An Innovations Approach to Least-Squares Estimation—Part VI: Discrete-Time Innovations Representations and Recursive Estimation

MICHEL R. GEVERS, MEMBER, IEEE, AND THOMAS KAILATH, FELLOW, IEEE

Abstract—The linear stochastic discrete-time realization problem is to find a white-noise driven finite-dimensional linear system whose output generates a specified separable covariance. The solution to this problem is presented in the form of a causal and causally invertible innovations representation (IR) whose existence depends only on the positive definite nature of the separable covariance. It is also shown that least-squares filtered and smoothed estimates of one process given observations of a related colored process can be expressed as linear combinations of the state vector of the IR of the observed process. The analogous continuous-time problems have been studied earlier, and it has been shown that an important role is played by what is known as the relative order of the covariance. Here this is defined as the number of differencing operations required to produce a delta function component in the differenced covariance. It is shown that, unlike the continuous-time case, the relative order of the covariance does not necessarily induce similar (relative order) constraints on the impulse response of all models whose responses to white noise have the given covariance. This fact is at the heart of certain differences between continuous-time and discrete-time results. It is shown, however, that the innovations representations obey a number of constraints equal to the relative order of the covariance.

I. Introduction and Outline of Results

THE linear stochastic realization problem is: "Given the covariance of a signal process, find a linear system driven by white noise whose output has this covariance." It has also been called the time-varying spectral factorization or, more correctly, the covariance factorization problem. The linear system will be said to generate the specified covariance.

In its easiest form, the problem is confined to stationary and often scalar processes that can be generated by finite-dimensional systems. In recent years, however, several authors have presented solutions to various aspects of the time-varying covariance factorization problem for continuous-time processes (see [1]–[9]).

The basic problem here is one in which the specified signal covariance, say $R(t,s)$, contains a white-noise component: in operator form

$$R = I + K. \tag{1}$$

For the special case of a separable K, Anderson, Moore, and Loo [5] and also Brandenburg [7] gave a solution, under certain conditions, to this problem in the form of a class of linear finite-dimensional systems that generate the covariance R. It has been known for a long time that there are essentially distinct minimal systems that can generate a given covariance. In independent work Kailath and Geesey [1], [8], obtained a unique (up to impulse response) solution by requiring that the linear system not only be causal but also causally invertible, leading to what have been called innovations models or innovations representations (IR's) [10]. The general properties of IR's were studied by Levy [11], Hida [12], and Cramer [13] who called them proper canonical representations.

The case in which the process does not explicitly contain white noise, namely, where its covariance R is smooth, is more difficult. Of particular interest for such a process is what has been called [2], [3] its definite relative order α. Essentially this is one half the number of times the covariance has to be differentiated in order that it contain an added delta function. The covariance of the αth derivative of the process is then of the form $I + K$. For such processes the most general solution of the singular problem has been given by Geesey and Kailath [2]. (For separable K, and with further assumptions, the problem was also solved independently by Moore and Anderson [6] and Brandenburg [7].) Geesey and Kailath described a Wiener–Hopf equation and, for separable K, a Riccati equation whose solution determined the IR. They also showed that the output of a state-variable system driven by white noise has a covariance of definite relative order α if and only if the parameters and the state variance of this system obey α constraints, which can be used to reduce the order of the Riccati equation by α. The point is that because of the smoothness of the process $y(\cdot)$, α different projections of the state are calculable without error as some linear combination of the signal $y(\cdot)$ and its first α-1 derivatives.

Although several of these results carry over essentially unchanged (derivatives must be replaced by differences, and so on) to the discrete-time case, there are some important differences. For example, the absence of a white-noise component does not make a great difference to the

Manuscript received July 7, 1972. Paper recommended by P. A. Frost, Chairman of the IEEE S-CS Estimation and Identification Committee. This work was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, under Contract AF 44-620-69-C-0101 and in part by the Joint Services Electronics Program under Contract N-00014-67-A-0112-0044.

M. R. Gevers was with the Department of Electrical Engineering, Stanford University, Stanford, Calif. 94305. He is now at Batiment Maxwell, B-1348 Louvain-la-Neuve, Belgium.

T. Kailath is with the Department of Electrical Engineering, Stanford University, Stanford, Calif. 94305.

discrete-time problem, as could have been anticipated from the fact that the Kalman filter formulas in discrete time are not critically modified, as they are in continuous time, by the absence of an additive white-noise term [14]. In fact, we shall show (in Section IV) that in discrete time, a causal and causally invertible lumped innovations representation exists whenever a separable positive-definite symmetric covariance is given. However, several different but closely related state-variable forms of the innovations representation will be presented, all obtained through the solution of a Riccati equation.[1] The existence and some properties of such forms were first pointed out in [15], where it was shown that the different forms essentially arise from using as states of the IR the filtered estimates $\hat{x}(k|k)$, say, or the λ-step predictions $\hat{x}(k|k-\lambda)$, $\lambda = 1$ or $2 \cdots$. This distinction, which cannot be made in continuous time without introducing a pure time delay (and thus destroying finite dimensionality), will be seen to be at the heart of some differences between continuous-time and discrete-time results.

A major difference is that while *every* realization of a continuous-time covariance of definite relative order α has to obey α constraints, different realizations are possible in discrete time, some of which have constrained states and some of which do not. We have shown [16] that this distinction explains the observation of Bucy, Rappaport, and Silverman [17], [18] that in certain discrete-time filtering problems, differencing the observations does not have the same computation-reducing consequences as the analogous differentiation process always has in continuous time. An equivalent way of stating the above distinction is that while in continuous time the relative order of a covariance is preserved in all impulse responses that generate it (see, e.g., [2, lemma 1]), in discrete time the relative order of the covariance and of the impulse response obtained by factorization are not necessarily equal. This fact can best be seen in the scalar stationary case. In continuous time, the relative order of a transfer function (respectively a covariance) is the difference (respectively one half the difference) between the degrees of the denominator and the numerator polynomials. This relative order is clearly preserved in the factorization since

$$S(s) = H(s)H(-s) \tag{2}$$

where $S(s)$ is the power spectral density of the process and $H(s)$ is the transfer function of a filter whose response to white noise has power spectral density $S(s)$. In discrete time the relative order of both the transfer function and the covariance is the difference between the degrees of the denominator and the numerator polynomials. That this relative order is not preserved is evidenced by the fact that the factorization in discrete time is determined only up to powers of z since

$$\begin{aligned} S(z) &= H(z)H(z^{-1}) \\ &= z^k H(z) z^{-k} H(z^{-1}). \end{aligned} \tag{3}$$

[1] We should point out that solving a Riccati equation is not necessarily the only way of obtaining a realization.

In continuous time, the analog of z^k is $e^{s\lambda}$, but this factor cannot be introduced without destroying the finite dimensionality of $H(s)$; there is no such problem in discrete time. In Sections V and VI we shall show how this possibility affects the structural properties of the realizations, and in particular the amount of reduction in the order of the Riccati equation associated with the factorization problem.

Finally, in Section VIII we shall show how knowledge of the IR of the observed process $y(\cdot)$ leads, almost by inspection, to the determination of the filtered and smoothed estimates of a related process for which only cross-covariance information, rather than a model, is available.

Historical Remarks

Earlier results on discrete-time realization were obtained in [19]–[21] and [15]. In [19], various forms were obtained by using an associated control problem (see the further discussion at the end of Section IV), but invertibility was not explicitly discussed. In [20], a form similar to the one in our Theorem 1 (15) was obtained by using an existence argument similar to one presented in continuous time by Kailath and Geesey [2] (see the further discussion in Section IV). In [21], the form (15) was obtained as we do here, viz., by a method similar to that used in Kailath and Geesey [1],[2]. The possibility of forms other than (15) was first noted in Kailath [15], where the close relationship between the estimation problem and the factorization problem was fully exploited. All these forms have the same impulse response, a fundamental property of IR's [10], but they differ by more than just a state transformation (see the further discussion in Section III). The detailed structural analysis of the various IR's and other models was carried out in the thesis [22] and is related to the studies made in [2] and [21]. We may also note here that the authors of [20] also state that their results on realization had first been given in an internal report by Colebatch (Tech. Rep. EE-6813, University of Newcastle, New South Wales, Australia, 1968), although the validation was much less direct than in [20], being similar to the one used in continuous time by Anderson, Moore, and Loo [5].

II. Relations Between Models and Covariances

We shall consider a p-vector valued discrete-time process $y(\cdot)$ over an interval $[0,N]$, with N possibly infinite, and with a separable covariance of the form

$$R_y(i,j) = A(i)B(j)1(i-j) + B'(i)A'(j)1(j-i-1) \tag{4}$$

where $1(\cdot)$ is the step function

$$\begin{aligned} 1(i-j) &= 1, \quad \text{if } i \geq j \\ &\quad 0, \quad \text{if } i < j \end{aligned} \tag{5}$$

and $A(\cdot)$ and $B(\cdot)$ have dimensions $p \times n$ and $n \times p$, respectively. This is not the most general form of a discrete-time separable covariance because it does not contain a delta component as, for example,

$$R_y(i,j) = A(i)B(j)1(i-j-1) + B'(i)A'(j)1(j-i-1) + L(i)\delta(i-j) \quad (6)$$

where

$$\delta(i-j) = 1, \quad \text{for } i = j, \ \delta(i-j) = 0, \quad \text{for } i \neq j.$$

This last form originates from a process that contains a white-noise component. However, as we noted earlier, the realization theory for smooth covariances exhibits somewhat greater differences from the corresponding continuous-time case. Therefore, we shall devote most of our attention to (4).

Now suppose $y(\cdot)$ is known to arise from some lumped model of the form

$$x(i+1) = \phi(i+1,i)x(i) + G(i+1)u(i+1), \quad x(0) = x_0 \quad (7a)$$

$$y(i) = H(i)x(i), \quad 0 \leq i \leq N \quad (7b)$$

where $\phi(\cdot,\cdot)$, $G(\cdot)$, $H(\cdot)$ are known functions and $\{x_0, u(\cdot)\}$ are zero-mean random variables with

$$E[x_0u'(\cdot)] = 0, \quad E[x_0x_0'] = \Pi_0, \quad (7c)$$

$$E[u(k)u'(l)] = I\delta(k-l). \quad (7d)$$

Such a model, with the outputs a linear combination of the states, has been called a *Markovian* representation of the process $y(\cdot)$. Note that $y(\cdot)$ is not Markov, but is the so-called projection of a Markov process, i.e., a linear combination of the components of a vector Markov process. The model (7) is causal, although it may not be causally invertible. Let us denote the variance matrix of the states at time i by

$$\Pi(i) = E[x(i)x'(i)]. \quad (8)$$

Then it is well known and is easy to show that $\Pi(\cdot)$ obeys the difference equation

$$\Pi(i+1) = \phi(i+1,i)\Pi(i)\phi'(i+1,i) + G(i+1)G'(i+1)$$
$$\Pi(0) = \Pi_0. \quad (9)$$

Therefore the covariance of the output $y(\cdot)$ can be expressed as

$$R_y(i,j) = H(i)\phi(i,j)\Pi(j)H'(j)1(i-j) + H'(i)\Pi(i)\phi'(j,i)H'(j)1(j-i-1). \quad (10)$$

Comparing (4) with (10), we conclude that one set of relations between the parameters of the system (7) and the covariance parameters of (4) is

$$H(i)\phi(i,0) = A(i) \quad (11a)$$

$$\phi(0,i)\Pi(i)H'(i) = B(i), \quad 0 \leq i \leq N. \quad (11b)$$

Finally, let

$$M(i) = A(i)\phi(0,i), \quad N(i) = \phi(i,0)B(i) \quad (12)$$

in terms of which we can also write (4) as

$$R_y(i,j) = M(i)\phi(i,j)N(j)1(i-j) + N'(i)\phi'(j,i)M'(j)1(j-i-1). \quad (13)$$

Then an obvious identification between (10) and (13) gives the relation

$$H(i) = M(i) \quad (14a)$$

$$\Pi(i)M'(i) = N(i), \quad 0 \leq i \leq N. \quad (14b)$$

These several formulas will be used presently.

A Comment on the Invertibility of $\phi(\cdot,\cdot)$

We should note that the identifications in (11) require that the state-transition matrix $\phi(i,j)$ of the given model (7) must be nonsingular[2] for $i > j$. In our work, however, this assumption is not restrictive because our goal is to construct models corresponding to given covariances. If the columns of $A(\cdot)$ and the rows of $B(\cdot)$ are linearly independent, and if we choose models of order not greater than the number of such rows or columns, then we can always choose a nonsingular $\phi(\cdot,\cdot)$.

III. Innovations Representations of a Lumped Markovian Process

In the first theorem we shall assume that the process is known to arise from some lumped model. This will be relaxed later when the existence of the IR will be shown to depend only upon the positive definiteness of the given covariance.

Theorem 1—Innovations Representation 1 (IR-1): Let $y(\cdot)$ have a positive definite covariance of the form (13) and assume that $y(\cdot)$ is known to arise from some lumped model. Then an innovations representation for $y(\cdot)$ can be written in the form

$$\theta(i+1) = \phi(i+1,i)\theta(i) + \phi(i+1,i)K(i)\varepsilon(i), \ \theta(0) = 0 \quad (15a)$$

$$y(i) = M(i)\theta(i) + \varepsilon(i) \quad (15b)$$

where $\varepsilon(\cdot)$ is a zero-mean white-noise process with covariance

$$E[\varepsilon(i)\varepsilon'(j)] = [M(i)N(i) - M(i)\Sigma_1(i)M'(i)]\delta(i-j) \quad (15c)$$

and

$$K(i) = [N(i) - \Sigma_1(i)M'(i)][M(i)N(i) - M(i)\Sigma_1(i)M'(i)]^{-1}. \quad (15d)$$

The matrix $\Sigma_1(i)$ is the state variance

$$\Sigma_1(i) = E[\theta(i)\theta'(i)] \quad (15e)$$

and obeys the matrix Riccati difference equation

[2] Notice that while the state-transition matrix is always nonsingular in continuous time, this is not necessarily so in discrete time. This is another source of the differences between continuous and discrete time, although it is not particularly so in this paper.

$$\Sigma_1(i+1) = \phi(i+1,i)\Sigma_1(i)\phi'(i+1,i) + \phi(i+1,i) \cdot [N(i) - \Sigma_1(i)M'(i)][M(i)N(i) - M(i)\Sigma_1 \cdot (i)M'(i)]^{-1} [N(i) - \Sigma_1(i)M'(i)]'\phi'(i+1,i)$$

$$\Sigma_1(0) = 0. \tag{15f}$$

Proof: The proof follows a procedure used in [1]. We begin by assuming that the process $y(\cdot)$ is the output of a known model of the form (7), and show how to find the IR for it. This IR is obtained as a rearrangement of the Kalman filter equations of the system. Since it is known, moreover (see [10]–[13]), that the IR is uniquely determined up to its impulse response by the covariance of the process, we should be able to express the IR entirely in terms of this covariance function.

The Kalman filter equations for the one-step prediction estimates of the state are (see [14])

$$\hat{x}(i+1|i) = \phi(i+1,i)\hat{x}(i|i-1) + \phi(i+1,i)K(i)\varepsilon(i) \tag{16a}$$

$$\varepsilon(i) = y(i) - M(i)\hat{x}(i|i-1) \tag{16b}$$

$$\hat{x}(0|-1) = 0 \tag{16c}$$

where

$$K(i) = P(i|i-1)M'(i)[M(i)P(i|i-1)M'(i)]^{-1}. \tag{16d}$$[3]

$P(i|i-1)$ is the covariance of the instantaneous error

$$P(i|i-1) = E[(x(i) - \hat{x}(i|i-1))(x(i) - \hat{x}(i|i-1))'] \tag{16e}$$

and satisfies the matrix Riccati equation

$$P(i+1|i) = \phi(i+1,i)P(i|i-1)\phi'(i+1,i) + G(i+1)G'(i+1) - \phi(i+1,i) \cdot P(i|i-1)M'(i)[M(i)P(i|i-1)M'(i)]^{-1} \cdot M(i)P(i|i-1)\phi'(i+1,i)$$

$$P(0|-1) = \Pi_0. \tag{16f}$$

The $\{\varepsilon(i)\}$ are the innovations of the process $y(\cdot)$. Rearranging (16b) and replacing $\hat{x}(i|i-1)$ by $\theta(i)$ shows that (15a),(15b), together with (16d)–(16f), is another model for the process $y(\cdot)$. It can also be verified directly, if so desired, that the outputs of both models have the same covariance.

Next we shall show how to express the IR (15a),(15b) directly in terms of the parameters of the covariance function (13). To do so, we define the variance matrix of the state estimates as

$$\Sigma_1(i) = E[\hat{x}(i|i-1)\hat{x}'(i|i-1)]. \tag{17}$$

By the projection theorem the state estimates are orthogonal to the error in the estimates, and hence

$$P(i|i-1) = \Pi(i) - \Sigma_1(i) \tag{18}$$

where $P(i|i-1)$ and $\Pi(i)$ obey (16f) and (16g), respectively. Substituting (18) in (16f) and subtracting (16g), using (14), gives (15f) for $\Sigma_1(\cdot)$. A similar substitution in (16d) gives the expression (15d) for the gain.

The innovations process $\varepsilon(\cdot)$ driving IR—1 [see (15a) and (15b)] is a whitened version of the observation process, as is well known [10],[14]. Its variance is easily obtained noting that

$$\varepsilon(i) = M(i)[x(i) - \hat{x}(i|i-1)]. \tag{19}$$

The expression (15c) follows immediately using (18).

The existence of a solution to the nonlinear equation (15f) follows from the existence of a model which guarantees the existence of Π and hence of $\Sigma_1 \leq \Pi$. We shall show in Section IV how this assumption of existence can be relaxed. ∎

Corollary 1—Whitening Filter: The innovations representation (IR-1) is causal and causally invertible, and the inverse serves as a whitening filter. Its equations are

$$\varepsilon(i) = y(i) - M(i)\theta(i) \tag{20a}$$

$$\theta(i+1) = \phi(i+1,i)\theta(i) + \phi(i+1,i)K(i)[y(i) - M(i)\theta(i)], \quad \theta(0) = 0 \tag{20b}$$

where $K(\cdot)$ is obtained through (15d)–(15f).

Proof: The proof is just a trivial rewriting of the equations (15) of the innovations representation. ∎

In the above we used the Kalman filter equations for the one-step predictions $\hat{x}(i|i-1)$. However, it is well known that the Kalman equations can also be described in terms of filtered estimates $\hat{x}(i|i)$. This distinction, which is effectively absent in continuous time, leads to an alternative representation. The use of IR's based on filtered or predicted estimates [cf. (31)] was first noted in [15].

Theorem 2—Innovations Representation 2 (IR-2): Let $y(\cdot)$ have a positive definite covariance of the form (13) and suppose that it is known to arise from some lumped model. Then an innovations representation for this process can be written in the form

$$\mathcal{C}(i+1) = \phi(i+1,i)\mathcal{C}(i) + K(i+1)\varepsilon(i+1), \quad \mathcal{C}(-1) = 0 \tag{21a}$$

$$y(i) = M(i)\mathcal{C}(i) \tag{21b}$$

where $\varepsilon(i)$ is a zero-mean white-noise process with covariance

[3] The existence of the inverse requires the assumption that the covariance $R_y(i,j)$ is strictly positive for $0 \leq i, j \leq N$ (see Section IV). This is a minimal assumption which is equivalent to the condition that no $y(i)$ be a linear combination of the other $y(\cdot)$; this assumption ensures that there is a *unique* relation between $y(\cdot)$ and its innovations sequence $\epsilon(\cdot)$. While the assumption can be relaxed by using Moore–Penrose pseudoinverses, we do not feel it useful to indulge in such pedantry here.

$$E[\mathcal{E}(i)\mathcal{E}'(j)] = [M(i)N(i) - M(i)\phi(i,i-1) \cdot \Sigma_2(i-1)\phi'(i,i-1)M'(i)]\delta(i-j) \quad (21c)$$

and

$$K(i) = [N(i) - \phi(i,i-1)\Sigma_2(i-1)\phi'(i,i-1)M'(i)] \cdot [M(i)N(i) - M(i)\phi(i,i-1)\Sigma_2(i-1) \cdot \phi'(i,i-1)M'(i)]^{-1}. \quad (21d)$$

$\Sigma_2(\cdot)$ is the state variance

$$\Sigma_2(i) = E[\mathcal{C}(i)\mathcal{C}'(i)] \quad (21e)$$

and obeys the Riccati equation

$$\begin{aligned} \Sigma_2(i+1) &= \phi(i+1,i)\Sigma_2(i)\phi'(i+1,i) + [N(i+1) \\ &\quad - \phi(i+1,i)\Sigma_2(i)\phi'(i+1,i)M'(i+1)] \\ &\quad \cdot [M(i+1)N(i+1) - M(i+1)\phi(i+1,i)\Sigma_2(i) \\ &\quad \cdot \phi'(i+1,i)M'(i+1)]^{-1} \cdot [N(i+1) - \phi(i+1,i) \\ &\quad \cdot \Sigma_2(i)\phi'(i+1,i)M'(i+1)]', \Sigma_2(-1) = 0. \end{aligned} \quad (21f)$$

Proof: The proof is completely analogous to the proof of Theorem 1. Alternatively, we can also obtain IR-2 from IR-1 directly by using the fact that

$$\hat{x}(i+1|i) = \phi(i+1,i)\hat{x}(i|i) \quad (22a)$$

and hence that

$$\Sigma_1(i) = \phi(i+1,i)\Sigma_2(i)\phi'(i+1,i) \quad (22b)$$

where $\Sigma_2(\cdot)$ denotes the variance of the filtered estimates

$$\Sigma_2(i) = E[\hat{x}(i|i)\hat{x}'(i|i)]. \quad (23)$$

Substituting these relations in (16a)–(16c) and (15d)–(15f), and replacing $\hat{x}(i|i)$ by $\mathcal{C}(i)$ gives

$$\mathcal{C}(i+1) = \phi(i+1,i)\mathcal{C}(i) + K(i+1)\mathcal{E}(i+1), \quad \mathcal{C}(-1) = 0 \quad (24a)$$

$$y(i) = M(i)\phi(i,i-1)\mathcal{C}(i-1) + \mathcal{E}(i) \quad (24b)$$

where $K(i)$ is obtained through (21d) and (21f).

Finally we notice that

$$M(i)K(i) = I. \quad (25)$$

Hence (24b) can also be written as (21b), using (24a) and (25).

It should go without saying (see footnote 3) that the inverse in (21d) exists because it is the innovations variance which is nonsingular because $R_y(i,j)$ is nonsingular. ■

Corollary 2—Whitening Filter: The innovations representation (IR-2) is causal and causally invertible. Its inverse, which is a whitening filter, is given by

$$\mathcal{E}(i) = y(i) - M(i)\phi(i,i-1)\mathcal{C}(i-1), \quad \mathcal{C}(-1) = 0 \quad (26a)$$

$$\mathcal{C}(i+1) = \phi(i+1,i)\mathcal{C}(i) + K(i+1) \cdot [y(i+1) - M(i+1)\phi(i+1,i)\mathcal{C}(i)] \quad (26b)$$

where $K(\cdot)$ is given by (21d) and (21f).

Proof: Equations (26) are easily obtained by a rearrangement of (21a) and (21b). ■

Normalization of the Input White Noise

Notice that the variance of the innovations process in both IR-1 and IR-2 is a time-varying function that depends on the solution of a Riccati equation [see (15c) and (15f) and (21c) and (21f)]. This is in contrast to the continuous-time case where the variance of the innovations is equal to the variance of the white-noise component in the covariance of the process. It may be convenient to have an IR that is driven by unit variance white noise. This can be easily achieved by defining normalized innovations $\nu(\cdot)$ that have unit variance. For IR-2 we define

$$\nu(i) = [M(i)N(i) - M(i)\phi(i,i-1)\Sigma_2(i-1) \cdot \phi'(i,i-1)M'(i)]^{-1/2}\mathcal{E}(i). \quad (27)$$

The equations of IR-2 are now

$$\mathcal{C}(i+1) = \phi(i+1,i)\mathcal{C}(i) + L(i+1)\nu(i+1), \quad \mathcal{C}(-1) = 0 \quad (28a)$$

$$y(i) = M(i)\mathcal{C}(i) \quad (28b)$$

where $\nu(\cdot)$ is unit variance white noise and

$$L(i) = [N(i) - \phi(i,i-1)\Sigma_2(i-1)\phi'(i,i-1)M'(i)] \cdot (M(i)N(i) - M(i)\phi(i,i-1) \cdot \Sigma_2(i-1)\phi'(i,i-1)M'(i)]^{-1/2}. \quad (28c)$$

Comparison of IR-1 and IR-2

We have obtained two different IR's for the same process, and we have shown that they can be written entirely in terms of the parameters of its covariance. Note that, as might be expected, these two representations have the same impulse response

$$y(i) = \sum_{j=0}^{i} M(i)\phi(i,j)K(j)\mathcal{E}(j). \quad (29)$$

The point is that the states of IR-2 are a transformation of the states of IR-1 [see (22a)]. However, this is not an ordinary coordinate transformation because there is a one-unit delay between corresponding states.

There is an interesting difference between the output equations (15b) and (21b) for IR-1 and IR-2, respectively. While IR-2 is of the form of the original model (7), the output equation in IR-1 contains an added white-noise term. However, notice that this last equation can also be rewritten, using (15a), in the form

$$y(i) = M(i)\phi(i,i+1)\theta(i+1). \quad (30)$$

While in this form $y(i)$ is written as a linear combination of the states, as in the model (7), it should be observed that there is now a one-unit delay between the states and the outputs. However, the noncausal relationship between outputs and states in this modified version of IR-1 does not affect the causal invertibility between outputs and inputs.

Other Representations

The ideas used to obtain IR-1 and IR-2 could be extended to obtain representations whose states are 2, 3, $\cdots$, λ-step predicted estimates of the form

$$\hat{x}(i+\lambda|i) = \phi(i+\lambda, i+\lambda-1)\hat{x}(i+\lambda-1|i-1) + K(i)\mathcal{E}(i) \quad (31a)$$

$$y(i) = M(i)\phi(i, i+\lambda)\hat{x}(i+\lambda|i) \quad (31b)$$

for some $\lambda > 0$. Similarly, in continuous time, innovations models can be obtained whose states are predicted estimates $\hat{x}(t+\lambda|t)$ for some fixed positive λ. However, when $\lambda > 0$ the models will contain pure delays and will not be finite dimensional.

IV. Existence of the Innovations Representation

We shall show now that the assumption that the process $y(\cdot)$ arises from some lumped model is not required. Just the basic assumption that $R_y(\cdot,\cdot)$ is a positive definite covariance on $[0,N]$ will suffice to demonstrate the existence of IR-2. The existence of IR-1 then follows by the relation (22a). The proof will parallel an argument used in continuous time by Kailath and Geesey (cf. [2, Appendix III) and [3, lemmas 3.6 and 3.7]). The continuous-time proof needs to start with covariances that have a delta-function component, but as we might expect, this is not necessary in discrete time. We may note that Son and Anderson apparently rediscovered the discrete-time analog of the just-cited proof in [2] and [3].

We consider again a process $y(\cdot)$ over $[0,N]$ with a separable covariance of the form (13). We first note that if $R(\cdot,\cdot)$ is a positive definite covariance, there exists a causal system that, when driven by white noise, generates this covariance. We shall subsequently show that this system is finite dimensional.

Proposition 3: Let $R_y(i,j)$ in (13) be a positive definite covariance on $[0,N]$. Then there exists a causal impulse response matrix $h(\cdot,\cdot)$ such that a system excited by unit variance white noise and with $h(\cdot,\cdot)$ as impulse response has output covariance $R_y(\cdot,\cdot)$, i.e.,

$$R_y(i,j) = \sum_{k=0}^{N} h(i,k)h'(j,k) \quad (32)$$

with

$$h(i,j) = 0 \qquad \text{for } i < j. \quad (33)$$

Proof: The result is well known. It relies on the fact that the $Np \times Np$ covariance matrix

$$R = [R(i,j)], \qquad 0 \le i,j \le N \quad (34)$$

can be factored into a lower-triangular matrix H and its transpose

$$R = H \cdot H'. \quad (35)$$

H is partitioned in $p \times p$ blocks $h(i,j)$, and the result follows. Notice that the positive definiteness of R implies that $h(i,i)$ is nonsingular for all i since

$$0 < \det R = \prod_{i=0}^{N} [\det h(i,i)]^2. \quad \blacksquare(36)$$

Theorem 4: Let $h(\cdot,\cdot)$ be defined as in Proposition 3. Then $h(i,j)$ has the form

$$h(i,j) = M(i)\phi(i,j)L(j)1(i-j), \qquad 0 \le i,j \le N \quad (37)$$

for some matrix $L(\cdot)$ and with $M(\cdot)$ and $\phi(\cdot,\cdot)$ defined by (13).

Proof: Since R is positive definite, so is H. Let V be the inverse of H. Then V is also a lower-triangular matrix. Let V be partitioned like H. Then, using (35), we have

$$H = R \cdot V' \quad (38)$$

with

$$V' = \begin{bmatrix} v'(0,0) & v'(1,0) & \cdots & v'(N,0) \\ 0 & v'(1,1) & & \\ & & \ddots & \\ 0 & \cdots & 0 & v'(N,N) \end{bmatrix}.$$

Then we can write, using (38), (13), and the causality of $h(\cdot,\cdot)$,

$$\begin{aligned} h(i,j) &= \sum_{k=0}^{j} R(i,k)v'(j,k) \\ &= \sum_{k=0}^{j} M(i)\phi(i,k)N(k)v'(j,k)1(i-j) \\ &= M(i)\phi(i,j)\left[\sum_{k=0}^{j} \phi(j,k)N(k)v'(j,k)\right]1(i-j) \\ &= M(i)\phi(i,j)L(j)1(i-j) \end{aligned}$$

with

$$L(j) \triangleq \sum_{k=0}^{j} \phi(j,k)N(k)v'(j,k). \quad \blacksquare(39)$$

We may observe that the above argument actually already gives a causal and causally invertible model for the process. It can be shown after some algebraic manipulations (see [22]) that $L(\cdot)$ as defined in (39) is identical to the gain $L(\cdot)$ of IR-2 [see (28c)] where $\Sigma_2(\cdot)$ obeys the Riccati equation (21f), which proves the existence of IR-1 and IR-2. However, we feel that the approach taken in Theorems 1 and 2 gives more insight and is more helpful in applications to estimation problems.

Another characterization of IR-2 is in its minimum properties. If one considers all models of the form (7) for a process $y(\cdot)$ with covariance (13), then the realization problem reduces to finding all solutions $\Pi(\cdot)$ and $G(\cdot)$ of the set of equations

$$\Pi(i+1) = \phi(i+1,i)\Pi(i)\phi'(i+1,i) + G(i+1)G'(i+1) \quad (40a)$$

$$\Pi(0) = \Pi_0 \quad (40b)$$

$$\Pi(i)M'(i) = N(i). \quad (40c)$$

Indeed, once $H(\cdot)$ and $\phi(\cdot,\cdot)$ in (7) have been identified with the corresponding parameters $M(\cdot)$ and $\phi(\cdot,\cdot)$ of (13), the only way models generating $R_y(\cdot,\cdot)$ can differ is the matrix $G(\cdot)$ and the initial conditions matrix Π_0.

In the stationary steady-state case, Faurre [19] has analyzed the realization problem from this point of view, namely: find all nonnegative matrices Π and Q that satisfy the sets of constraints

$$\Pi = \phi\Pi\phi' + Q \quad (41a)$$

$$\Pi H' = N. \quad (41b)$$

There are, in general, infinitely many solutions (Π,Q) to these equations. To obtain a unique solution, Faurre introduced the usual partial ordering on the set of nonnegative definite solutions Π, viz., $\Pi_1 \geq \Pi_2$ if $\Pi_1 - \Pi_2$ is nonnegative definite. He then obtained a smallest and a largest solution for Π with the help of an associated optimal control problem. These solutions were found to be the asymptotic solution of a Riccati equation obtained by applying a matrix inversion lemma. It turns out that Faurre's Riccati equation for the smallest solution Π of (41) is precisely (21f), and that the model that could be obtained by factorization of his steady-state solution for Q into GG' is precisely the steady-state version of IR-2. In [22] a simple proof has been given of the fact that of all solutions $\{\Pi(i),G(i+1), 0 \leq i \leq N\}$ of (40), the realization with the smallest variance is IR-2, in the sense that if $\Pi^*(i)$ is the state-variance of any other realization, then

$$\Pi^*(i) - \Pi(i) \geq 0. \qquad 0 \leq i \leq N.$$

The proof of this result and some further implications of this minimality property will be presented elsewhere.

V. Relative Order Properties of Covariances and Their Realizations

So far we have shown how to obtain IR's for separable covariances that do not contain a (Kronecker) delta function component. It may be remembered, however, that in continuous time the realization problem is significantly more complicated when the covariance of the process is smooth [2] than when it contains a delta function component [1]. In the former case it is generally necessary to introduce what has been called the definite relative order of the process (see [2] and [3]), which is the number of differentiations required to produce a delta function term in the differentiated covariance function. It has been shown (see, e.g., [2, lemma 1]) that in continuous time the relative order of the process (or the covariance) is equal to the relative order of all models whose response to white noise has the given covariance; the *relative order of a model* is defined as the number of output differentiations required to produce a term proportional to the input in the differentiated output of the model. The relative order of the model induces constraints that can be used in the realization problem to reduce the order of the Riccati equation associated with the covariance factorization [2]. Thus, in continuous time, the realization of a separable covariance of definite relative order α can be achieved through the solution of a Riccati equation of order $n-\alpha$ where n is the dimension of the minimal realization.

In discrete time, we shall similarly define the relative order of a covariance as the number of differencing operations required to produce a (Kronecker) delta function component of the differenced covariance; for stationary processes it is also equal to the difference between the degrees of the denominator and numerator polynomials of the power spectral density function (the z-transform of the covariance function). The relative order of a model is equal to the number of differencing operations required to produce a term proportional to the input in the differenced output; for constant systems it is also equal to the difference between the degrees of the denominator and numerator polynomials of the transfer function (the z-transform of its impulse response). As in continuous time, the relative order, say q, of a model induces constraints on the parameters of the model that can be used to reduce by q the order of the Riccati equation in the covariance factorization problem.

However, as we shall see now, the relative order α of a given covariance function has no unique relationship, in discrete time, to the relative order q of a model whose response to white noise has the given covariance. Whereas in continuous time $q = \alpha$ for all state-variable representations, in discrete time one can only say that $q \leq \alpha$. However, we shall show that there exists a class of representations whose relative order equals the relative order of the process, and that the innovations representations belong to that class. For these models the order of the Riccati equation associated with the realization problem can be reduced by α.

Before turning to the proofs of these results, let us note that although our results are stated for scalar processes, they can be essentially carried over, although with a heavy notational burden, to the vector case.

We shall say, following [2] and [3], that the covariance (4) has definite relative order α if there exists a finite integer α, $\alpha > 0$ such that

1) $A(i-k)B(i) - B'(i-k)A'(i) = 0,$
$$k = 1,2,\cdots,\alpha-1 \qquad k \leq i \leq N \quad (42a)$$

2) $A(i-\alpha)B(i) - B'(i-\alpha)A'(i) = C(i)C'(i) > 0,$
$$\alpha \leq i \leq N. \quad (42b)$$

Let (4) be the covariance of the output of a Markovian model

$$x(i) = x(i-1) + G(i)u(i), \qquad x(0) = x_0 \quad (43a)^4$$

$$y(i) = H(i)x(i), \qquad 0 \le i \le N \quad (43b)$$

where

$$E[u(i)] = 0, \qquad E[u(i)u'(j)] = I\delta(i-j) \quad (43c)$$

$$E[u(i)x_0'] = 0, \quad E[x_0x_0'] = \Pi_0. \quad (43d)$$

The state variance obeys the recursive equation

$$\Pi(i) = \Pi(i-1) + G(i)G'(i), \qquad \Pi(0) = \Pi_0. \quad (44)$$

The covariance of the outputs of (43) is given by

$$R_y(i,j) = H(i)\Pi(j)H'(j)1(i-j) + H(i)\Pi(i)H'(j)1(j-i-1). \quad (45)$$

Comparing with (4) we have

$$H(i) = A(i) \quad (46a)$$

$$\Pi(i)A'(i) = B(i), \qquad 0 \le i \le N. \quad (46b)$$

Multiplying (44) to the left by $A(i-1)$ and to the right by $A'(i)$ and using (46) and the relative order property (42) of the covariance, we find

$$A(i-1)B(i) - B'(i-1)A'(i) \equiv 0 = A(i-1)G(i)G'(i)A'(i). \quad (47)$$

However, contrary to what happened in continuous time (see, e.g., [2, eq. (23)], we cannot conclude here that either

$$A(i-1)G(i) = 0, \qquad 1 \le i \le N \quad (48a)$$

or

$$A(i)G(i) = 0, \qquad 0 \le i \le N. \quad (48b)$$

The only conclusion from (47) is that $A(i-1)G(i)$ and $A(i)G(i)$ are orthogonal for all i. Similarly, the following relations, which could have been expected to hold by direct analogy to the continuous-time results, do not necessarily hold in discrete time:

$$A(i-k)G(i) = 0, \qquad k = 1,2,\cdots,\alpha-1 \quad (49a)$$

$$A(i-k)x(i) = y(i-k), \qquad k = 0,1,\cdots,\alpha-1 \quad (49b)$$

$$A(i-k)\Pi(i) = B'(i-k), \qquad k = 0,1,\cdots,\alpha-1 \quad (49c)$$

$$k \le i \le N.$$

If (49a) holds for $k = 1,\cdots,q-1$, $(q \le \alpha)$, then q will be called the relative order of the representation (43) by analogy to the continuous-time definition (see, e.g., [3, definition B.1, p. 22]). We have just shown that in discrete time there is no unique relationship between the relative order q of the representation and the relative order α of the covariance of a given process. Clearly, by (47), we always have $q \le \alpha$.

[4] The choice of $u(i)$ rather than the more common $u(i-1)$ is more convenient here [see also (7a) and (21a)].

Notice that the difference between continuous and discrete time is due to the required distinction in discrete time between i and $i-1$; if these could be identified, we would have, instead of (47),

$$A(i)G(i)G'(i)A'(i) = 0 \quad (50)$$

which would imply

$$A(i)G(i) = 0 \quad (51)$$

just as in continuous time.

Next we show that a process whose covariance has definite relative order α can always be represented as the output of a state-variable model that does obey the constraints (49).

Theorem 5: Let the covariance (4) of a process $y(\cdot)$ over an interval $[0,N]$ have definite relative order α, $\alpha > 0$. Then there exists at least one representation of the form (43) that obeys the constraints (49).

Proof: We notice first that the matrix $H(i)$ of the realization (43) is determined by (46a). Hence the set of all realizations (43) for the process $y(\cdot)$ is determined by the set of all solutions $\{\Pi(i),G(i+1),\ i = 0,1,\cdots,N\}$ to the equations (44) and (46b) where $A(\cdot)$ and $B(\cdot)$ obey the constraints (42). We show that there exists a subset of solutions $\{\Pi(\cdot),G(\cdot)\}$ that obey the constraints (49). All we need to prove is that we can choose $\Pi(\cdot)$ and $G(\cdot)$ to obey (49) without violating the constraints (44), (46b), and (42). To show this, we choose $G(\cdot)$ and $\Pi(\cdot)$ such that

$$A(i-1)G(i) = 0 \qquad 1 \le i \le N \quad (52)$$

$$A(i-1)\Pi(i) = B'(i-1), \qquad 1 \le i \le N. \quad (53)$$

We show that this choice is consistent with the constraints on $G(\cdot)$ and $\Pi(\cdot)$. Multiplying (44) to the left by $A(i-1)$, using (52) and (53), we get

$$0 = A(i-1)G(i)G'(i) = A(i-1)\Pi(i) - A(i-1)\Pi(i-1) = B'(i-1) - A(i-1)\Pi(i-1). \quad (54)$$

Thus (46b) follows, and post-multiplication by $A'(i)$ shows that (42a) follows similarly. If $\alpha > 2$, the same procedure can be pursued to show that the constraints (43) on $G(\cdot)$ and $\Pi(\cdot)$ are consistent with the conditions (42), (44), and (46b). ∎

We have established that although not all representations of a process with definite relative order α obey all α constraints (49), there exists a class of models that does obey these constraints. Before identifying such models, it will be useful to restate the relative order conditions (42) and the constraints (49) for a process whose covariance is given by (13) instead of (4). A covariance of the form (13) has relative order α if

1) $$M(i-k)\phi(i-k,i)N(i) - N'(i-k)\phi'(i,i-k)M'(i) = 0, \qquad k = 1,2,\cdots,\alpha-1;\ k \le i \le N \quad (55a)$$

2) $M(i-\alpha)\phi(i-\alpha,i)N(i) - N'(i-\alpha)\phi'(i,i-\alpha)M'(i) = C(i)C'(i) > 0, \quad \alpha \le i \le N. \quad (55b)$

The constraints (49) for a model of the form (7) are

$$H(i-k)\phi(i-k,i)G(i) = 0, \quad k = 1,\cdots,\alpha-1 \quad (56a)$$

$$H(i-k)\phi(i-k,i)x(i) = y(i-k), \quad k = 0,\cdots,\alpha-1 \quad (56b)$$

$$H(i-k)\phi(i-k,i)\Pi(i) = N'(i-k)\phi'(i,i-k), \quad k = 0,\cdots,\alpha-1 \quad k \le i \le N. \quad (56c)$$

Corollary 3: Let a process $y(\cdot)$ with covariance (13) have definite relative order α, $\alpha > 0$. Then the innovations representation IR-2 of this process has definite relative order α, i.e., it obeys the constraints (56a)–(56c).

Proof: The proof is an easy but lengthy verification that the equations (56a)–(56c) hold for the model (21) with $G(\cdot)$ replaced by $K(\cdot)$, $x(\cdot)$ by $\theta(\cdot)$, $\Pi(\cdot)$ by $\Sigma_2(\cdot)$, and $H(\cdot)$ by $M(\cdot)$. ■

As we noticed in Section III, IR-1 is not of the form (7). It does not obey the constraints (55), but obeys the same number of equivalent constraints.

Corollary 4: Let a process $y(\cdot)$ with covariance (13) have definite relative order α, $\alpha > 0$. Then the innovations representation IR-1 of this process obeys the constraints

$$M(i-k)\phi(i-k,i)K(i) = 0, \quad k = 1,\cdots,\alpha-1 \quad (57a)$$

$$M(i-k)\phi(i-k,i)\theta(i) = y(i-k), \quad k = 1,\cdots,\alpha \quad (57b)$$

$$M(i-k)\phi(i-k,i)\Sigma_1(i) = N'(i-k)\phi'(i,i-k), \quad k = 1,\cdots,\alpha \quad (57c)$$

$$k \le i \le N.$$

Proof: The proof is again a straightforward verification. ■

VI. Reduced Order Model for a Scalar Process

We shall show now how the constraints (56) can be used to obtain a reduced order Riccati equation for IR-2 [cf. (21f)]. This will require a coordinate transformation. For simplicity of notation, we shall consider a scalar process and we shall assume a "uniform observability" assumption [defined below, after (60)].

Thus consider a scalar process $y(\cdot)$ over an interval $[0,N]$ with covariance

$$R_y(i,j) = m(i)\phi(i,j)N(j)1(i-j) + N'(i)\phi'(j,i)m'(j)1(j-i-1) \quad (58)$$

where $m(\cdot)$ is $1 \times n$, $\phi(\cdot,\cdot)$ is $n \times n$ and nonsingular, and $N(\cdot)$ is $n \times 1$.

Let the covariance (58) have definite relative order α, $\alpha > 0$. Then we know by Corollary 3 that the representation IR-2 given by (21), with $M(\cdot)$ replaced by $m(\cdot)$, obeys the constraints (56). To exploit the constraints (56c) on $\Sigma_2(\cdot)$ and thereby reduce the order of the Riccati equation (21f), we shall make the state transformation

$$x_{\text{new}}(i) = T(i)x_{\text{old}}(i) \quad (59)$$

with

$$T(i) = \begin{bmatrix} m(i) \\ m(i-1)\phi(i-1,i) \\ \cdot \\ \cdot \\ \cdot \\ m(i-n+1)\phi(i-n+1,i) \end{bmatrix}. \quad (60)$$

The assumption of uniform observability is that $T(i)$ is uniformly nonsingular, i.e., nonsingular for $n-1 \le i \le N$.[5] In the stationary case this assumption is the same as complete observability.

In the new coordinate system we have

$$m(i) = [1 \quad 0 \quad \cdots \quad 0] \quad (61a)$$

$$\phi(i+1,i) = \begin{bmatrix} & & a(i+1) & & \\ \hline 1 & 0 & & \cdots & 0 \\ 0 & 1 & 0 & & \\ \cdot & & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ 0 & \cdots & & 0 \quad 1 & 0 \end{bmatrix} \quad (61b)$$

where $a(i+1)$ is $1 \times n$:

$$a(i+1) = [a^1(i+1) \cdots a^n(i+1)]. \quad (61c)$$

Now let $N^j(i)$ be the jth component of the vector $N(i)$ in the new coordinate system. Then define

$$M(i) = \begin{bmatrix} N^1(i) & N^2(i) & \cdots & N^n(i) \\ N^2(i) & N^1(i-1) & N^2(i-1) & N^{n-1}(i-1) \\ \cdot & & N^1(i-2) & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ & & \cdot & \\ N^n(i) & & \cdots & N^1(i-n+1) \end{bmatrix}$$

$$\alpha - 1 \le i \le N \quad (62)$$

with $N^j(i) = 0$ for $i < 0$ and for any j. Let the $(n-\alpha) \times (n-\alpha)$ lower right-hand submatrix of $\phi(i+1,i)$ be denoted by

$$F \triangleq \begin{bmatrix} 0 & & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & & \cdot & \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & & \cdot \\ 0 & \cdots & 0 \quad 1 & 0 \end{bmatrix}. \quad (63)$$

[5] This assumption can easily be avoided. However, without this assumption, the simple form (61) cannot be obtained (cf. the analogous situation in continuous time [2]).

Finally, let the last $(n - \alpha)$ elements of $a(i+1)$ be denoted by

$$b(i+1) \triangleq [a^{\alpha+1}(i+1) \cdots a^{n}(i+1)]. \tag{64}$$

With these notations we can describe an $(n - \alpha)$th order Riccati equation for the innovations representation (IR-2) of a process with definite relative order α.

Theorem 6: Let a scalar process $y(\cdot)$ over $[0,N]$ have covariance (58) with definite relative order α and let $\Sigma_2(\cdot)$ be the solution of the Riccati equation (21f) of IR-2. Then for $i \geq \alpha - 1$, $\Sigma_2(i)$ can be partitioned into

$$\Sigma_2(i) = M(i) + D(i), \qquad \alpha - 1 \leq i \leq N \tag{65}$$

with

$$D(i) = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & E(i) \end{array}\right] \tag{66}$$

where $E(i)$ is $(n - \alpha) \times (n - \alpha)$ and obeys the equation

$$E(i+1) = FE(i)F' + FE(i)b'(i+1)[N^1(i+1) - a(i+1)M(i)a'(i+1) - b(i+1)E(i)b'(i+1)]^{-1} \cdot b(i+1)E(i)F', \qquad \alpha - 1 \leq i \leq N \tag{67a}$$

$$E(\alpha - 1) = \text{lower right-hand submatrix of } \Sigma_2(\alpha - 1) - \text{lower right-hand submatrix of } M(\alpha - 1). \tag{67b}$$

Proof: That the variance of the states of IR-2 can be written in the form (65) and (66) follows after lengthy algebraic manipulations from the constraints (49c) on $\Sigma_2(\cdot)$ and the special form that these constraints take in the new coordinate system. The expressions (67) follow by substitution of (65) and (66) in (21f). More details can be found in [22]. ■

Corollary 5: Let a scalar process $y(\cdot)$ over $[0,N]$ have covariance (58) with definite relative order α. Then for all $i \geq \alpha - 1$, IR-2 is state equivalent to

$$\left[\; \theta(i+1) \;\right] = \left[\begin{array}{ccccc} & & a(i+1) & & \\ \hline 1 & 0 & & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdot & & & \cdot & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & \cdots & 0 & 1 & 0 \end{array}\right] \cdot \left[\; \theta(i) \;\right] + \left[\begin{array}{c} 1 \\ \hline 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \hline \hat{K}(i+1) \end{array}\right] \mathcal{E}(i+1) \tag{68a}$$

$$y(i) = [1 \; 0 \; \cdots \; 0][\theta(i)] \tag{68b}$$

where $\mathcal{E}(\cdot)$ is a white-noise process with

$$E[\mathcal{E}(i)\mathcal{E}(j)] = [N^1(i) - a(i)M(i-1)a'(i) - b(i)E(i-1)b'(i)]\delta(i-j). \tag{68c}$$

$\hat{K}(i)$ is $(n - \alpha) \times 1$ and

$$\hat{K}(i) = FE(i-1)b'(i)[N^1(i) - a(i)M(i-1)a'(i) - b(i)E(i-1)b'(i)]^{-1} \tag{68d}$$

where $E(i)$ obeys (67).

Proof: From (21d) it follows directly that $m(i)K(i) = 1$. From the form of $T(i)$ [see (60)] it follows that the first component of $K(i)$ in the new coordinate space is 1. The zeros in $K(i)$ follow from the constraints (49a) with $G(\cdot)$ replaced by $K(\cdot)$. The other relations follow from Theorem 6.

Remark 1: From the form of F in (63) we notice that the first component of $\hat{K}(\cdot)$ is zero. This, together with (68a), shows that the realization has a pure delay line of order α and a feedback structure of order $n - \alpha$.

Remark 2: We have obtained a reduced order Riccati equation for IR-2 when the process has relative order α. Naturally, since IR-1 also obeys α constraints [see (57)], a similar reduction can be obtained for it. Some results along these lines were already obtained in [21], patterned on analogs of the continuous-time results of [2] and [3].

VII. Example

We shall study a second-order scalar process whose covariance has definite relative order 2. We shall give two different realizations, of which only one is causally invertible and obeys the constraints (56). For this last realization we shall see that the Riccati equation (21f) is totally degenerate. A similar example is studied in continuous time in [2].

We consider a scalar process $y(\cdot)$ over $[0,\infty)$ whose covariance matrix R is given by

$$R = [(R_{ij})] = \begin{bmatrix} 16 & 1 & 4 & \frac{1}{4} & \cdots \\ 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \cdots \\ 4 & \frac{1}{4} & \frac{5}{4} & \frac{1}{16} & \cdots \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \frac{9}{32} & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \end{bmatrix}. \tag{69}$$

This covariance can be factored into

$$R(i,j) = h\phi^{i-j}N(j)1(i-j) + N'(i)\phi'^{j-i}h'1(j-i-1) \tag{70}$$

where h is a row vector with entries 0 and 1, ϕ is a 2×2 matrix with zeros on the main diagonal and $\frac{1}{2}$ elsewhere, and $N(i)$ is given by the recurrence relations

$$N^1(i+1) = \tfrac{1}{4}N^1(i)$$
$$N^2(i+1) = \tfrac{1}{16}N^2(i-1) + \tfrac{1}{4}, \qquad i \geq 1 \tag{71a}$$

together with the initial values

$$N'(0) = [2 \quad 16], \quad N'(1) = [\tfrac{1}{2} \quad \tfrac{1}{2}]. \tag{71b}$$

Notice that

$$\lim_{i\to\infty} N'(i) = [0 \quad \tfrac{1}{4}]. \tag{72}$$

It is easy to verify that h, ϕ, and $N(\cdot)$ do indeed produce the covariance $R(i,j)$ given in (69). Using (55) we check that the relative order of the covariance is 2:

$$h\phi^{-1}N(i) - N'(i-1)\phi'h' \equiv 0, \qquad i \geq 1 \quad (73a)$$

$$h\phi^{-2}N(i) - N'(i-2)\phi'^2h' = \tfrac{1}{4} > 0, \qquad i \geq 2. \quad (73b)$$

Thus we know by Corollary 3 that the projection of $\Sigma_2(\cdot)$, the state variance of IR 2, on the vectors $h\phi^{-1}$ and $h\phi^{-2}$ can be computed directly from $N(\cdot)$ and ϕ [see (56c)]. But $h\phi^{-1}$ and $h\phi^{-2}$ are linearly independent:

$$h\phi^{-1} = [2 \quad 0]; \quad h\phi^{-2} = [0 \quad 4].$$

Hence $\Sigma_2(\cdot)$ can be completely determined without solving the Riccati equation (21f), as we shall see. But first we show that there exist other state-variable models of the form (7) that can yield the same covariance.

A first realization is

$$x(i+1) = \phi x(i) + \Gamma\nu(i+1) \quad (74a)$$

$$y(i) = [0 \quad 1]\, x(i) \quad (74b)$$

where

$$\Gamma' = [1 \quad 0] \text{ and } \phi = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$$

$$E[\nu(i)\nu(j)] = \delta(i-j),\; E[x_0x_0'] \triangleq \Pi_0 = \begin{bmatrix} 2 & 2 \\ 2 & 16 \end{bmatrix}. \quad (74c)$$

$\Pi(\cdot)$ can be computed by solving the state-variance equation

$$\Pi(i+1) = \phi\Pi(i)\phi' + GG', \qquad \Pi(0) = \Pi_0. \quad (75)$$

It is easy to check that

$$\Pi(i)h' = N(i), \qquad i = 0,1,\cdots. \quad (76)$$

The relation (73a) is, of course, satisfied, but notice that none of the constraints (56) holds. The input–output relation is

$$y(i) = \tfrac{1}{4}y(i-2) + \tfrac{1}{2}\nu(i-1), \quad (77)$$

or in transfer function form

$$H(z) = 2z^{-1}/4 - z^{-2}. \quad (78)$$

Notice that while both poles of the transfer function are inside the unit circle, the only zero is at infinity. It follows, as is otherwise easy to see, that this system is not causally invertible (or minimum phase). The power spectral density of the output, in the steady state, is

$$R(z) = H(z)H(z^{-1}) = 4[17 - 4(z^2 + z^{-2})]^{-1}. \quad (79)$$

Notice that an alternate factorization of $R(z)$ gives

$$H(z) = 2/4 - z^{-2}. \quad (80)$$

The fact that different factorizations of a given power spectral density are possible in discrete time and the structural properties of these different realizations is studied in [16].

For the same covariance (69) IR-2 is

$$x(i+1) = \phi x(i) + K(i+1)\mathcal{E}(i+1) \quad (81a)$$

$$y(i) = [0 \quad 1]\, x(i) \quad (81b)$$

with

$$E[\mathcal{E}(i)\mathcal{E}(j)] = [hN(i) - h\phi\Sigma_2(i-1)\phi'h'] \quad (81c)$$

$$\Sigma_2(0) = N(0)[hN(0)]^{-1}N'(0) = \begin{bmatrix} \frac{1}{4} & 2 \\ 2 & 16 \end{bmatrix} \quad (81d)$$

$$x(0) = K(0)\mathcal{E}(0) = \begin{bmatrix} \frac{1}{8} \\ 1 \end{bmatrix}\mathcal{E}_0.$$

The expressions of $\Sigma_2(0)$ and $x(0)$ follow from (21f) and (21a). The gains $K(i)$ are computed from (21d). But notice that for $i \geq 1$, $\Sigma_2(i)$ can be calculated directly by its known projections on h and $h\phi^{-1}$, using the constraints (56c):

$$h\Sigma_2(i) = [0 \quad 1]\Sigma_2(i) = N'(i), \qquad i \geq 0 \quad (82a)$$

$$h\phi^{-1}\Sigma_2(i) = [2 \quad 0]\Sigma_2(i) = N'(i-1)\phi', \qquad i \geq 1. \quad (82b)$$

The Riccati equation (21f) need not be solved. Notice finally that this last model is invertible (the inverse is obtained by Corollary 2), and that the constraints (56a) and (56b) also hold. In particular,

$$hx(i) = y(i);\; h\phi^{-1}x(i) = y(i-1). \quad (83)$$

VIII. Applications to Least-Squares Estimation

In Part V [2] of this series of papers we showed how the IR's and the interpretations thereof were helpful in providing solutions, almost by inspection, to a number of filtering and smoothing problems, including problems with covariance information rather than the more common state-model information. Similar discrete-time applications can be made here, and we outline them very briefly. They exploit once again the intimate relation that exists between least-squares estimation and the IR's.

Suppose we are given observations of a lumped process $y(\cdot)$ with a positive-definite separable covariance function

$$R_y(i,j) = M(i)\phi(i,j)N'(j), \qquad i \geq j \quad (84a)$$

$$= N'(i)\phi'(i,j)M'(j), \qquad i < j. \quad (84b)$$

Since we know that $R_y(i,j)$ is the covariance of some lumped process, we shall assume that we have the model

$$x(i+1) = \phi(i+1,i)x(i) + G(i+1)u(i+1) \quad (85a)$$

$$y(i) = M(i)x(i) \quad (85b)$$

with the usual assumptions on x_0 and $u(\cdot)$. To be consistent with (84), the state variance $\Pi(\cdot)$ has to obey the constraint

$$\Pi(i)M'(i) = N(i). \quad (86)$$

We wish to find the least-squares filtered and smoothed estimates of a related signal process $w(\cdot)$ for which the following assumption holds:

$$w(i) = M_w(i)x(i) + y^{\perp}(i) \quad (87)$$

where $y^{\perp}(\cdot)$ is any process uncorrelated with $y(\cdot)$. It is easy to see that this assumption is equivalent to the following cross-covariance information about $w(\cdot)$:

$$E[w(i)y'(j)] = M_w(i)\phi(i,j)N(j), \qquad i \geq j \quad (88a)$$

$$= N_w'(i)\phi'(j,i)M'(j), \qquad i < j \quad (88b)$$

with the constraint

$$\Pi(i)M_w'(i) = N_w'(i). \tag{89}$$

We first show that the filtered estimate $\hat{w}(i|i)$ of $w(\cdot)$ can be expressed as a known linear combination of the state $\mathcal{C}(i)$ of the IR-2 of the process $y(\cdot)$:

$$\hat{w}(i|i) = M_w(i)\mathcal{C}(i). \tag{90}$$

To show this, we observe that (87) implies

$$\hat{w}(i|i) = M_w(i)\hat{x}(i|i). \tag{91}$$

But as noted earlier (cf. Corollary 2 and the proof of Theorem 2), $\hat{x}(i|i)$ is just the state vector of the whitening filter (26) associated with IR-2, and as shown in Theorem 2, it can be expressed entirely in terms of the parameters $M(\cdot)$, $\phi(\cdot,\cdot)$, and $N(\cdot)$ of the covariance function $R_y(i,j)$.

It follows similarly from (87) that

$$\hat{w}(i|i-1) = M_w(i)\hat{x}(i|i-1) \tag{92}$$

where $\hat{x}(i|i-1)$ is the state vector of IR-1 (cf. Corollary 1 and the proof of Theorem 1).

For the smoothed estimate, we have, again using (87),

$$\hat{w}(i|j) = M_w(i)\hat{x}(i|j). \tag{93}$$

The smoothed estimate $\hat{x}(i|j)$ can be calculated by using a general innovations formula (cf. Part II of this series [23]):

$$\hat{x}(i|j) = \hat{x}(i|i) + \sum_{k=i+1}^{j} E[x(i)\varepsilon'(k)]E[\varepsilon(k)\varepsilon'(k)]^{-1}\varepsilon(k). \tag{94}$$

By the orthogonality property of the innovations, we have

$$E[x(i)\varepsilon'(k)] = E[\tilde{x}(i|i-1)\tilde{x}'(k|k-1)]M'(k)$$

$$= P(i,k)M'(k), \qquad i < k \tag{95}$$

where

$$P(i,k) \triangleq E[\tilde{x}(i|i-1)\tilde{x}'(k|k-1)]. \tag{96}$$

It is easy to calculate that

$$P(i,k) = P(i|i-1)\psi'(k,i) \tag{97}$$

where $P(i|i-1)$ is defined by (16e) and $\psi(k,i)$ is the state-transition matrix of the error equation

$$\tilde{x}(i+1|i) = \phi(i+1,i)[I - K(i)M(i)]\tilde{x}(i|i-1) + G(i+1)u(i+1)$$

$$= \psi(i+1,i)\tilde{x}(i|i-1) + G(i+1)u(i+1). \tag{98}$$

With these notations we can now write

$$\hat{x}(i|j) = \hat{x}(i|i) + P(i|i-1) \sum_{k=i+1}^{j} \psi'(k,i)M'(k) \cdot [M(k)P(k|k-1)M'(k)]^{-1}\varepsilon(k)$$

$$= \hat{x}(i|i) + P(i|i-1)\lambda(i), \text{ say} \tag{99}$$

where $\lambda(i)$ is the so-called adjoint variable

$$\lambda(i) = \sum_{k=i+1}^{j} \psi'(k,i)M'(k)[M(k)P(k|k-1)M'(k)]^{-1}\varepsilon(k)$$

$$= \psi'(i+1,i)\lambda(i+1) + \psi'(i+1)M'(i+1) \cdot [M(i+1)N(i+1) - M(i+1)\Sigma_1(i+1) \cdot M'(i+1)]^{-1}\varepsilon(i+1) \tag{100a}$$

$$\lambda(j) = 0. \tag{100b}$$

Substituting (99) into (93) yields, through (18) and (89),

$$\hat{w}(i|j) = \hat{w}(i|i) + M_w(i)P(i|i-1)\lambda(i)$$

$$= \hat{w}(i|i) + [N_w'(i) - M_w(i)\Sigma_1(i)]\lambda(i) \tag{101}$$

where $\Sigma_1(i)$ is the solution of (15f). To complete the rewriting of the smoothing formula for $\hat{w}(i|j)$ in terms of the parameters of the covariance function, notice that, by (98) and (15d),

$$\psi(i+1,i) = \phi(i+1,i)\{I - [N(i) - \Sigma_1(i)M'(i)] \cdot [M(i)N(i) - M(i)\Sigma_1(i)M'(i)]^{-1}M(i)\}. \tag{102}$$

The equations (100)–(102), together with the recursive equation for $\Sigma_1(i)$ and the relation (90) for $\hat{w}(i|i)$, constitute a complete solution of the smoothing problem for $w(\cdot)$ in terms of the parameters of the covariance function (88). The formula (101) is the analog of the Bryson–Frazier form for the fixed-interval smoothing estimate. The continuous-time version of (101) was first published in [2] [see, e.g., (137)].

IX. Concluding Remarks

We have shown how to find innovations representations for discrete-time observation processes with specified separable covariance. From these representations, expressions can be derived, almost by inspection, for the filtered and smoothed estimates of a related process for which either model or covariance information is given.

Our results show that there exists an important difference between the continuous- and discrete-time realization problems in the case where the covariance has a nonzero definite relative order. Whereas in continuous time the relative order is a unique property of both its covariance function and any of its realizations, in discrete time the relative order of the covariance of a process and a lumped realization of this process are not necessarily the same. This means that whether or not the realization inherits the constraints imposed by the relative order property on the covariance depends on how the factorization is performed. In [16] and [22] we have shown how these facts explain certain phenomena [17],[18] in estimation problems for given discrete-time models.

Another important reason for studying the representation problem is as a prelude to system *identification*, as opposed to system *realization*, problems. The distinction is that in realization problems knowledge of the true covariance is assumed, while in the identification problem only a finite observed record $\{y(i), 0 \leq i \leq N\}$ is available. One procedure is to assume stationarity, use the record to estimate the covariance, and then apply the algorithms of this paper. However, perhaps because of the difficulty in obtaining good covariance estimates, most attention seems to have been paid to maximum likelihood (ML) techniques that work directly with the given data record. This may very well be the best procedure (and we may note incidentally that use of the IR gives a computationally convenient form of the ML equations [24]), but we do not believe that this fact has been conclusively demonstrated. In fact, our result that not all the inherent smoothness of a covariance function need be present in all models suggests that the operation of forming a covariance may smooth our irrelevant roughness in a data record. This whole question is a topic of continuing inquiry.

References

[1] T. Kailath and R. A. Geesey, "An innovations approach to least-squares estimation—Part IV: Recursive estimation given lumped covariance functions," *IEEE Trans. Automat. Contr.* (Special Issue on Linear-Quadratic-Gaussian Problem), vol. AC-16, pp. 720-727, Dec. 1971.

[2] ——, "An innovations approach to least-squares estimation—Part V: Innovations representations and recursive estimation in colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 435-453, Oct. 1973.

[3] R. Geesey, "Canonical representations of second-order processes with applications," Ph.D. dissertation, Dep. Elec. Eng., Stanford Univ., Stanford, Calif., Dec. 1968; also SEL Tech. Rep. 7050-17, June 1969.

[4] B. D. O. Anderson and J. B. Moore, "The Kalman-Bucy filter as a true time-varying Wiener filter," *IEEE Trans. Syst., Man Cybern.*, vol. SMC-1, pp. 119-128, Apr. 1971.

[5] B. D. O. Anderson, J. B. Moore, and S. G. Loo, "Spectral factorization of time-varying covariance functions," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 550-557, Sept. 1969.

[6] J. B. Moore and B. D. O. Anderson, "Spectral factorization of time-varying covariance functions: The singular case," *Math. Syst. Theory*, vol. 4, no. 1, pp. 10-23, 1970.

[7] L. H. Brandenburg, "Shaping filter models for nonstationary random processes," Ph.D. dissertation. Dep. Elec. Eng., Columbia Univ., New York, N.Y., June 1968.

[8] T. Kailath, "Fredholm resolvents, Wiener-Hopf equations, and Riccati differential equations," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 665-672, Nov. 1969.

[9] R. Geesey and T. Kailath, "Applications of the canonical representation to the estimation and detection in colored noise," in *Proc. Symp. Comput. Processing on Communications.* Brooklyn, N.Y.: Polytechnic Inst. Brooklyn Press, 1970, pp. 435-457.

[10] T. Kailath, "The innovations approach to detection and estimation theory," *Proc. IEEE*, vol. 58, pp. 680-695, May 1970.

[11] P. Lévy, *Brownian Movement and Stochastic Processes.* Paris: Gauthier-Villars, 1964, Appendix II.

[12] T. Hida, "Canonical representations of Gaussian processes and their applications," *Mem. Coll. Sci. Kyoto Univ.*, ser. A, vol. 33, pp. 109-155, 1960.

[13] H. Cramér, "A contribution to the multiplicity theory of stochastic processes," in *Proc. 5th Symp. Math. Statist. and Probability*, vol. 2. Berkeley, Calif.: Univ. California Press, 1967, pp. 215-221.

[14] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME* (J. Basic Eng.), ser. D, vol. 83, pp. 95-107, Dec. 1961.

[15] T. Kailath, "Seminar on discrete-time stochastic realization," Stanford Univ., Stanford, Calif., Feb. 1971.

[16] M. Gevers and T. Kailath, "Constant, predictable and degenerate directions of discrete-time Riccati equations," *Automatica*, Nov. 1973.

[17] R. S. Bucy, D. Rappaport, and L. M. Silverman, "Correlated noise filtering and invariant directions for the Riccati equation," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 535-540, Oct. 1970. See also D. Rappaport, "Constant directions of the Riccati equation," Ph.D. dissertation, Dep. Elec. Eng., Univ. Southern California, Los Angeles, Aug. 1969.

[18] D. Rappaport, "Constant directions of the Riccati equation," *Automatica*, vol. 8, pp. 175-186, Mar. 1972.

[19] P. Faurre, "Identification par minimisation d'une représentation Markovienne de processus aléatoire," Centre d'Automatique de l'Ecole National Supérieure des Mines de Paris, Ref. E. 21, July 1969; also *Lecture Notes in Mathematics* (from Nice Symp. Optimization), vol. 132. New York: Springer, pp. 85-106, 1970.

[20] L. H. Son and B. D. O. Anderson, "Design of Kalman filters using signal model output statistics," submitted for publication.

[21] R. Geesey, "Notes on discrete-time innovations representations," lecture notes, Univ. Washington, St. Louis, Mo., 1970.

[22] M. Gevers, "Structural properties of realizations of discrete-time Markovian processes," Ph.D. dissertation, Stanford Univ., Stanford, Calif., May 1972; also Stanford Univ., SEL Tech. Rep. 7050-19.

[23] T. Kailath and P. Frost, "An innovations approach to least-squares estimation—Part II: Linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 655-660, Dec. 1968.

[24] R. K. Mehra, "Identification of stochastic linear dynamic systems," *AIAA J.*, Jan. 1971.

Michel R. Gevers (S'67-M'72) was born in Antwerp, Belgium, on June 29, 1945. He received the M.S. degree in electrical engineering from the Catholic University of Louvain, Louvain, Belgium, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif., in 1968 and 1972, respectively. In 1969 he was awarded a Harkness Fellowship, and in 1971 an ESRO/NASA International Fellowship. He was also a fellow of the Belgian American Educational Foundation.

In 1972 he joined the Laboratoire d'Automatique et Dynamique des Systemes, Catholic University of Louvain, Louvain-la-Neuve, Belgium. His current research interests are in estimation theory, system identification and adaptive systems, and the applications of these techniques to ecological systems.

Thomas Kailath (S'57-M'62-F'70), for a photograph and biography, see page 453 of the October 1973 issue of this TRANSACTIONS.