtering in additive white noise, IEEE Trans. AC–13, 1968, pp. 646–655.

26. P.G. Kaminski, A.E. Bryson and S.F. Schmidt, Discrete square root filtering: A survey of current techniques, IEEE Trans. AC–16, 1971, pp. 727–736.