Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS Vol. 35, No. 3, September 1971

The Choice of Signal-Process Models in Kalman-Bucy Filtering*

B. D. O. ANDERSON

The University of Newcastle, Australia

AND

THOMAS KAILATH

Stanford University, Palo Alto, Calif.

Submitted by L. A. Zadeh

Kalman and Bucy have shown how to obtain the linear least-squares estimate of a signal, given observations of the signal plus independent white noise, and given a lumped-parameter or state-variable model for the process. The filter producing the signal estimate produces it as a linear functional of an estimate of the state of the model; and although the variance in the error of the signal estimate is independent of that particular model out of the infinitely many possible assumed to generate the signal, the associated covariance of the estimation error in the system states is dependent on the choice of model. The paper establishes that there is one particular model yielding a smallest error-variance in a sense to be described, and that this model is causally invertible. In the particular case where the signal process is stationary and observed over a semi-infinite time interval, this means that the model has the minimum-phase property.

I. INTRODUCTION

Given a lumped-parameter system (state-variable) model for a scalar signal process, Kalman and Bucy [1] showed how to obtain the linear least-squares filtered estimate of the signal observations of the signal plus independent white noise. The Kalman-Bucy method yields the signal estimate as an appropriate linear combination of the estimates of the states of the

* This work was supported by the Applied Mathematics Division of the Air Force Office of Scientific Research under Contract AF 49(638)1517, by the JSEP at Stanford University under Contract Nonr 225(83), and in part by the Australian Research Grants Committee.

model. However, there can be many state-variable models for a process with a given covariance: these models must, of course, lead to the same mean-square-error in the estimate of the signal, but they can have different error-covariance matrices for the estimates of the different state variables. *We prove the existence of a smallest (in a sense to be described) error* covariance matrix, and then establish a conjecture of Kalman (private communication) that the model that yields the smallest such error-covariance matrix must be a "causally invertible" model. (The term causally invertible will be defined subsequently; however, we note one property, viz. that such models are unique). In the special case of a stationary scalar signal process observed over a semi-infinite time interval, the model has the so-called minimum-phase property.

To describe our results more fully, we shall need to define some notation and the Kalman-Bucy problem. Given scalar observations $y(\cdot)$ of a scalar signal process $z(\cdot)$ corrupted by an uncorrelated additive white noise process $v(\cdot)$, $y(t) = z(t) + v(t)$, with

$$E[z(t)\,z(s)] = R(t, s), \qquad E[v(t)\,v(s)] = \delta(t - s), \qquad t_0 \leqslant s, t. \qquad (1)^1$$

Kalman and Bucy [1] described a method for finding

$$\hat{z}(T) = \text{the linear least-squares estimate (l.l.s.e.) of } z(T) \text{ given } \{y(t),\, t_0 \leqslant t \leqslant T\}. \qquad (2)$$

The application of their method requires knowledge of a dynamical (state-variable) model for the process $z(\cdot)$ in the form

$$\dot{x}(t) = F(t)\,x(t) + G(t)\,u(t), \qquad z(t) = h'(t)\,x(t) \qquad (3)$$

where

$$E[u(t)\,u'(s)] = I\delta(t - s), \quad E[x(t_0)\,x'(t_0)] = P_0, \quad E[u(t)\,x'(t_0)] \equiv 0, \quad E[v(t)\,x'(t_0)] \equiv 0. \qquad (4)$$

Evidently

$$E[z(t)\,z(s)] = h'(t)\,E[x(t)\,x'(t)]\,h(s) = R(t, s). \qquad (5)$$

Then the l.l.s.e. $\hat{z}(t)$ is given by the equations

$$\hat{z}(t) = h'(t)\,\hat{x}(t), \quad \dot{\hat{x}}(t) = F(t)\,\hat{x}(t) + P(t)\,h(t)[y(t) - \hat{z}(t)], \quad \hat{x}(t_0) = 0, \qquad (6)$$

[1] All processes are assumed, for convenience, to have zero mean. $E[\cdot]$ denotes expectation and prime denotes transposition.

where $P(\cdot)$ is the solution of the Riccati equation

$$\dot{P} = FP + PF' - Phh'P + GG', \qquad P(t_0) = P_0 . \tag{7}$$

It turns out that $P(\cdot)$ is the covariance of the error in estimating the states $x(\cdot)$:

$$P(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}. \tag{8}$$

We note that the mean-square-error in the estimate of $z(\cdot)$ is given by

$$E[z(T) - \hat{z}(T)]^2 = h'(T)\, P(T)\, h(T) \tag{9}$$

at time T.

The chief advantage of this solution is that it gives a computationally efficient, recursive solution for a large class of nonstationary signal processes. On the other hand, it requires prior knowledge of a model for the signal process. [However it may be argued that in several problems such a model may be directly available from physical considerations. Further, even if no model is directly available, the obvious nonuniqueness of the model (3)–(5) may simplify the determination of such a model from the given covariance function $R(t, s)$. This is in apparent contrast to the familiar Wiener-Hopf spectral factorization, in which a unique minimum phase factorization is required.] An interesting question associated with the nonuniqueness of the model (3)–(5) is how one might distinguish between differing models. Evidently all the models must give the same error covariance for estimating $z(T)$, but different models will generally have different covariances $P(\cdot)$ for the errors in the estimates of the states and, of course, different sets of $\{F, G, h\}$ matrices. Kalman (personal communication) conjectured that, in the scalar-input, stationary, steady-state problem (i.e., G a vector, F, G, h and P constant, and $t_0 = -\infty$):

> The (unique) minimum-phase model for the stationary process $z(\cdot)$ over $(-\infty, T)$ is the one that has the smallest associated P matrix, in the sense that if $\hat{P}$ is the error covariance associated with any other model, then $\hat{P} - P$ is nonnegative definite. (10)

A minimum-phase model is one with a transfer function that has all of its zeros (and all of its poles, because of causality) in the left-half plane.

We shall establish this result, but in a general form applicable also to nonstationary signal processes observed over a finite-time interval. In the nonstationary case, causal invertibility takes the place of the minimum-phase property. Note that it is not *a priori* obvious that there should exist a minimum P matrix, as the notion of nonnegative definiteness induces at best a partial ordering on the set of symmetric matrices.

Before stating and proving the main result, we need some preliminary definitions and remarks. A covariance function is called separable if it is of the form

$$R(t, s) = h'(t \vee s)\, k(t \wedge s) \tag{12}$$ [2]

where $h(\cdot)$ and $k(\cdot)$ are vectors. Processes generated by passing white-noise through a lumped-parameter system are of this form, though the converse is not necessarily true. However, we shall assume that

there is at least one lumped-parameter system model for the process $z(\cdot)$. (13)

This assumption can be replaced by some further simple conditions on $R(t, s)$ (cf., [2–5]); but the present assumption seems most appropriate for this paper.

A covariance function of the form (12) is said [2, 5] to have definite relative-order α if there exists a finite nonnegative integer α such that

(i) The α-th derivatives of $h(\cdot)$ and $k(\cdot)$ exist and are continuous; (14)

(ii) $h'^{(i-1)}k^{(i)} - k'^{(i-1)}h^{(i)} \equiv 0, i = 1, 2, \ldots, \alpha - 2, \alpha - 1$ (15)[3]

and

$$r^2(t) \triangleq h'^{(\alpha-1)}k^{(\alpha)} - k'^{(\alpha-1)}h^{(\alpha)} \text{ is positive for all } t; \tag{16}$$

(iii) the matrix $[h'^{(i)}(0)\, k^{(j)}(0)], i, j = 1, \ldots, \alpha - 1$ is nonsingular. (17)

Direct calculation will show that a process whose covariance has definite relative order α has $(\alpha - 1)$ mean-square derivatives, but the α-th "derivative" contains white noise (hence the adjective definite). This definition provides a class of nonstationary processes that are the analogue of stationary processes with rational power spectral density.

Finally we shall need to define precisely what we mean by the causal invertibility of a lumped-parameter system. A system of the form

$$\dot{x}(t) = G(t)\, u(t), \qquad z(t) = h'(t)\, x(t), \qquad t_0 \leqslant t \tag{18}$$

is called *causally invertible* if given $\{z(t), t_0 \leqslant t \leqslant T\}$ we can uniquely recover $\{u(t), t_0 \leqslant t \leqslant T\}$ *and* $x(t_0)$ for all T. For stationary processes over $(-\infty, T)$, the initial conditions are simple: they are zero. In (18) we have assumed that the $F(\cdot)$ matrix is zero. However, there is no loss of generality in doing this: there exists an easily determined nonsingular transformation[4]

[2] $t \wedge s$ will denote the minimum of t and s and $t \vee s$ their maximum. The use of the vector $h(\cdot)$ in (12) is intentional, as is made clear subsequently.

[3] $h^{(i)}(t) = d^i h(t)/dt^i$.

[4] If $\Phi(t, t_0)$ is the transition matrix of $F(t)$, the coordinate transformation $\mathrm{x}(t) = \Phi(t_0, t)\, x(t)$ has the required properties, as may easily be checked.

that converts a system (3) with a given $F(\cdot)$ matrix into another system with the same impulse response but with $F(\cdot) = 0$. From the input-output point of view, the model is unaltered by this transformation. With these definitions we can state our main result:

THEOREM. *Consider the estimation by the Kalman-Bucy method of a signal process* $z(\cdot)$, *whose covariance satisfies the conditions of* (12)–(17), *in a background of additive white noise. Then of all lumped-parameter models for* $z(\cdot)$ *of the form* (3)–(5), *with a fixed but arbitrary* $F(\cdot)$ *matrix, the one that has the smallest associated* $P(\cdot)$ *matrix* (*the error-covariance matrix in estimating the states of the model*) *is the unique* (*causal and*) *causally invertible model*: *i.e., if the error covariance associated with an arbitrary model is simply* $P(\cdot)$, *while that associated with the causally invertible model is* $P_{\min}(\cdot)$, *then* $P(t) - P_{\min}(t)$ *is nonnegative definite for all* t.

The proof, given in Section II, is short, largely because it relies on several other recently published results of some of our colleagues and ourselves. We are especially indebted to R. Geesey, J. Moore and L. Silverman for the many discussions that developed the background of fact and intuition that formed the basis of our present brief solution.

II. Proof of the Main Theorem

The proof, which is presented in five steps, essentially consists of putting together a number of previous results. The assumptions are stated in (12)–(17). Until Step 5, we assume that the input u of any model generating the signal is scalar. The signal itself is of course scalar throughout the discussions.

Step 1. *Some Relations for a Lumped-Parameter Process*

By our assumption that the process $z(\cdot)$ has at least one scalar input lumped-parameter system model, we can write

$$z(t) = h'(t)\,x(t) \qquad \dot{x}(t) = g(t)\,u(t) \tag{19}$$

where the various quantities are as in Eqs. (3)–(5) except that we have exploited our freedom of choice for $F(\cdot)$ to set it equal to zero; to emphasize the scalar nature of the input, we have replaced $G(t)$ by $g(t)$. We define

$$\Pi(t) = E[x(t)\,x'(t)], \qquad \text{with} \quad \Pi(t_0) = P_0 \text{ given.} \tag{20}$$

Then it is well known [1] that $\Pi(\cdot)$ is characterized by the differential equation

$$\dot{\Pi}(t) = g(t)\,g'(t), \qquad \Pi(t_0) = P_0\,. \tag{21}$$

The covariance of $z(\cdot)$, earlier denoted by $R(t, s)$, is computable using (19) and (20) as

$$E[z(t)\, z(s)] = R(t, s) = h'(t)\, \Pi(t \Lambda s)\, h(s) = h'(t V s)\, k(t \Lambda s) \tag{22}$$

where $k(t)$ is defined via

$$\Pi(t)\, h(t) = k(t). \tag{23}$$

Note. Since knowledge of $R(t, s)$ implies knowledge of $h(\cdot)$—at least to within a constant nonsingular transformation, see (22)—and because we are assuming that $F(t) \equiv 0$, we see that the only way models generating $R(t, s)$ can differ is in the vector $g(\cdot)$, and the initial conditions matrix P_0.

If the process $z(\cdot)$ is observed in additive white noise, the Kalman-Bucy solution for the least-squares estimate $\hat{z}(t)$ is determined by the solution $P(\cdot)$ of the following Riccati equation (see Eq. (7))

$$\dot{P} = gg' - Phh'P, \qquad P(t_0) = P_0. \tag{24}$$

Let us define a new matrix variable

$$\Sigma(t) = \Pi(t) - P(t). \tag{25}$$

Then, from the equations for $\dot{\Pi}$ and $\dot{P}$ and by use of (23), we can write

$$\begin{aligned} \dot{\Sigma} &= (\Pi h - \Sigma h)(\Pi h - \Sigma h)' \\ &= (k - \Sigma h)(k - \Sigma h)', \qquad \Sigma(t_0) = 0. \end{aligned} \tag{26}$$

It can be shown [6] that the assumption that the process $z(\cdot)$ with separable covariance $h'(tVs)\, k(t\Lambda s)$ has at least one lumped-parameter system model ensures that the Riccati equation (26) has a unique continuous solution $\Sigma(\cdot)$. The important factor to note is that $\Sigma(\cdot)$ *is completely determined by the given covariance function whereas both* $\Pi(\cdot)$ *and* $P(\cdot)$ *vary with the model chosen to represent the process* $z(\cdot)$.[5] *However, the relation* (25) *and the non-dependence of* $\Sigma(\cdot)$ *on the model show that the model that yields the minimum* $P(\cdot)$ (assuming there is such a model) *must be the one with minimum* $\Pi(\cdot)$.

Step 2. *Models with Minimum* $\Pi(\cdot)$

For a fixed $F(\cdot)$ matrix, the different models for $z(\cdot)$ are determined by our choices for g and P_0 (of course, consistent with the covariance requirements (20)–(23)). However, the finite definite-relative order assumptions on

[5] We may note in passing that Ph is independent of the particular model since it is equal to $k - \Sigma h$. This fact shows that the Kalman-Bucy filter equation $\dot{\hat{x}} = Ph'(y - h'\hat{x})$ is also independent of the model! However, P, the error-covariance in estimating the states x, does depend upon the model. Note also that Σ is the covariance matrix of $\hat{x}$.

the covariance of $z(\cdot)$ yield the result that the only free variable is the initial covariance matrix P_0. To see this result, which is the second key step in our proof, we use a result proved in [2, 4, 5]. These references show that all single-input lumped models for $z(\cdot)$ can be determined from the solution of the following Riccati equation for $\Pi(\cdot)$:

$$\dot{\Pi} = (\Pi h^{(\alpha)} - k^{(\alpha)})\, r^{-2}(\Pi h^{(\alpha)} - k^{(\alpha)})' \qquad (27)^6$$

with initial condition matrix $\Pi(t_0)$, nonnegative definite symmetric, and satisfying

$$\Pi(t_0)\, h^{(i)}(t_0) = k^{(i)}(t_0), \qquad i = 0,..., \alpha - 1, \qquad (28)$$

where (cf., Eqs. (14)–(17))

$$\alpha = \text{the definite relative order of } R(t, s)$$

and

$$h^{(i)}(t_0) = \partial^i h(t)/\partial t^i \,|_{t=t_0}, \qquad r^2 \triangleq h'^{(\alpha-1)}k^{(\alpha)} - k'^{(\alpha-1)}h^{(\alpha)}.$$

The vector $g(t)$ is given by $[\Pi(t)\, h^{(\alpha)}(t) - k^{(\alpha)}(t)]\, r^{-1}(t)$, but this does not concern us here.

Put another way, if we find any nonnegative definite symmetric $\Pi(t_0)$ satisfying (28), and if the Riccati equation (27) possesses a solution with this $\Pi(t_0)$ as its initial condition, we obtain thereby a single-input model generating $z(t)$. (This fact may actually be verified by direct calculation). In addition, we obtain all single-input models this way.

Step 3. *Minimum $\Pi(t_0)$ yields Minimum $\Pi(t)$, $t \geqslant t_0$*

The next step is to see how variations in $\Pi(t_0)$ can affect the values of $\Pi(t)$, $t \geqslant t_0$. Anderson and Moore [7], exploiting some formulas of McReynolds and Bryson [8], have shown that a larger initial condition matrix results in a larger covariance matrix at every subsequent instant.[7] That is, if $\Pi_1(t_0)$ and $\Pi_2(t_0)$ are two initial covariances with $\Pi_1(t)$ and $\Pi_2(t)$ the corresponding covariances at time t, and if $\Pi_1(t_0) - \Pi_2(t_0)$ is nonnegative definite, then $\Pi_1(t) - \Pi_2(t)$ is nonnegative definite for all t. Therefore, we conclude *that the model with minimum initial covariance matrix $\Pi_{\min}(t_0) = P_{0,\min}$—if such exists—yields the smallest error-covariance matrix, $P_{\min}(t)$ for all $t \geqslant t_0$.*

[6] If we had not chosen $F(\cdot) \equiv 0$, this and later formulas become much more complicated.

[7] In our problem this result can be restated in the following terms: in the least-squares estimation of a process in additive white noise, a smaller initial mean-square-error implies a smaller mean-square-error at all subsequent instants.

In [4], it is proved that as a result of the assumption that there exists at least one model for $z(\cdot)$, there exists a model with an initial condition matrix $P_{0,\min}$ which is unique and satisfies the minimality property. Further, [2, 4] establish the formula

$$P_{0,\min} = B[A'B]^{-1}B' \tag{29}$$

where the matrices A and B are given by

$$A = [h(0)\, h^{(1)}(0) \cdots h^{(\alpha-1)}(0)], \tag{30}$$

$$B = [k(0)\, k^{(1)}(0) \cdots k^{(\alpha-1)}(0)], \tag{31}$$

then $P_{0,\min}$ is the smallest nonnegative definite symmetric matrix such that (28) holds. (We note that the definite-relative-order property (cf., (17)) ensures that $A'B$ is (of rank α and) invertible.)

Step 4. $P_{0,\min}$ *defines an Invertible Model*

The final step in the proof is to show that the (unique) choice $P_{0,\min}$ yields an invertible model for the process $z(\cdot)$. As in the other steps, this fact has also been established in the literature (cf., Silverman [9], who extended the results of Brockett, and Sivan and Weiss). A good presentation of Silverman's results in the context of the modelling problem for random processes has been given by Geesey [2], and in [10]. The important point is as follows: A model

$$\dot{x}(t) = g(t)\, u(t), \qquad z(t) = h'(t)\, x(t)$$

for a process with covariance of definite-relative-order α possesses a well-defined inverse[8] if and only if the initial conditions $x(t_0)$ are such that

$$E[x(t_0)\, x'(t_0)] \qquad \text{has rank } \alpha. \tag{32}$$

But the matrix $P_{0,\min}$ defined in Eq. (29) obviously has rank α and hence the model with $P_{0,\min}$ is (causally) invertible.

Finally, we recall the standard result [11] that a causally invertible model generating a covariance is uniquely specified to within multiplication of the output of the model by a factor $\epsilon(\cdot)$, with $\epsilon(t) = +1$ or -1 for any t. By insisting that the state-space equations describing the model contain matrices with continuous elements, the uniqueness is to within a factor $\epsilon(\cdot)$ with $\epsilon(t) = +1$ for all t or -1 for all t.

[8] That is, a causal system that given $\{y(t),\, t_0 \leqslant t \leqslant T\}$ yields $\{u(t),\, t_0 \leqslant t \leqslant T\}$ *and* $x(t_0)$ for all t.

Step 5. Extension to Vector Inputs

It may be that the scalar process $z(\cdot)$ is generated by passing vector white noise into a linear system. We show here that $\Pi_{\min}(t)$ remains minimum, when compared with the $\Pi(\cdot)$ matrices associated with vector input models.

A slight extension of the results of references [2], [4] and [5] shows that all (vector input) lumped models for $z(\cdot)$ can be determined from solutions of the inequality

$$\dot{\Pi} \geqslant (\Pi h^{(\alpha)} - k^{(\alpha)})\, r^{-2}(\Pi h^{(\alpha)} - k^{(\alpha)})' \tag{33}$$

with the linear constraints

$$\Pi(t)\, h^{(i)}(t) = k^{(i)}(t), \qquad i = 0,..., \alpha - 1, \tag{34}$$

(cf. (27)–(28)) where α, $h^{(i)}$ etc., are as before. This means that there is a nonnegative symmetric matrix $Q(\cdot)$ such that

$$\dot{\Pi} = (\Pi h^{(\alpha)} - k^{(\alpha)})\, r^{-2}(\Pi h^{(\alpha)} - k^{(\alpha)})' + Q \tag{35}$$

(of course, Q cannot be arbitrary, because the side conditions (34) must still hold.) Now denote by $X(t)$ the difference $\Pi(t) - \Pi_{\min}(t)$, which must be nonnegative at t_0 by the very definition of $P_{0,\min}$ as the minimum matrix satisfying (34) evaluated at t_0. Subtraction of Eq. (27) for $\Pi_{\min}(t)$ from (35) yields an equation for $X(t)$ of the form

$$\dot{X} = XAX + BX + XB' + Q$$

where A is nonnegative definite and symmetric. It is then easy to check that nonnegativity of $X(t_0)$ implies nonnegativity of $X(t)$, which is the desired minimality property.

III. Concluding Remarks

The original conjecture of Kalman was for time-invariant systems operating from $t_0 = -\infty$. The material presented has, however, considered time-varying systems operating from some finite initial time t_0. To specialize these systems so that they become time-invariant is not difficult. Let t_0 approach $-\infty$; then $\Pi(t)$ (and $P(t)$) become constant, but still are such that minimality corresponds to invertibility.

The case when the $z(\cdot)$ process is no longer a scalar process is apparently a good deal more difficult technically. The main stumbling block is to obtain the appropriate generalizations of (27) and (28); this has been done in unpublished work of Silverman and Anderson. Despite these difficulties, it seems reasonably safe to assert that our basic conclusions also hold for vector processes.

Representation of covariances using causally invertible models is essential in solving some estimation and detection problems. (Both [10] and [11] contain some applications). Accordingly, it would seem important to determine general properties of this particular class of models. In view too of the many parallels between linear filtering and linear optimal control problems, including those for which the Riccati equation is a commonly used tool in arriving at a solution, the possibility of application of the main result of this paper could be envisaged.

Finally we may remark that though our basic result is simple to state and looks very plausible, we were surprised at the number of different recent results that we used to obtain our proof. It should be of interest to obtain a more direct proof.

References

1. R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, *J. Basic Engr.* 83 (1961), 95–108.
2. R. Geesey, Canonical representations of second-order processes with applications, Ph.D. Dissertation, Department of Electrical Engineering, Stanford University, Stanford, California, December 1968.
3. T. Kailath and R. Geesey, Covariance factorization—An explication via examples, *in* "Proc. Second Asilomar Conference on System Sciences," November 1968.
4. J. B. Moore and B. D. O. Anderson, Spectral factorization of time-varying covariance functions: the singular case, *Math. Systems Theory* (1970), to appear.
5. L. H. Brandenburg, Shaping Filter Models for Nonstationary Random Processes, Ph.D. Dissertation, Department of Electrical Engineering, Columbia University, June 1968.
6. B. D. O. Anderson, J. B. Moore, and S. G. Loo, Spectral factorization of time-varying covariance functions, *IEEE Trans. Information Theory* **15** (1969), 550–557.
7. B. D. O. Anderson and J. B. Moore, Convergence properties of Riccati equation solutions, *IEEE Trans. Automatic Control* 13 (1968), 732–733.
8. S. McReynolds and A. E. Bryson, A successive sweep method for solving optimal programming problems (presented at the Joint Automatic Control Conference) *Proc. JACC* (1965), 551–560.
9. L. Silverman, Properties and application of inverse systems, *IEEE Trans. Automatic Control* 13 (1968), 436–437.
10. R. A. Geesey and T. Kailath, Applications of the canonical representation to estimation and detection in colored noise, *in* "Proceedings of the Symposium on Computer Processing in Communications," Polytechnic Institute of Brooklyn, April, 1969.
11. T. Hida, Canonical representations of Gaussian processes and their applications, *Mem. Col. Sci. Kyoto* XXXIII (1960), 109–155.

Printed in Belgium by the St. Catherine Press Ltd., Tempelhof 37, Bruges