On Fredholm Integral Equations, Toeplitz Equations and Kalman–Bucy Filtering*

ANDERS LINDQUIST

Department of Mathematics
University of Kentucky
Lexington, Kentucky 40506

Communicated by W. Fleming

ABSTRACT

In this paper we consider some new algorithms for computing the Kalman–Bucy gain in stationary systems requiring a number of equations of order n (rather than n^2) whenever the order n of the system is much larger than the dimension of the output. These equations were independently obtained by Kailath and Lindquist in continuous and discrete time respectively. We briefly discuss the relations with some recent related results due to Casti, Kalaba & Murthy and Rissanen. Some of the reasons for these reductions are inherent in the properties of general stationary processes, and therefore a considerable portion of the paper is devoted to exploring the connections with some previous work by Levinson, Whittle and Wiggins & Robinson, and also with the Szegö theory of polynomials orthogonal on the unit circle and some continuous analogs due to Krein. We demonstrate that the Bellman–Krein formula is the fundamental relation in continuous time, the trick being to introduce a "reversed time" counterpart of the weighting function (Fredholm resolvent). This is suggested by the "forward and backward innovation" approach in a previous paper by the author, the essential relations of which we reformulate in terms of Fredholm integral equations (in continuous time) and Toeplitz equations (in discrete time). Therefore we also derive the discrete-time Bellman–Krein formulas of which there are actually two—one corresponding to the one-step predictor and one to the pure filter. In this way we shall be able to pin down the reasons for the striking discrepancies between the continuous-time and the discrete-time cases. Finally we clarify the relations between Levinson's equations and Chandrasekhar's X- and Y-functions.

1. Introduction. In applying Kalman–Bucy filtering to practical problems one often encounters the situation for which the order n of the system is much larger than the dimension m of the observed process. Although strictly speaking one only needs mn scalar functions—namely the components of the so-called gain-matrix—to determine the filter, the classical approach requires the solution of an $n \times n$-matrix Riccati equation, i.e., due to symmetry, $\frac{1}{2}n(n+1)$ scalar equation. However, recently a different approach has been developed which for a stationary

* This research was carried out while the author held a visiting position at the Division of Applied Mathematics, Brown University, Providence, Rhode Island, and was supported by the National Science Foundation under grant NSF–46614.

APPLIED MATHEMATICS & OPTIMIZATION, Vol. 1, No. 4

system replaces the Riccati equation by a set of $2mn$ scalar equations in the continuous-time case or $2mn+\frac{1}{2}m(m+1)$ in the discrete-time case. (Actually for $m = 1$ only $2n$ equations are really required.) These equations were independently derived by Kailath [9] in continuous time and Lindquist [15] in discrete time. However the problem appears to have a longer history which we shall briefly review and in which particularly the work of Casti, Kalaba and Murthy [4], Casti and Tse [6], and Rissanen [19] play an important part.

Let us first point out that (unlike the Riccati equation) the equations mentioned above only hold for (wide sense) stationary systems. By this we mean that the system process is stationary, while of course the filtering estimate is nonstationary, the estimation interval being finite and increasing. As demonstrated by Kailath a *similar* reduction in the number of equations required can be obtained for certain special types of nonstationary processes. Although these equations usually are different in number, Kailath's method [9] to a certain extent provides a unified framework for these equations and the stationary ones. However in this paper we shall not discuss these nonstationary cases since they do not naturally fit into our treatment.

As already mentioned, the continuous-time version of the new equations were first obtained by Kailath [9] who derived them from the Riccati equation. (Also see [10].) Basically Kailath's contribution consists in the simple but important observation that a certain $n \times n$-matrix in a formula due to Bucy [3] for the derivative $\dot{\Sigma}$ of the Riccati equation solution has at most rank m (which in the sequel we shall always assume to be much less than n). Then $\dot{\Sigma}$ has also (at most) rank m and can therefore be decomposed to yield the desired equations. However it has been demonstrated by Lindquist [15] that this low rank property basically has nothing to do with the Riccati equation since in fact it holds for all wide sense stationary systems and not only those realized by a finite dimensional system. It should however be pointed out that the above mentioned $n \times n$-matrix in Bucy's formula can have low rank also in certain nonstationary systems and therefore, as pointed out above, Kailath's method also works for these situations.

The discrete-time equations were first derived (independently of Kailath's work on the continuous-time case) by Lindquist [15]. This was done without resort to the Riccati equation. Instead we used certain recursions for the solution of Toeplitz equations with counterparts in the theory of orthogonal polynomials [1, 8] and previously derived in various degrees of completeness by Levinson [14], Whittle [21] and Robinson and Wiggins [20]. None of these references is however sufficiently complete for our purposes* and we gave our own proof in [15] which suggested the method of "forward and backward innovations", further streamlined in our paper [16]. (Some rudiments of this approach were at least implicitly already contained in Whittle's paper [21].) The equations of [15] can of course also be derived from the Riccati equation (which was done by Kailath in the subsequent paper [11]). However, since somewhat surprisingly the discrete-time equations are more complicated, in this case Kailath's method loses much of its

* After the submission of [15], a complete treatment [18] due to Rissanen has appeared in print. His proof is however different in spirit.

elegance and unfortunately appears rather ad hoc, the decomposition not being unique. Unlike our method [15] and the discrete-time version of our innovation method [16] the extra equations obtained in the discrete-time case are given no natural interpretation and therefore the Riccati approach gives little physical insight. In passing we may add that at an A.M.S. meeting in Cleveland, Ohio, on 24 November, 1972, R. E. Kalman announced similar results ($2n$ equations for the gain in the scalar output case.) We are as yet not aware of the actual details of these results although we have been told that his method is independent of ours.

There are, however, some similar results in the literature which were actually obtained somewhat earlier. Casti, Kalaba and Murthy [4] consider the class of continuous-time filtering problems for which the (scalar) system process has the covariance function

$$C(t) = \int_0^1 e^{-|t|\lambda}\, \omega(\lambda) d\lambda \tag{1.1}$$

where ω is a suitable weighting function. Then the filter can be determined by solving a certain pair of coupled differential equations for two functions $X(t, \lambda)$ and $Y(t, \lambda)$. Fredholm integral equations with kernels of type (1.1) were studied in other contexts already during the 1940's by Chandrasekhar [7], and in fact the two functions mentioned above are usually called Chandrasekhar's X- and Y-functions. Now by choosing ω to be a weighted sum of n δ-functions, (1.1) will be a sum of exponentials and the X- and Y-functions only need to be determined for n different values of λ. This covariance function corresponds to a scalar output Kalman–Bucy process with a simple coefficient matrix (the eigenvectors span the whole space) and indeed the corresponding equations of [9, 10, 16] can with some effort be derived from these results. (The generalization to vector outputs is fairly straightforward.) Certain connections with the Kalman–Bucy filter were discussed in Casti and Tse [6]. We should certainly also mention the work of Rissanen [19], who was probably the first to give a fast discrete-time algorithm [$2(n+1)m^2$ equations] of the type discussed in this paper, although this work concerns ARMA-models rather than Kalman–Bucy models.

The purpose of this paper is to discuss how the above mentioned fast algorithms are related to certain results concerning Fredholm integral equations, Toeplitz equations and orthogonal polynomials. Linear estimation of arbitrary stationary processes leads to Fredholm integral equations in continuous time and Toeplitz equations in discrete time, the above mentioned algorithms being obtained by introducing a special structure on the covariance function. Nevertheless, as pointed out above, certain properties of the new Kalman–Bucy algorithms are actually intrinsic to stationary processes in general and therefore to pinpoint these we shall first discuss the general case.

The Toeplitz equations obtained in discrete-time filtering of *arbitrary* stationary processes can be solved recursively by means of the *Levinson-type equations* [14, 15, 18, 20, 21] also found in the Szegö theory of *orthogonal polynomials* [1, 8]. The continuous-time counterparts of these equations were presented in [16], and they are, as we shall demonstrate in this paper, actually restrictions of the *Bellman–Krein formula* [2, 12] to suitably parametrized planes in the 3-dimensional argument space of the resolvent. Hence, as also the treatment in [5] suggests, it appears that the Bellman–Krein formula is the fundamental tool in obtaining the results mentioned above, and therefore in Section 3 we shall develop its discrete-time counterpart, from which we derive the Levinson-type equations. With the aid of the Bellman–Krein formula we shall also construct the matrix versions of certain continuous analogs of the Szegö orthogonal polynomials introduced by Krein [13]. From these equations Chandrasekhar's X- and Y-functions are obtained as a *special case* by using the covariance function (1.1). Hence we are anxious to point out that Chandrasekhar's equations are not the continuous analogs of Levinson's equation which somewhat vaguely has been suggested in the literature. We shall also show that the relation between the Kalman–Bucy gain and the orthogonal polynomial type equations is similar in discrete and continuous time and therefore the connection to [6] should be clear.

The main emphasis of this paper is on the connections to our own previous work [15, 16], the treatment being in the same spirit giving the pertinent facts of our *forward and backward innovations* approach [16] in terms of Fredholm and Toeplitz equations. We shall also try to clarify (as has been done before) that the extensions of Levinson's [14], Krein's [13] and Geronimus' [8] equations (which were developed for the scalar situation) to the vector case do introduce some non-trivial aspects, the scalar equations being degenerated versions of the vector ones in that the "backward" and "forward" relations are identical. Also by developing the simpler continuous-time case and the more complicated discrete-time case parallelly in Sections 2 and 3 respectively, we shall be able to pin down the reasons for the differences between the two cases.

2. The Continuous-time Case. *2.1* Let $\{y(t); 0 \leq t \leq T\}$ be an m-dimensional wide sense stationary stochastic process with zero mean, and let C be the (bounded) $m \times m$-matrix covariance function

$$C(t-s) = E\{y(t)y(s)'\}. \tag{2.1}$$

Prime denotes transposition and of course we have

$$C(-t) = C(t)'. \tag{2.2}$$

Furthermore, let $\{w(t); 0 \leq t \leq T\}$ be an m-dimensional process with zero mean and orthogonal increments:

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

Also, for simplicity, assume that the two processes y and w are uncorrelated.

Consider the wide sense conditional mean

$$\hat{y}(t|r) = \hat{E}\{y(t) \mid z(s); 0 \leq s \leq r\} \tag{2.4}$$

where the observation process z is defined as

$$z(t) = \int_0^t y(s)ds + w(t). \tag{2.5}$$

Then, denoting the estimation error

$$\tilde{y}(t|r) = y(t) - \hat{y}(t|r) \tag{2.6}$$

and the error covariance function

$$G_r(t, s) = E\{\tilde{y}(t|r)\tilde{y}(s|r)'\}, \tag{2.7}$$

we have the following representation:

Proposition 2.1:
$$\hat{y}(t|r) = \int_0^r G_r(t, s)dz(s). \tag{2.8}$$

Proof: Since (2.4) is the linear least squares estimate it has the form (2.8) (see e.g. [17]), and it only remains to determine G_r. However, the projection theorem implies that

$$E\{\tilde{y}(t|r)z(s)'\} = \int_0^s [E\{\tilde{y}(t|r)y(\tau)'\} - G_r(t, \tau)]d\tau$$

is zero for all $s \in [0, r]$ and therefore

$$G_r(t, s) = E\{\tilde{y}(t|r)y(s)'\} \tag{2.9}$$

from which (2.7) follows. □

Now G_r is actually the resolvent of a Fredholm integral equation and has the following properties:

Proposition 2.2: *The function G defined by (2.7) satisfies*

$$G_r(s, t) = G_r(t, s)', \tag{2.10}$$

and is the unique L_2 solution of each of the following two integral equations:

$$G_r(t, s) + \int_0^r G_r(t, \tau)C(\tau - s)d\tau = C(t-s) \tag{2.11}$$

$$G_r(t, s) + \int_0^r C(t-\tau)G_r(\tau, s)d\tau = C(t-s). \tag{2.12}$$

Moreover, it satisfies the Bellman–Krein formula

$$\frac{\partial G_r}{\partial r}(t, s) = -G_r(t, r)G_r(r, s). \tag{2.13}$$

In fact, (2.10) follows directly from the definition (2.7). Inserting (2.8) into (2.9) yields (2.11). By transposing (2.11) and applying (2.10) and (2.2), we have (2.12). Since C is bounded G_r is bounded too. (For $\tilde{y}$ is a projection of y.) Hence G_r is the unique L_2 solution of (2.11) and (2.12), the integral operators of which are positive. It is easy to derive the Bellman–Krein formula from (2.11), but we refer the reader to [2, 12].

Note that the well-known formula

$$\hat{y}(t|r) = \hat{y}(t|t) + \int_t^r G_s(t, s)[dz(s) - \hat{y}(s|s)ds] \tag{2.14}$$

for the smoothing estimate readily follows from the Bellman–Krein equation. In fact, by formally differentiating (2.8) with respect to r and using (2.13) we have

$$d_r\hat{y}(t|r) = G_r(t, r)[dz(r) - \int_0^r G_r(r, s)dz(s)dr]$$

which yields (2.14). (Also see [22].) Here the differentiation is equivalent to changing the order of integration in a stochastic integral which is permitted since the integrand is bounded. (See [17] where this argument is used repeatedly.)

It is clear that there is one resolvent G for each choice of the kernel C and to remind ourselves of this fact we may write $G[C]$, although we shall refrain from this whenever there is no reason for misunderstanding. Now, define the "backward" resolvent G* as follows:

$$G_r^*(t, s) = G_r(r-t, r-s). \tag{2.15}$$

Then it is immediately seen (by using (2.2) and suitably changing the variable of integration) that G_r^* satisfies (2.11) with C exchanged for C':

Proposition 2.3: $G^*[C] = G[C']$.

Clearly $G^{**} = G$, and in the scalar case ($m = 1$) the star operation degenerates so that $G^* = G$.

2.2 We shall now turn to the filtering problem and following [16] we define the weighting function

$$F(t, s) = G_t(t, t-s) \tag{2.16}$$

in terms of which we can write the *filtering estimate*

$$\hat{y}(t|t) = \int_0^t F(t, t-s)dz(s). \tag{2.17}$$

We also define the starred version of F as

$$F^*(t, s) = G_t^*(t, t-s) \tag{2.18}$$

which by (2.15) equals $G_t(0, s)$, and is therefore the weighting function for the *initial point smoothing estimate*

$$\hat{y}(0|t) = \int_0^t F^*(t, s)dz(s). \tag{2.19}$$

The reversed time definition of F is natural in the context of [16] where $\hat{y}(t|t)$ was expressed in terms of the "backward" observation process. Since by (2.15)

$$F(t, s) = G_t^*(0, s), \tag{2.20}$$

in our present setting this definition will enable us to apply the Krein–Bellman formula to obtain

$$\frac{\partial G_t^*}{\partial t}(0, s) = -G_t^*(0, t)G_t^*(t, s). \tag{2.21}$$

This motivates us to define the function

$$\Gamma(t) = G_t(t, 0) \tag{2.22}$$

which by (2.15) equals $G_t^*(0, t)$, and therefore (2.21) yields

$$\frac{\partial F}{\partial t}(t, s) = -\Gamma(t)F^*(t, t-s). \tag{2.23}$$

In the scalar case ($m = 1$) we can remove the star in (2.23), and we have the continuous-time Levinson equation. In the general case we also need the starred version of (2.23):

$$\frac{\partial F^*}{\partial t}(t, s) = -\Gamma^*(t)F(t, t-s) \tag{2.24}$$

which follows by analogy, defining Γ^* from G^* as in (2.22). However, we are also fortunate to have the following important relation

$$\Gamma^*(t) = \Gamma(t)', \tag{2.25}$$

for by (2.14) $G_t^*(t, 0)$ is equal to $G_t(0, t)$, which by (2.10) equals $G_t(t, 0)'$. Also notice that Γ is the restriction of F to the diagonal:

$$\Gamma(t) = F(t, t), \tag{2.26}$$

an analogous starred statement holding for Γ^*.

Equations (2.23) and (2.24) together with relation (2.25), which were derived in [16] by means of the "forward and backward innovations", constitute the continuous-time version of the (matrix) Levinson-type equations [15, 18, 20, 21]. We have derived these equations for the sake of comparison and our interest in them is mainly theoretical. They show that the filter is completely characterized by the parameter function Γ. Indeed by (2.13) and (2.10) we can express $G_r(t, s)$ in terms of the F-function, so that Γ in fact characterizes any linear estimate (2.8). There is a close relationship between Γ and the filtering error covariance

$$R(t) = G_t(t, t). \tag{2.27}$$

In fact, observing that $R(t) = G_t^*(0, 0)$, (2.21) and (2.25) yield

$$\dot{R}(t) = -\Gamma(t)\Gamma(t)'; \; R(0) = C(0). \tag{2.28}$$

The starred version of this equation is

$$\dot{R}^*(t) = -\Gamma(t)'\Gamma(t); \; R^*(0) = C(0) \tag{2.29}$$

where of course $R^*(t) = G_t^*(t, t) = G_t(0, 0)$ is the error covariance of the initial point smoothing estimate (2.19).

We can now apply these equations to the Kalman–Bucy filter as outlined in the end [16]. However, for the sake of comparison with the discrete-time case and the development in [4], we shall *formally* take a somewhat different course. Let us define the functions Φ and Φ^* as follows:

$$\begin{cases} \Phi(t, z) = I - \int_0^t e^{zs} G_t^*(s, 0)ds & (2.30) \\ \Phi^*(t, z) = e^{zt}\left[I - \int_0^t e^{-zs} G_t(s, 0)ds\right]. & (2.31) \end{cases}$$

The scalar ($m = 1$) versions of these functions (in which case $G^* = G$) were defined in a paper [13] by Krein, who pointed out that in a certain sense the functions $\lambda \to \Phi^*(t, i\lambda)$ are the continuous-time analogs of the Szegö polynomials orthogonal on the unit circle. Note that in the scalar case we have the particularly simple relationship

$$\Phi^*(t, z) = e^{zt} \Phi(t, -z) \tag{2.32}$$

which however does not hold in the general case.

Then, by a straight-forward application of the Bellman–Krein formula and (2.25), we have

$$\begin{cases} \frac{\partial \Phi}{\partial t}(t, z) = -\Phi^*(t, z)\Gamma(t)' & (2.33) \\ \frac{\partial \Phi^*}{\partial t}(t, z) = z\Phi^*(t, z) - \Phi(t, z)\Gamma(t) & (2.34) \end{cases}$$

with initial conditions $\Phi(0, z) = \Phi^*(0, z) = I$. We shall later allow z to be a constant $n \times n$-matrix, so the reader should convince himself that the equations hold for this case also. Note that the functions Φ and Φ^* are completely determined by the parameter function Γ.

2.3 So far we have assumed that C is an arbitrary covariance function. However, if we require that C be a certain type of function, we also introduce a structure on Γ. Indeed, from (2.12) and (2.22), we have

$$\Gamma(t) = C(t) - \int_0^t C(t-s)G_t(s, 0)ds. \tag{2.35}$$

For example with the covariance function (1.1)

$$C(t) = \int_0^1 e^{-|t|\lambda} \omega(\lambda)d\lambda \tag{2.36}$$

we have

$$\Gamma(t) = \int_0^1 \Phi^*(t, -s)\omega(s)ds, \tag{2.37}$$

in which case

$$\begin{cases} X(t, s) = \Phi(t, -s) \\ Y(t, s) = \Phi^*(t, -s) \end{cases}$$

are the Chandrasekhar X- and Y-functions. Casti, Kalaba and Murthy [4] showed that the corresponding filtering estimate can be expressed in terms of these functions.

Now, let us assume that the process y is defined by

$$y(t) = Hx(t) \tag{2.38}$$

where H is a constant $m \times n$-matrix and x an n-dimensional wide sense stationary process described by the stochastic differential equation

$$dx = Axdt + Bdv; \; x(0) = x_0. \tag{2.39}$$

Here x_0 has zero mean and covariance matrix P_0 and v is a zero mean process of type (2.3). Assume that x_0, v and the observation noise w in (2.5) are pairwise uncorrelated. Therefore, since for $t \geq s$

$$x(t) = e^{A(t-s)} x(s) + \int_s^t e^{A(t-\tau)} Bdv(\tau),$$

x being stationary it is not hard to see that

$$E\{x(t)x(s)'\} = e^{A(t-s)}P_0 \tag{2.40}$$

and consequently we have

$$C(t) = He^{At}P_0H'. \tag{2.41}$$

If the matrix A is simple, i.e. its eigenvectors span R^n, and $m = 1$ (a restriction which however can be removed), this covariance function is of type (2.36) with the weighting function being a sum of δ-functions, and then we can use the method of [4]. However with the apparatus developed above we can easily handle the general case, the simple trick being the substitution of the matrix A for z in the functions Φ and Φ^*.

In fact, by plugging (2.41) into (2.35), we have

$$\Gamma(t) = H\, e^{At}\left[P_0H' - \int_0^t e^{-As}\, P_0H'G_t(s, 0)ds\right],$$

which in view of (2.31) suggests the notation

$$\Gamma(t) = H(P_0H'\Phi^*)(t, A). \tag{2.42}$$

Here $(P_0H'\Phi^*)(t, z)$ is the function obtained by premultiplying (2.31) by P_0H' (while formally "z is still a scalar"). Then insert A with the exponentials in the left-most position to obtain $(P_0H'\Phi^*)(t, A)$.

It is well-known that the wide sense conditional mean

$$\hat{x}(t) = \hat{E}\{x(t) \mid z(s);\ 0 \le s \le t\}$$

is generated by the Kalman–Bucy filter

$$d\hat{x} = A\hat{x}dt + K(t)[dz - H\hat{x}dt];\ \hat{x}(0) = 0$$

where the gain-matrix

$$K(t) = \Sigma(t)H' \tag{2.43}$$

is usually determined by solving a matrix Riccati differential equation for the $n \times n$ error covariance matrix

$$\Sigma(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}. \tag{2.44}$$

However, here we shall derive the new equations [9, 10, 16].

Now, by (2.8) and (2.15), we have

$$\hat{y}(t|t) = \int_0^t G_t^*(0, t-s)dz(s)$$

which inserted into

$$K(t) = E\{x(t)[y(t) - \hat{y}(t|t)]'\}$$

(obtained from (2.38), (2.43) and (2.44)) yields

$$K(t) = P_0H' - \int_0^t e^{A(t-s)}\, P_0H'G_t^*(t-s, 0)ds,$$

where (2.38), (2.40) and also (2.10) have been used. Hence, again observing the above mentioned notational convention, (2.30) gives

$$K(t) = (P_0H'\Phi)(t, A). \tag{2.45}$$

Therefore, define K^* to be

$$K^*(t) = (P_0H'\Phi^*)(t, A), \tag{2.46}$$

so that we can exploit the recursions (2.33) and (2.34), properly interpreted, to obtain

$$\begin{cases} \dot{K}(t) = -K^*(t)\Gamma(t)' & (2.47) \\ \dot{K}^*(t) = AK^*(t) - K(t)\Gamma(t). & (2.48) \end{cases}$$

However, from (2.42) and (2.46) we have

$$\Gamma(t) = HK^*(t) \tag{2.49}$$

and hence we have the $2mn$ equations for the gain

$$\begin{cases} \dot{K}(t) = -K^*(t)K^*(t)'H' & (2.50) \\ \dot{K}^*(t) = [A - K(t)H]K^*(t) & (2.51) \end{cases}$$

with initial conditions $K(0) = K^*(0) = P_0H'$. Of course we do not need Φ and Φ^* to derive these equations. However, we feel that there are some conceptual advantages which will become apparent on comparing with the results of Section 3. The reader can easily convince himself that the formal step from (2.33) and (2.34) to (2.47) and (2.48) is indeed correct.

The equations (2.50) and (2.51) were first obtained by Kailath [9] by decomposing the Riccati equation. Our own approach [16] was by means of the "forward and backward innovations".

3. The Discrete-time Case. In order to explain the differences between the discrete-time and the continuous-time estimation problems, in this section we shall develop the discrete-time equations by closely following the procedure in Section 2, and to facilitate a comparison we shall as far as possible use the same notations.

3.1 We consider the problem to determine the wide sense conditional mean

$$\hat{y}(t|r) = \hat{E}\{y(t) \mid z(s);\, s = 0, 1, \ldots, r\} \tag{3.1}$$

where

$$z(t) = y(t) + w(t). \tag{3.2}$$

Here $\{y(t); t = -1, 0, 1, 2, \ldots\}$ is an m-dimensional wide sense stationary process with zero mean and covariance function C defined as in (2.1). Of course C satisfies (2.2). The error process w is a zero mean white noise sequence:

$$E\{w(t)w(s)'\} = I\delta_{ts} \tag{3.3}$$

where δ_{ts} is the Kronecker symbol, and as in Section 2 we assume that y and w are uncorrelated.

For the one-step prediction problem ($r = t-1$), in which we are primarily interested, a nontrivial estimation problem can sometimes be formulated with the identity matrix I in (3.3) exchanged for a singular matrix or even with $w = 0$. (Like in the continuous-time problem, exchanging I for a *non-singular* matrix does not introduce any nontrivial complications.) However, for the moment we wish to retain the analogy with the continuous-time case, and we shall discuss the necessary modifications in the end of this section (part 3.3).

Definition 3.1: With $\tilde{y}$ defined as in (2.6), let

$$G_r(t, s) = E\{\tilde{y}(t|r)\tilde{y}(s|r)'\} \tag{3.4}$$

for $t, s = -1, 0, 1, 2, \ldots$ and $r = 0, 1, 2, \ldots$

Proposition 3.2: $\hat{y}(t|r) = \sum_{s=0}^{r} G_r(t, s)z(s).$ (3.5)

Proof: Since indeed $\hat{y}$ is linear in z, we only need to show that G as defined by (3.5) is given by (3.4). The projection theorem implies that

$$E\{\tilde{y}(t|r)z(s)'\} = E\{\tilde{y}(t|r)y(s)'\} - G_r(t, s)$$

is equal to zero for all $s = 0, 1, \ldots, r$, and our assertion follows. □

Note, that Definition 3.1 defines G_r for arguments ($s = -1$, $s > r$) which are not needed in the representation (3.5). The reason for this will soon become apparent. Proposition 2.2 has the following counterpart in discrete time:

Proposition 3.3: *The function G satisfies*

$$G_r(s, t) = G_r(t, s)' \tag{3.6}$$

and is the unique solution of each of the following systems of equations:

$$G_r(t, s) + \sum_{i=0}^{r} G_r(t, i)C(i-s) = C(t-s) \tag{3.7}$$

$$G_r(t, s) + \sum_{i=0}^{r} C(t-i)G_r(i, s) = C(t-s). \tag{3.8}$$

Moreover, it satisfies the discrete Bellman–Krein equations

$$G_{r+1}(t, s) = G_r(t, s) - G_{r+1}(t, r+1)G_r(r+1, s) \tag{3.9}$$

and

$$G_{r+1}(t, s) = G_r(t, s) - G_r(t, r+1)G_{r+1}(r+1, s). \tag{3.10}$$

Proof: Relation (3.6) follows immediately from the definition. Equation (3.7) is obtained by inserting (3.5) into

$$G_r(t, s) = E\{\tilde{y}(t|r)y(s)'\}$$

and equation (3.8) by transposing (3.7) and applying (3.6) and (2.2). Both (3.7) and (3.8) have unique solutions since the Toeplitz matrix

$$T_{ts} = C(t-s) + I\delta_{ts}$$

is positive definite. For $s = 0, 1, 2, \ldots r$ we can write (3.7) in the following way:

$$\sum_{i=0}^{r} G_r(t, i)T_{is} = C(t-s) \tag{3.11}$$

and for the same values of s we also have

$$\sum_{i=0}^{r} G_{r+1}(t, i)T_{is} + G_{r+1}(t, r+1)C(r+1-s) = C(t-s)$$

which together yields

$$\sum_{i=0}^{r} [G_{r+1}(t, i) - G_r(t, i) + G_{r+1}(t, r+1)G_r(r+1, i)]T_{is} = 0. \tag{3.12}$$

Therefore, since $T_{ij}(i, j = 0, 1, \ldots, r)$ is positive definite, (3.9) holds for $s = 0, 1, 2, \ldots, r$. To see that (3.9) also holds for $s = -1$ and $s > r$, exchange T_{is} for $C(i-s)$ in (3.12). We can do this since the square brackets are all zero. Then apply (3.7) to cancel all sums, which will yield the desired result. Equation (3.10) is obtained by transposing (3.9) and using (3.6). □

The Toeplitz equations (3.7) and (3.8) are the discrete-time counterparts of the Fredholm resolvent equations (2.11) and (2.12), and (3.9) or (3.10) is the discrete Bellman–Krein equation. Notice the lack of symmetry in the quadratic term which causes the discrete-time equations derived below to be more complicated than their continuous-time counterparts.

The well-known smoothing formula

$$\hat{y}(t|r+1) = \hat{y}(t|r) + G_{r+1}(t, r+1)[z(r+1) - \hat{y}(r+1|r)] \tag{3.13}$$

is an immediate consequence of (3.9) and (3.5). In fact, multiply (3.9) by $z(s)$ and sum over $s = 0, 1, \ldots, r$. Similarly, by multiplying (3.10) by $z(s)$ and summing over $s = 0, 1, \ldots, r+1$, we have

$$\hat{y}(t|r+1) = \hat{y}(t|r) + G_r(t, r+1)[z(r+1) - \hat{y}(r+1|r+1)]. \tag{3.14}$$

We can of course reformulate (3.13) and (3.14) to resemble (2.14) which therefore has two counterparts—one for the one-step predictor and one for the pure filter. It is however worth noting that although it is the first Bellman–Krein formula (3.9) which corresponds to the one-step prediction problem, below we shall in fact use the second formula (3.10) in deriving the new equations for this very same problem. (Similarly we could use (3.9) in treating the pure filtering problem but we shall not go into this here.)

This suggests that the reason for the discrepancies between the discrete-time case and the continuous-time case is to be sought in the fact that the continuous time filtering estimate has two "counterparts" in discrete-time, namely $\hat{y}(t|t-1)$ and $\hat{y}(t|t)$. We shall pursue this point by determining the relationship between the weighting functions $G_{t-1}(t, s)$ and $G_t(t, s)$ of the two estimates. Such a relation is readily obtained from the second Bellman–Krein formula and we have the following lemma:

Lemma 3.4: *The weighting functions $G_{t-1}(t, s)$ and $G_t(t, s)$ for the one-step pre-*

dictor and the pure filter are related in the following way:

$$G_{t-1}(t, s) = R_t G_t(t, s) \tag{3.15}$$

or equivalently

$$G_{t-1}(s, t) = G_t(s, t)R_t \tag{3.16}$$

where

$$R_t = G_{t-1}(t, t)+I. \tag{3.17}$$

Proof: Put $r = t-1$ in (3.10) to obtain

$$G_t(t, s) = G_{t-1}(t, s)-G_{t-1}(t, t)G_t(t, s)$$

which is the same as (3.15). Then apply (3.6) to obtain (3.16). □

Since, by Definition 3.1, $G_{t-1}(t, t)$ is the covariance of $\tilde{y}(t|t-1)$, R_t should be interpreted as the covariance matrix of the innovation process

$$\nu(t) = z(t)-\hat{y}(t|t-1) \tag{3.18}$$

which (by definition) is a white noise sequence. Now by multiplying (3.15) by $z(s)$ and summing over $s = 0, 1, \ldots, t$, we have

$$\hat{y}(t|t-1)+G_{t-1}(t, t)z(t) = R_t\hat{y}(t|t), \tag{3.19}$$

which provides us with the following relation

$$\nu(t) = R_t\mu(t) \tag{3.20}$$

where μ is the other innovation process

$$\mu(t) = z(t)-\hat{y}(t|t). \tag{3.21}$$

We may add in passing that (3.20) gives us

$$E\{\mu(t)\mu(s)'\} = R_t^{-1}\delta_{ts} \tag{3.22}$$

and

$$E\{\mu(t)\nu(s)'\} = I\delta_{ts}, \tag{3.23}$$

which shows that only the asymmetric form (3.23) can be normalized as in the continuous-time case. We shall make further use of Lemma 3.4 below after having introduced the matrix versions of Szegö's orthogonal polynomials.

However first, as in Section 2, we define

$$G_r^*(t, s) = G_r(r-t, r-s) \tag{3.24}$$

and it should be clear that Proposition 2.3 holds for the discrete-time case as well, so that all formulas in Proposition 3.3 and Lemma 3.4 are valid also for G^* provided that we exchange C for C' whenever this function occurs. Of course we should also exchange R_t in Lemma 3.4 by R_t^* with the obvious definition. (In the context of [16] this is the covariance of the backward innovation process.) As in Section 2, $G^{**} = G$, and in the scalar case ($m = 1$) we have $G^* = G$.

3.2 Now define the matrix polynomials $\Phi_t(z)$ and $\Phi_t^*(z)$ in analogy with the functions $\Phi(t, z)$ and $\Phi^*(t, z)$ of Section 2:

$$\begin{cases} \Phi_t(z) = I - \sum_{s=1}^{t} z^s G_{t-1}^*(s-1, -1) & (3.25) \\ \Phi_t^*(z) = z^t[I - \sum_{s=1}^{t} z^{-s} G_{t-1}(s-1, -1)]. & (3.26) \end{cases}$$

In the scalar case ($m = 1$) the polynomials $\{\Phi_0^*, \Phi_1^*, \Phi_2^*, \ldots\}$ are the well-known Szegö polynomials orthogonal on the unit circle, as the subsequent development will show, and the unstarred polynomials (3.25) are given by

$$\Phi_t(z) = z^t \Phi_t^* \left(\frac{1}{z}\right). \tag{3.27}$$

However for $m > 1$ the simple relation (3.27) is not true, but we shall shortly see that certain useful recursions due to Geronomus [8] can be generalized to the vector case.

In fact, the discrete Bellman–Krein equation (3.10) yields

$$G_t(s-1, -1) = G_{t-1}(s-1, -1) - G_{t-1}^*(t-s, -1)G_t(t, -1), \tag{3.28}$$

where we have also used (3.24). Likewise we have the starred version of (3.28)

$$G_t^*(s-1, -1) = G_{t-1}^*(s-1, -1) - G_{t-1}(t-s, -1)G_t^*(t, -1). \tag{3.29}$$

Then, if we define

$$\Gamma_t = G_t(t, -1) \tag{3.30}$$

and Γ_t^* analogously in terms of G^*, it is not hard to see that the following recursions hold:

$$\begin{cases} \Phi_{t+1}(z) = \Phi_t(z) - z\Phi_t^*(z)\Gamma_t^* & (3.31) \\ \Phi_{t+1}^*(z) = z\Phi_t^*(z) - \Phi_t(z)\Gamma_t & (3.32) \end{cases}$$

with $\Phi_0(z) = I$ and $\Phi_0^*(z) = I$.

The $m \times m$ matrix coefficients of these polynomials are essentially the weighting matrices for the data $\{z(0), z(1), \ldots, z(t-1)\}$ in linearly estimating one step forward and backward respectively. In fact, if for the moment we write $\Phi_t(z)$ and $\Phi_t^*(z)$ as $\sum \Phi_{ti} z^i$ and $\sum \Phi_{ti}^* z^i$ respectively, (3.5) and (3.24) may be invoked to see that

$$\hat{y}(t|t-1) = -\sum_{i=0}^{t-1} \Phi_{t,\, t-i} z(i) \tag{3.33}$$

and that

$$\hat{y}(-1|t-1) = -\sum_{i=0}^{t-1} \Phi_{t,\, t-i}^* z(i). \tag{3.34}$$

(Note that the error covariance of (3.34) is $G_{t-1}(-1, -1)$, which by (3.24) is the same as $R_t^* - I$.)

Analogously with the continuous-time case there are some parameter sequences Γ and Γ^* which completely characterize the weighting pattern of the one-step predictor (and in fact any other linear estimate). It is then natural to ask whether there is a simple relation such as (2.25) also in the discrete-time case. Unfortunately the answer of this question is "no". Except for the scalar case ($m = 1$) when of course $\Gamma_t^* = \Gamma_t$, in general we have $\Gamma_t^* \neq \Gamma_t'$. In fact,

$$\Gamma_t^* = G_t^*(t, -1) = G_t(0, t+1) = G_t(t+1, 0)',$$

which, compared with (3.30), is seen to be "one-step out of phase". However if instead we define the function

$$S_t = G_{t-1}(t, -1), \tag{3.35}$$

by Lemma 3.4 we have

$$S_t = R_t \Gamma_t, \tag{3.36}$$

obtained by putting $s = -1$ in (3.15). Of course, we also have a starred version of (3.36):

$$S_t^* = R_t^* \Gamma_t^*. \tag{3.37}$$

With this transformation we have established a counterpart of (2.25), for by (3.24) and (3.6), we have

$$S_t^* = S_t', \tag{3.38}$$

so that instead S_t is the parameter sequence characterizing the predictor.

This of course leaves us with the task to determine R_t and R_t^* to which we shall finally proceed. Therefore first note that (3.24) can be used to reformulate (3.17) as follows

$$R_t = G_{t-1}^*(-1, -1) + I.$$

Then again apply the Bellman–Krein formula (3.10) to obtain

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

which is the same as

$$R_{t+1} = R_t - S_t^{*\prime} \Gamma_t^*. \tag{3.39}$$

We can write this in a more symmetric form, for by (3.37),

$$R_{t+1} = R_t - \Gamma_t^{*\prime} R_t^* \Gamma_t^*, \tag{3.40}$$

and of course the corresponding recursion for R_t^* is

$$R_{t+1}^* = R_t^* - \Gamma_t' R_t \Gamma_t. \tag{3.41}$$

The initial conditions are $R_0 = R_0^* = C_0 + I$.

Hence the equations (3.31), (3.32), (3.36), (3.37), (3.38), (3.40) and (3.41) provide us with a means to recursively determine the matrix coefficients of the one-step predictor, that is granted that we know the sequence S_t. However, by (3.8)

$$S_t = C(t+1) - \sum_{i=0}^{t-1} C(t-i) G_{t-1}(i, -1), \tag{3.42}$$

so that in fact to update S_t we must keep a record of all old matrix coefficients. By imposing a suitable structure on C, we can however obtain a recursion for S_t also. This can be achieved by an ARMA-model [19] or by a Kalman–Bucy model, to the latter of which we shall return below (part 3.4). The equations for the general case in (for our purposes) somewhat incomplete versions have been in the literature for some time. For the scalar case ($m = 1$) they can be found in Levinson [14] and in the theory of orthogonal polynomials [1, 8]. In this case of course (3.31) and (3.32) are really equivalent since we have (3.27). In the general case ($m > 1$), the equations can be found in Whittle [21] and Robinson and Wiggins [20]. However, the relation (3.38) is missing in [21], and in [20] there is no proof for it. In our present context the proof of this relation is immediate. Complete treatments can be found in the recent papers by Rissanen [19] and Lindquist [15]. However the motive for our present exposition is not to give another proof for these simple relations, but rather to expose the interconnection with the continuous-time case and the central role played by the discrete-time Bellman–Krein formula (3.10).

3.3 Finally let us point out how the one-step prediction problem as it is usually stated in the literature can be reformulated in the framework of this section: Consider a wide sense stationary vector process $\{u(s);\, s = -1, 0, 1, 2, \ldots\}$ with zero mean and covariance function

$$E\{u(t)u(s)'\} = T(t-s).$$

Then the problem is to determine

$$\hat{u}(t|t-1) = \hat{E}\{u(t) \mid u(s);\, s = 0, 1, \ldots, t-1\}.$$

Furthermore, assume that the process is full rank in the sense that the block Toeplitz matrix $\{T_{ij}\}$ is positive definite. (Here we have defined $T_{ij} = T(i-j)$.) Then $\{T_{ij}\}$ can be decomposed as follows

$$T(t-s) = \tilde{T}(t-s) + P\delta_{ts}$$

where $\{\tilde{T}_{ts}\}$ is positive semidefinite and P positive definite. (E.g., take $P = \epsilon I$ where ϵ is sufficiently small.) Then

$$C(t-s) = P^{-\frac{1}{2}}\tilde{T}(t-s)P^{-\frac{1}{2}'}$$

is the covariance matrix of some process $\{y(t);\, t = -1, 0, 1, 2, \ldots\}$, and therefore the process $z(t) = P^{-\frac{1}{2}}u(t)$ can be written as

$$z(t) = y(t) + w(t)$$

where y and w are defined as in the beginning of this section. Of course

$$\hat{u}(t|t-1) = P^{\frac{1}{2}}\hat{z}(t|t-1).$$

3.4 We shall now impose a special structure on the covariance matrix sequence C of the process y. Assume that y is given by

$$y(t) = Hx(t) \tag{3.43}$$

where H is a constant $m \times n$-matrix (m being usually much smaller than n) and x is an n-dimensional wide sense stationary process defined by the difference equation

$$x(t+1) = Ax(t) + Bv(t);\; x(-1) = x_0.$$

Here A and B are constant matrices, v is a zero mean white noise vector sequence

(3.3) of arbitrary dimension and x_0 is a zero mean random vector with covariance matrix P_0. (Since x is stationary, P_0 is in fact the covariance matrix of $x(t)$ for all t.) Also assume that x_0, v, and the observation noise w are pairwise uncorrelated. Then it is readily seen that

$$E\{x(t)x(s)'\} = A^{t-s}P_0 \quad (t \geq s) \tag{3.44}$$

so that the covariance matrix of y is

$$C(t) = HA^tP_0H' \quad (t \geq 0). \tag{3.45}$$

It is well-known that the one-step predictor

$$\hat{x}(t) = \hat{E}\{x(t) \mid z(s);\ s = 0, 1, \ldots, t-1\}$$

is generated by the Kalman–Bucy formula

$$\hat{x}(t+1) = A\hat{x}(t) + K_t[z(t) - H\hat{x}(t)];\ \hat{x}(0) = 0$$

in which it remains to determine the gain matrix

$$K_t = AQ_t(HQ_t + I)^{-1} \tag{3.46}$$

where the $n \times m$-matrix Q_t is defined by

$$Q_t = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}H'. \tag{3.47}$$

Of course the estimate $\hat{y}(t|t-1)$ is given by

$$\hat{y}(t|t-1) = H\hat{x}(t), \tag{3.48}$$

$\hat{E}$ being a linear operator.

Proceeding along the same lines as in the original proof [15], we shall now derive the fast algorithm first obtained by Lindquist. To this end, first note that by (3.5) and (3.24) we have

$$\hat{y}(t|t-1) = \sum_{i=0}^{t-1} G_{t-1}^*(-1, t-1-s)z(s)$$

which inserted into (3.47)

$$Q_t = E\{x(t)[y(t) - \hat{y}(t|t-1)]'\}$$

yields

$$Q_t = P_0H' - \sum_{s=1}^{t} A^sP_0H'G_{t-1}^*(s-1, -1)$$

where we have used (3.44) and (3.6). Analogously with Section 2 and in view of (3.25), we shall write this as

$$Q_t = (P_0H'\Phi_t)(A) \tag{3.49}$$

by which we mean that $\Phi_t(z)$ should first be premultiplied by P_0H' regarding z as a scalar, after which A is inserted with z^s in left-most position.

Then if we define Q_t^* to be

$$Q_t^* = A(P_0H'\Phi_t^*)(A) \tag{3.50}$$

we obtain in analogy with (3.31) and (3.32)

$$\begin{cases} Q_{t+1} = Q_t - Q_t^* \Gamma_t^* & (3.51) \\ Q_{t+1}^* = AQ_t^* - AQ_t \Gamma_t & (3.52) \end{cases}$$

with initial conditions $Q_0 = P_0 H'$ and $Q_0^* = AP_0 H'$, and therefore it only remains to determine Γ and Γ^*. However inserting (3.45) into (3.42), we immediately have

$$S_t = HA(P_0 H' \Phi_t^*)(A)$$

that is

$$S_t = HQ_t^*. \tag{3.53}$$

Now from (3.47), (3.43) and (3.48) we see that

$$G_{t-1}(t, t) = HQ_t$$

so that by (3.17)

$$R_t = HQ_t + I. \tag{3.54}$$

Unfortunately R_t^* cannot easily be expressed in terms of Q_t or Q_t^*, and therefore we shall need to use the recursion (3.41). Then, by using (3.36), (3.37), (3.38), (3.53) and (3.54), we can completely eliminate Γ_t, Γ_t^* and R_t from the recursions (3.51), (3.52) and (3.41) to obtain

$$\begin{cases} Q_{t+1} = Q_t - Q_t^*(R_t^*)^{-1} Q_t^{*\prime} H' & (3.55) \\ Q_{t+1}^* = AQ_t^* - AQ_t(HQ_t + I)^{-1} HQ_t^* & (3.56) \\ R_{t+1}^* = R_t^* - Q_t^{*\prime} H'(HQ_t + I)^{-1} HQ_t^* & (3.57) \end{cases}$$

with initial conditions $Q_0 = P_0 H'$, $Q_0^* = AP_0 H'$ and $R_t^* = HP_0 H' + I$. Since R_t^* is symmetric, these are $2mn + \frac{1}{2}m(m+1)$ equations to determine Q_t. If $m = 1$ we do not need (3.57) since R_t^* is then equal to R_t which is given by (3.54). This leaves us with $2n$ equations, and if we prefer we can express them directly in the gain. (See [16].) However, for computional purposes the number of equations is not our main concern. Instead we wish to minimize the number of arithmetic operations and to achieve this, our equations have to be reformulated somewhat. We refer the reader to [15] for the details on this problem.

References

[1] N. I. AKIEZER, *The Classical Moment Problem*, Hafner Publishing Company, New York, 1965.

[2] 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.

[3] R. S. BUCY and P. D. JOSEPH, *Filtering for Stochastic Processes with Applications to Guidance*, Interscience Publishers, New York, 1968.

[4] 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.

[5] J. CASTI and R. KALABA, *Imbedding Methods in Applied Mathematics*, Addison-Wesley, Reading, Mass., 1973.

[6] J. Casti and E. Tse, Optimal linear filtering theory and radiative transfer: Comparisons and interconnections, *J. Math. Anal. and Appl.*, vol. 40, October 1972, pp. 45–54.
[7] S. Chandrasekhar, *Radiative Transfer*, Oxford University Press, 1950.
[8] L. Ya. Geronimus, *Orthogonal Polynomials*, Consultant Bureau, New York, 1961.
[9] T. Kailath, Some Chandrasekhar-type algorithms for quadratic regulators, *Proc. IEEE* Decision and Control Conference, Dec. 1972.
[10] T. Kailath, Some new algorithms for recursive estimation in constant linear systems, *IEEE Trans. Information Theory*, Vol. IT–19, November 1973, pp. 750–760.
[11] 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.
[12] 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.
[13] M. Krein, Continuous analogs of propositions concerning polynomials orthogonal on the unit circle, *Dokl. Akad. NAUK. SSSR*, **105**, 1955, pp. 637–640.
[14] 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.
[15] A. Lindquist, A new algorithm for optimal filtering of discrete-time stationary processes, *SIAM J. Control*, **4**, 1974.
[16] A. Lindquist, Optimal filtering of continuous-time stationary processes by means of the backward innovation process, *SIAM J. Control*, **4**, 1974.
[17] A. Lindquist, Optimal control of linear stochastic systems with applications to time lag systems, *Information Sci.*, Vol. 5 (Jan. 1973), pp. 81–126.
[18] J. Rissanen, Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with applications to factoring positive matrix polynomials, *Mathematics of Computation*, **27** (1973), pp. 147–154.
[19] J. Rissanen, A fast algorithm for optimum linear prediction, *IEEE Transactions Automatic Control*, vol. AC-18, p. 555, Oct. 1973.
[20] R. A. Wiggins and E. A. Robinson, Recursive solutions to the multichannel filtering problem, *J. Geophys. Research*, **70**, pp. 1885–1891 (1965).
[21] 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.
[22] T. Kailath, Application of a resolvent identity to a linear smoothing problem, *SIAM J. Control*, vol. 7, February 1969, pp. 68–74.

(*Received March 2, 1974*)