IEEE CDC Dec. 1973, pp168-171

C3

CDC
PAPER No. WP2-6

THE DISCRETE-TIME STOCHASTIC REALIZATION PROBLEM: MINIMUM VARIANCE PROPERTY OF THE INNOVATIONS REPRESENTATION

M. Gevers
University of Louvain, Belgium
and
T. Kailath
Stanford University

Abstract

The linear stochastic realization problem for a time-varying process with a smooth separable covariance is briefly described. It is shown that finding all Markovian realizations of the process is equivalent with finding all solutions to a set of constraints on the state-variances. Introducing a partial ordering on this set of nonnegative definite solutions (viz., $\pi_1 \geq \pi_2$ if $\pi_1 - \pi_2$ is nonnegative definite) it is shown that the smallest solution, obtained with the help of a matrix minimality property, is the unique causal and causally invertible Markovian representation. A stochastic interpretation is given based on the fact that the state of the IR is the filtered estimate of the state of any other model.

I. Introduction

The linear stochastic realization problem, also called covariance factorization problem, is to find a white-noise driven finite-dimensional linear system whose output generates a specified separable covariance. Whereas deterministic realization theory shows that all minimal-dimensional realizations are essentially equivalent, it is well-known that in the stochastic case there exist essentially distinct minimal-dimensional realizations. This problem has received a great deal of attention in the last few years, and several authors have presented solutions to various aspects of the time-varying covariance factorization problem for continuous-time processes [1] - [6].

First to be solved was the class of continuous-time problems in which the 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 [3] and Brandenburg [5] showed that the class of linear finite-dimensional systems that generate the covariance R are determined by the class of solutions of a certain Riccati equation. In independant work Kailath and Geesey [1] obtained a unique solution (up to impulse response) by requiring that the linear system be an innovations representation (IR) (i.e. a causal and causally invertible model) and using the general property that the IR of a process $y(.)$ depends only upon the covariance of $y(.)$. They obtained a Wiener-Hopf equation, and for separable K, a Riccati equation whose solution determined the IR.

The case in which the process does not explicitly contain white noise, namely where its covariance R is smooth, is more difficult, at least in continuous time. A solution to this problem, has been obtained by first differentiating the process $y(.)$ a number of times, say α times, until the differentiated process contains a white noise component. The covariance of the α-th derivative of the process is then of the form I + K. The number α was called the definite relative order of the process. For this singular case and for separable K, Moore and Anderson [4] and Brandenburg [5] once again obtained a class of systems that are determined by the set of solutions of a Riccati equation, while Geesey and Kailath [2] obtained a unique innovations representation.

Many of the continuous-time results carry over unchanged to the discrete-time case, with derivatives replaced by differences. Actually the case where the covariance has the form I + K has its exact discrete-time counterpart [7]. The case of a smooth covariance brings out some important differences with the continuous-time case, however, in that no differencing is needed here to obtain a model. In addition, unlike the continuous time case, the definite relative order of the covariance does not necessarily induce constraints on the states of the realizations (see [8], [9]). In [8] a solution has been given to the discrete-time factorization problem for a smooth covariance, in the form of a causal and causally invertible IR obtained via the solution of a Riccati equation.

It follows from the preceding remarks that the class of all finite-dimensional realizations that generate some separable covariance are determined by the set of all nonnegative definite solutions π to a certain matrix Riccati equation and that the IR is a particular member of this class, π being the state-variance of the model. Various authors [10] - [14] have investigated this set of solutions of the Riccati equation for the stationary case, and they have shown that, when solutions exist (viz., when the covariance is positive definite) there exists a largest solution π_{max} and a smallest solution π_{min}, and that all other solutions π are such that $\pi_{min} \leq \pi \leq \pi_{max}$, where $A-B \geq 0$ means A-B is nonnegative definite.

In this paper we shall treat the case of a discrete time smooth time-varying covariance function and we shall show that the realization with the smallest state-variance is the IR. In section II we describe the set of constraints that the state-variance $\pi(.)$ of a model must satisfy in order to generate the specified covariance. In Section III we use a matrix inequality to find the smallest solution $\Sigma(.)$ of these equations, and we show that the corresponding model is exactly the IR obtained via an alternate route in [8]. This result is given a stochastic interpretation in Section IV based on the fact that the state of the IR is the filtered estimate of the state of any other realization.

II. The stochastic realization problem

We shall consider a p-vector valued discrete-time zero-mean Gaussian process $y(i)$, $i = 0, 1, \ldots, N$, with N possibly infinite, and with a separable covariance of the form

$$R(i,j) = M(i)F(i,j)N(j) \quad , i \geq j \qquad (2a)$$

$$= N'(i)F'(j,i)M'(j), \; i < j \qquad (2b)$$

where $F(i,j)$ has the composition property (i.e. $F(i,j) = F(i,k)F(k,j)$ for $i \geq k \geq j$) and has dimension $n \times n$. We shall assume that the covariance $R(i,j)$ is strictly positive definite for $0 \leq i, j \leq N$.

The stochastic realization problem can then be formulated as follows : "Given the parameters $M(i)$, $F(i,j)$, $N(j)$ of the signal covariance for all i,j in $[0, N]$, find a Markovian representation for the process $y(i)$, i.e. find a triplet $H(i)$, $F(i)$, $G(i)$ and the appropriate initial conditions x_0 and $E\{x_0 x_0\}' \triangleq \pi_0$ such that

$$y(i) = H(i)x(i) \qquad (3a)$$

$$x(i) = F(i-1)x(i-1) + G(i)u(i) \qquad (3b)$$

where $u(i)$ is white Gaussian noise of unit variance, uncorrelated with x_0".

A unique (up to the impulse response) solution to this problem has been given in [8], with the existence guaranteed by the positive definite nature of the covariance function $R(.,.)$. Here we shall characterize the set of all solutions.

To do so we first describe the relations that must exist between the parameters of a model (3) and the covariance (2). This is easily done by identifying the output covariance of the model (3) with the given covariance (2), which leads to the following obvious (but nonunique) identifications.

$$M(i) = H(i), \; F(i,j) = F(i-1) \ldots F(j) \qquad (4a)$$

$$N(i) = \pi(i)M'(i) \qquad 0 \leq i,j \leq N \qquad (4b)$$

$\pi(.)$ is the variance of the states of the model and obeys the difference equation.

$$\pi(i) = F(i-1)\,\pi(i-1)F'(i-1) + G(i)G'(i) \; ; \quad 0 \leq i \leq N \qquad (4c)$$

Since knowledge of $R(.,.)$ implies knowledge of $M(.)$ and $F(.,.)$ - at least to within a nonsigular transformation - all Markovian models are determined (up to a state-transformation) by the set of all solutions $\pi(.)$, $G(.)$ to the constraints (4b,c). Clearly the only way models can differ is in the matrix $G(.)$ and the initial conditions.

In the stationary case, Faurre [13] has approached the realization problem from this point of view, namely "Find all nonnegative definite matrices π and Q that satisfy the constraints

$$\pi = F\pi F' + Q \qquad , \pi M' = N \qquad (5)$$

He showed that the set of solutions for π is convex, and he obtained a smallest and a largest π as the asymptotic value of a Riccati equation, that resulted from an associated optimal control problem.

In the next section we shall obtain the smallest time-varying $\pi(.)$ that obeys the equations (4b,c) by directly solving these equations for their smallest solution $\Sigma(.)$. We shall show that $\Sigma(.)$ is the state-variance of the IR obtained in [8].

III. Minimum Variance Realization

We shall say that $\Sigma(.)$ is the smallest solution of the constraints (4b,c) if for any other solution $\pi(.)$ we have

$$\pi(i) - \Sigma(i) \geq 0 \quad \text{for all i in } [0,N] \qquad (6)$$

We now show how to obtain $\Sigma(.)$.
Our proof follows a completely different line of argument than Faurre's, and relies heavily on the following well-known lemma.

Lemma
Let A and B be two $p \times n$ matrices with $p \leq n$ and let AB' be nonsingular. Then the smallest $n \times n$ symmetric nonnegative definite matrix Q that satisfies $AQ = B$ is

$$Q_0 = B'(AB')^{-1} B \qquad (7)$$

in the sense that for any other solution Q we have

$$Q - Q_0 \geq 0 \qquad (8)$$

Proof : $Q - Q_0 =$

$$[I-QA'(AQA')^{-1}A]\,Q\,[I-QA'(AQA')^{-1}A]' \geq 0 \qquad (9)$$

We next give a recursive equation for the smallest solution of (4b,c).

Theorem : The smallest $\pi(.)$ that obeys the constraints (4b,c) is given by the following Riccati equation :

$$\pi(i) = F(i-1)\pi(i-1)F'(i-1) + [N(i) - F(i-1)\pi(i-1)F'(i-1)M'(i)]$$
$$. [M(i)N(i) - M(i)F(i-1)\pi(i-1)F'(i-1)M'(i)]^{-1}$$
$$. [N(i) - F(i-1)\pi(i-1)F'(i-1)M'(i)]' \qquad (10a)$$

$$\pi(-1) = 0 \qquad (10b)$$

Proof : The proof contains three parts.
Part 1 : Let $\pi(i-1)$ satisfy the constraints (4b) for some fixed i-1 in [o,N]. We first show that the smallest $\pi(i)$ that obeys the constraints (4b,c) is given by (10a). For convenience we replace $G(i)G'(i)$ by $Q(i)$ in (4c). Clearly the smallest $\pi(i)$ will be obtained by adding to the known matrix $F(i-1)\pi(i-1)F'(i-1)$ the smallest symmetric nonnegative definite matrix $Q(i)$ such that (4b) is satisfied, i.e. such that :

$$Q(i)M'(i) = N(i) - F(i-1)\pi(i-1)F'(i-1)M'(i) \qquad (11)$$

By the previous lemma, this $Q(i)$ is given by the last term of (10a).

Part 2 : Next we show that if $\overline{\pi}(i-1)$ and $\pi(i-1)$ are two solutions of (4b,c) and if $\overline{\pi}(i)$ and $\pi(i)$ are the corresponding solutions at time i obtained from (10a), then $\overline{\pi}(i-1) \geq \pi(i-1)$ implies

$$\overline{\pi}(i) \geq \pi(i) \qquad (12)$$

Suppose (12) does not hold. Then there exists a vector c such that

$$c'\,\pi(i)\,c > c'\,\overline{\pi}(i)\,c \qquad (13)$$

or equivalently

$$c'Q(i)c > c'\overline{Q}(i)c \qquad (14)$$

where :

$$\underline{Q}(i) = \pi(i) - F(i-1)\,\pi(i-1)F'(i-1) \qquad (15a)$$
$$\overline{Q}(i) = \overline{\pi}(i) - F(i-1)\,\pi(i-1)F'(i-1) \qquad (15b)$$

It is easy to see that $\overline{Q}(i)$ is nonnegative definite. Now (14) and (15b) imply that there exists a $\overline{Q}(i)$, not larger than $Q(i)$, that is also a solution of the constraint (11). This contradicts the fact, shown in Part 1, that $Q(i)$, as given by the last term of (10a) is unique and minimal.

Part 3 : Parts 1 and 2 imply that the smallest $\pi(.)$ will be obtained by solving the Riccati equation (10a) with the smallest possible initial condition $\pi(o)$. By the lemma, the smallest $\pi(o)$ that obeys the constraint (4b) is

$$\pi_o = N(o)\,[M(o)N(o)]^{-1}\,N'(o) \qquad (16)$$

This is equivalent with (10b). This completes the proof.

The minimum variance Markovian representation is now completely determined as follows

$$C(i) = F(i-1)C(i-1) + K(i)\nu(i) \quad , C(-1) = o \qquad (17a)$$

$$y(i) = M(i)C(i) \qquad (17b)$$

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

$$K(i) = [N(i) - F(i-1)\Sigma(i-1)F'(i-1)M'(i)]$$
$$.[M(i)N(i) - M(i)F(i-1)\Sigma(i-1)F'(i-1)M'(i)]^{-\frac{1}{2}} \qquad (17c)$$

$\Sigma(.)$ is the state-variance, $\Sigma(i) = E\{C(i)C'(i)\}$, and obeys the Riccati equation (10) with $\pi(.)$ replaced by $\Sigma(.)$. The existence of $\Sigma(.)$ has been proved in [8].

It turns out that this representation is exactly the normalized IR that was obtained in [8] by the usual method of rewriting the Kalman filter equations, using the orthogonality property and the covariance relations (4a,b) (see eqs. (28) and (21) in [8]).

It should be noticed that for the stationary case Faurre [13] obtained the smallest solution as the asymptotic value of the Riccati equation (10) with F, M and N constant.

IV. Stochastic interpretation of the minimum variance property.

We have shown that a direct search for the minimum variance solution of the constraint equations (4b,c) leads to an IR. This fact has a very simple stochastic interpretation. As we have shown in Section II, the variance $\pi(i)$ of the state $x(i)$ of any Markovian representation must satisfy (4b,c). But as shown in [8] the state of the IR (17) is the filtered estimate $\hat{x}(i|i)$ of the state of any other representation, and its variance $\Sigma(i)$ depends only upon the covariance function. Therefore by the orthogonality of $\hat{x}(i|i)$ and the error $\tilde{x}(i) = x(i) - \hat{x}(i|i)$, it follows that :

$$P(i) = E\{\tilde{x}(i)\tilde{x}(i)\} = \pi(i) - \Sigma(i) \geq 0 \qquad (18)$$

where the inequality follows from the fact that $P(i)$ is a covariance. This relation shows clearly why the state-variance $\Sigma(i)$ of the IR must be smaller than the state-variance of any other model. The state of the IR being the projection of the state of any other model on the space of all past and present observations, it is the smoothest of all possible states, consistent with the output constraint (3a).
Stated otherwise, the state-variance of the IR contains the smallest possible amount of energy consistent with the constraint (4b).

Moreover, while all Markovian models must lead to the same mean-square-error covariance in the predicted signal, the IR has the smallest error covariance matrix for the predicted states

$$P(i|i-1) \triangleq E\{(x(i) - \hat{x}(i|i-1))(x(i) - \hat{x}(i|i-1))'\} \quad (19)$$

Indeed by the orthogonality between $\hat{x}(i|i-1)$ and the error $x(i) - \hat{x}(i|i-1)$ we have

$$P(i|i-1) = \pi(i) - \Sigma(i|i-1) \quad (20)$$

where $\Sigma(i|i-1)$ is the variance of $\hat{x}(i|i-1)$. But $\Sigma(i|i-1)$ depends only on the covariance function (see [8]). Therefore, since the IR has the smallest $\pi(i)$, it has the smallest $P(i|i-1)$.

The stochastic interpretation of the state of the IR as a filtered estimate has the important consequence that filtered or smoothed estimates of the signal process or a related process can be derived, almost by inspection, from the IR. If any other model is used, however, an additional Riccati equation will have to be solved.

V. Conclusion

The class of all Markovian realizations of a time-varying separable covariance is determined by the class of all state-variances that obey a set of constraints. The solution with the smallest state-variance is the IR. We have shown that it can be obtained by directly solving the constraint equations for their smallest solution using a matrix property. Alternatively the IR can be obtained in the usual way via the Kalman filter equations. That the IR has the smallest variance of all Markovian models follows from the stochastic interpretation of its state as the filtered estimate of the state of any other model. The stochastic interpretation of the maximum variance model is a topic of continuing inquiry (see e.g., [14])

References

1. T. Kailath and R.A. Geesey, "An innovations approach to least-squares estimation - Part IV : Recursive estimation given lumped covariance functions", I.E.E.E. Trans. Automat. Control, 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", I.E.E.E. Trans. Automat. Contr., vol. AC-18, Oct. 1973.

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

4. J.B. Moore and B.D.O. Anderson, "Spectral factorization of time-varying covariance functions : The singular case," Math. Syst. Theory, vol. 4, n°1, pp. 10-23, 1970.

5. L.H. Brandenburg, "Shaping filter models for nonstationary random processes", Ph. D. dissertation. Dep. Elec. Eng., Columbia University, New-York, N.Y., June 1968

6. B.D.O. Anderson and T. Kailath, "The choice of signal-process models in Kalman-Bucy filtering", J. of Math. Anal. and Appl., Vol 35, pp. 659-668, Sept. 1971.

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

8. M. Gevers and T. Kailath, "An innovations approach to least-squares estimation - part VI : Discrete-time innovations representations and recursive estimation", I.E.E.E. Trans. Automat. Contr., vol AC-18, Dec. 1973.

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

10. J.C. Willems, "Least-squares stationary optimal control and the algebraic Riccati equation", I.E.E.E. Trans. Automat. Contr., vol AC-16, pp. 621-634, Dec. 1971.

11. B.D.O. Anderson, "Algebraic properties of minimal degree spectral factors", Automatica, vol. 9, pp. 491-500, July 1973.

12. V. Kucera, "On-nonnegative definite solutions to matrix quadratic equations ", Proc. IFAC Fifth World Congress, paper 29.3, Paris, 1972.

13. P. Faurre, "Réalisations Markoviennes de processus stationnaires", IRIA, France, Res. Rept. N° 13, March 1973.

14. H. Akaike, "Markovian representation of stochastic processes by canonical variables" to be published.