ON THE MAYNE-FRASER SMOOTHING FORMULA AND STOCHASTIC REALIZATION THEORY FOR NONSTATIONARY LINEAR STOCHASTIC SYSTEMS*

Faris Badawi, Anders Lindquist, and Michele Pavon

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

ABSTRACT

This paper is a shortened version of [1], its basic purpose being to provide an easily accessible introduction to the results of [1], many of which are presented here without proofs. However, we have tried to rearrange the material of [1], changing the logical order in which various topics are introduced, and occasionally we regard the results from a somewhat different angle. This has been done to increase the present paper's usefulness as a complement to [1].

The work reported here is aimed at providing a theory of smoothing in the context of stochastic realization theory. This approach enables us to obtain stochastic interpretations of many important smoothing formulas and to explain the relationship between them. In this paper, however, we shall only consider one such formula, namely the Mayne-Fraser two-filter formula, which has a very natural interpretation in the stochastic realization setting; we refer the reader to [1] for further results. As a by-product, we also obtain certain results on the stochastic realization problem itself.

1. INTRODUCTION

Consider a linear stochastic system

$$(S)\quad \begin{cases} dx = A(t)x(t)dt + B(t)dw \; ; \; x(0) = \xi & (1.1a) \\ dy = C(t)x(t)dt + D(t)dw \; ; \; y(0) = 0 & (1.1b) \end{cases}$$

defined on the interval $0 \le t \le T$, where x is the n-dimensional *state process*, y is the m-dimensional *output process*, w is a p-dimensional process with orthogonal increments such that

$$E\{dw\} = 0; \quad E\{dwdw'\} = Idt \tag{1.2}$$

(prime denotes transposition), ξ is a centered random vector with finite covariance $\Pi := E\{\xi\xi'\}$ and uncorrelated with w, and A, B, C, and D are matrices of bounded functions with properties to be further specified below. We shall consider two problems related to such systems:

Problem 1. For an arbitrary $t \in [0,T]$, find the linear least-squares estimate $\hat{x}(t)$ of the state x(t) given the output record $\{y(t) \; ; \; 0 \le t \le T\}$, i.e., find the *wide sense conditional mean*

$$\hat{x}(t) = \hat{E}\{x(t) | y(\tau) \; ; \; 0 \le \tau \le T\}. \tag{1.3}$$

This is the *smoothing problem*, which has generated a rather extensive literature [1-18, 48], and is of considerable importance in applications.

Problem 2. Given the stochastic process $\{y(t) \; ; \; 0 \le t \le T\}$, find all possible systems (1.1) (in some suitable class of models S) having this process as its output process. This is the *stochastic realization problem* discussed in [20-33], and each such model S is called a *stochastic realization of* $\{y(t) \; ; \; 0 \le t \le T\}$. Note that we are only considering *proper* stochastic realizations [20], i.e., models S whose outputs not merely have the same covariance properties as the given process (the only requirement in the earlier realization theory [34-38]), but are equal to it a.s. for each t.

As we shall see in this paper, these two problems are intimately connected to each other. In fact, all the well-known smoothing formulas found in [2-18] have natural interpretations in the stochastic realization setting; see [1] for a more complete discussion of these results. Here we shall only consider the so-called *Mayne-Fraser two-filter formula* [5,6], on which topic a large number of papers have been written [7-9,12-17]. The many attempts to motivate this formula stochastically have, in our opinion, been less than convincing. We refer the reader to [48] for a well-written account of these matters. In our realization setting, however, the two filters have a natural interpretation: they are simply the minimum- and maximum-variance realizations respectively. Hence, the latter is not a "backward filter" as suggested in the literature (although it can be reformulated as such), but a "forward filter" just as its structure suggests.

The concept of *backward realization* is an essential tool in this paper. A similar approach was applied to the smoothing problem in the earlier papers [14-17], but, since only "wide sense" backward representations were used, some subtle points were overlooked. The fundamental idea of this paper, to embed the given system (1.1) into a class of stochastic realizations, was motivated by the results in [20-22]. Note that restricting our analysis to models (1.1) for which $BD' = 0$ (as in [14-17]), would render the natural class of realizations incomplete, since it would exclude the minimum- and maximum-variance realizations.

This paper is essentially a shortened conference version of [1], but the last section contains some aspects on the stochastic realization problem not included in [1]. Whenever a proof has been omitted, it can be found in [1].

2. SOME NOTATIONS

Let H be the space of all centered stochastic variables (on an underlying probability space) with finite second-order moments. Then H is a Hilbert space with inner product $(\xi,\eta) = E\{\xi\eta\}$. For an arbitrary k-dimensional stochastic process $\{z(t); \; 0 \le t \le T\}$ with components in H, define $H_t(z)$ to be the (closed) subspace spanned by the random variables $\{z_1(t), z_2(t), \ldots, z_k(t)\}$, and let H(z) be the closed linear hull in H of the subspaces $\{H_t(z); \; 0 \le t \le T\}$; we shall write this as $H(z) = \bigvee_{t\in[0,T]} H_t(z)$. Similarly define the *past space*

* This research was supported partially by the Airforce Office of Scientific Research, USAF Systems Command, under grant No. AFOSR-78-3519, and partially by the National Science Foundation under grant No. ENG-79-03731.

$H_t^-(z) := \bigvee_{\tau\in[0,t]} H_\tau(z)$ and the *future space* $H_t^+(z) := \bigvee_{\tau\in[t,T]} H_\tau(z)$. Sometimes we shall be more interested in spaces spanned by the *increments* of z. Hence, we define $H(dz)$, $H_t^-(dz)$ and $H_t^+(dz)$ to be the closed linear hulls in H of $\{z(\tau) - z(\sigma);\ \tau, \sigma \in I\}$ where I is the interval $[0,T]$, $[0,t]$ and $[t,T]$ respectively.

For each $\eta \in H$ and subspace $K \subset H$ let $\hat{E}\{\eta | K\}$ be the projection of η onto K, i.e., the *wide sense conditional mean*. Let u be a stochastic vector with components in H, and let H(u) be the closed linear span in H of the components of u. Then, for any $\eta \in H$, we shall often write $\hat{E}\{\eta | u\}$ in place of $\hat{E}\{\eta | H(u)\}$, and, for any subspace $K \subset H$, $\hat{E}\{u | K\}$ will denote the vector with components $\hat{E}\{u_i | K\}$.

3. FORWARD AND BACKWARD STOCHASTIC REALIZATIONS

Assuming that $R := DD'$ has a bounded inverse on $[0,T]$, it is well-known that the linear least-squares estimate

$$x_*(t) = \hat{E}\{x(t) | H_t^-(dy)\} \tag{3.1}$$

of the state process x of S is generated on $[0,T]$ by the *Kalman-Bucy filter*

$$\begin{aligned} dx_* &= Ax_*dt + B_*R^{-1/2}(dy - Cx_*dt); \\ x_*(0) &= 0 \end{aligned} \tag{3.2a}$$

where $R^{1/2}(t)$ is the symmetric square root of R(t), and the gain function B_* is given by

$$B_* = (Q_*C' + BD')R^{-1/2}, \tag{3.2b}$$

the error covariance matrix

$$Q_*(t) = E\{[x(t) - x_*(t)][x(t) - x_*(t)]'\} \tag{3.2c}$$

being the solution of the matrix Riccati equation

$$\begin{cases} \dot{Q}_* = AQ_* + Q_*A' \\ \quad - (Q_*C' + BD')R^{-1}(Q_*C' + BD')' + BB' \\ Q_*(0) = \Pi. \end{cases} \tag{3.2d}$$

Note that the filter (3.2a), and hence the estimate x_*, is completely determined by the matrices A, C, R and B_*. Clearly there are many models S having the same Kalman-Bucy filter.

In the sequel we shall only consider models S which are *minimal*, i.e., there is no other realization of $\{y(t);\ 0 \le t \le T\}$ with a state process x of smaller dimension n, and *analytic*, i.e., the coefficient matrices A, B, C, D and R^{-1} are analytic on $[0,T]$. Both these assumptions are purely technical and are introduced to insure that a certain matrix function is invertible; they could probably be removed at the price of a less elegant theory. Now, let the initial realization S used in forming (3.2) be minimal and analytic, and define $\mathcal{S}$ to be the class of all analytic realizations of $\{y(t);\ 0 \le t \le T\}$ having (3.2) as its Kalman-Bucy filter. Then all realizations of class $\mathcal{S}$ are minimal. Clearly A, C and $R := DD'$ are the same for all $S \in \mathcal{S}$, while B, D and the *state covariance function*

$$P(t) := E\{x(t)x(t)'\} \tag{3.3}$$

will differ over the class $\mathcal{S}$. Of course, different $S \in \mathcal{S}$ will have completely different stochastic processes x and w.

Furthermore, from (3.2b) and the fact that $Q_* = P - P_*$, where $P_*(t) := E\{x_*(t)x_*(t)'\}$, it follows that also the function

$$G := PC' + BD' \tag{3.4}$$

is an invariant over $\mathcal{S}$. In fact,

$$G = P_*C' + B_*R^{1/2}. \tag{3.5}$$

It is easy to see that P satisfies the matrix differential equation

$$\dot{P} = AP + PA' + BB'\ ;\ P(0) = \Pi, \tag{3.6}$$

which has the solution

$$P(t) = \Phi(t,0)\Pi\Phi(t,0)' + \int_0^t \Phi(t,\tau)B(\tau)B(\tau)'\Phi(t,\tau)'d\tau \tag{3.7}$$

where Φ is the transion matrix of the system $\dot{z} = Az$ of differential equations. Hence, $P(t) > 0$ for all $t \in [0,T]$ if and only if S belongs to the subclass $\mathcal{S}_+ = \{S \in \mathcal{S} | \Pi > 0\}$. [For symmetric matrices P and Q, $P \ge Q$ ($P > Q$) means that P-Q is nonnegative (positive)definite.] It can be shown [1] that $\mathcal{S}_+$ is nonempty.

Let $S \in \mathcal{S}_+$. Then $\bar{x}(t) := P(t)^{-1}x(t)$ is a well-defined stochastic vector process on all of $[0,T]$, and it can be shown [1; Lemma 2.3] that it satisfies the *backward Markovian representation*

$$d\bar{x} = -A'\bar{x}dt + \bar{B}d\bar{w}\ ;\ \bar{x}(T) = \bar{\xi} \tag{3.8a}$$

where $\bar{B} = P^{-1}B$, $\bar{w}$ is a p-dimensional orthogonal increment process of type (1.2) defined by

$$d\bar{w} = dw - B'P^{-1}xdt, \tag{3.8b}$$

and $\bar{\xi} := P(T)^{-1}x(T)$ is uncorrelated with $\bar{w}$. Then $H_t^-(d\bar{w}) \perp H_t^+(\bar{x})$ for all $t \in [0,T]$; this is what characterized the *backward* property of (3.8). Moreover, the state covariance function $\bar{P}(t) := E\{\bar{x}(t)\bar{x}(t)'\}$ satisfies $\bar{P} = P^{-1}$.

Representation (3.8) is a strict sense version of a similar result presented in [15,16,42]. (The last paper contains an alternative justification of the formulas of [15,16] using the techniques of [12,13].) The version given in these papers is however insufficient for our purposes since it provides a representation up to second-order properties only. Modulo some trivial technicalities, the proof of (3.8) above [1; Lemma 2.3] is the same as the one presented in [20]. (In this context it should be mentioned that *all* the basic ideas of a recent paper coauthored by Kailath [IEEE Trans. IT-25(1979), p.121-124] are contained in [20,21], and that, three months prior to the submission of Kailath's paper and at his request, [20,21] were personally handed over to him by one of the authors.)

Together with (3.4), representation (3.8) yields a *backward realization of* $\{y(t)\ ;\ 0 \le t \le T\}$, namely

$$(\bar{S}) \begin{cases} d\bar{x} = -A'\bar{x}dt + \bar{B}d\bar{w}\ ;\ \bar{x}(T) = \bar{\xi} \\ dy = G'\bar{x}dt + Dd\bar{w} \end{cases} \tag{3.9}$$

whose state covariance matrix function $\bar{P}$ satisfies

$$\dot{\bar{P}} = -A'\bar{P} - \bar{P}A - \bar{B}\bar{B}';\ \bar{P}(T) = \bar{\Pi}, \tag{3.10}$$

where $\bar{\Pi} := P(T)^{-1}$. By this procedure each $S \in \mathcal{S}_+$ gives rise to a backward realization $\bar{S}$; the class of all $\bar{S}$

generated in this way will be denoted $\bar{S}_+$.

Next we proceed to enlarge the class $\bar{S}_+$. Given any $\bar{S} \in \bar{S}_+$, (by symmetry with the forward setting) the linear least squares estimate

$$\bar{x}_*(t) = \hat{E}\{\bar{x}(t) | H_t^+(dy)\} \qquad (3.11)$$

is generated by the *backward Kalman-Bucy filter*

$$d\bar{x}_* = -A'\bar{x}_* dt + \bar{B}_* R^{-1/2}(dy - G'\bar{x}_* dt) \; ; \quad \bar{x}_*(T) = 0, \qquad (3.12)$$

where the gain function $\bar{B}_*$ can be determined via a matrix Riccati equation [1]. Now, in complete analogy with the forward setting, we define $\bar{S}$ to be the class of all analytic backward realizations of $\{y(t); 0 \le t \le T\}$ whose backward Kalman-Bucy filter is given by (3.12). It can be shown [1] that $\bar{S}_+ \subset \bar{S}$, and hence we have obtained the required extension. All realizations of class $\bar{S}$ are minimal.

Unfortunately, there is no one-one correspondence between models in S and $\bar{S}$. For this we need to enlarge these classes even further. This leads to *generalized stochastic realizations*.

4. GENERALIZED STOCHASTIC REALIZATIONS

In order to extend the one-one correspondence between forward and backward realizations beyond S_+ and $\bar{S}_+$ we shall have to enlarge S and $\bar{S}$ slightly in the following way. Let $\hat{S}$ be the class of all systems (1.1) which for any $\varepsilon > 0$ is an analytic realization of $\{y(t); 0 \le t \le T - \varepsilon\}$ having (3.2a), restricted to the interval $[0,T-\varepsilon]$, as its Kalman-Bucy filter. Similarly, we define $\hat{\bar{S}}$ to be the class of all models (3.9) which for any $\varepsilon > 0$ is an analytic realization of $\{y(t); \varepsilon \le t \le T\}$ such that (3.12), restricted to $[\varepsilon,T]$, is its backward Kalman-Bucy filter. The elements of $\hat{S}$ and $\hat{\bar{S}}$ will be called *generalized realizations* and *generalized backward realizations of* $\{y(t); 0 \le t \le T\}$ respectively. Clearly $S \subset \hat{S}$ and $\bar{S} \subset \hat{\bar{S}}$.

Then to each realization $S \in S$ there corresponds a generalized backward realization $\bar{S} \in \hat{\bar{S}}$. In fact, it can be shown that, since S is minimal, (A,B) is completely controllable. This together with the analyticity implies that (A,B) is *totally* controllable [40,41]. Consequently P(t) has an analytic inverse on any interval $[\varepsilon,T]$, for the last term of (3.7) is the controllability gramian. (Cf [1]; Lemma 2.2). Hence the procedure leading to the backward model (3.9) can always be carried out on the restricted interval $[\varepsilon,T]$. Similarly there corresponds a generalized realization $S \in S$ to any backward realization $\bar{S} \in \bar{S}$. We collect these observations in the following theorem.

THEOREM 4.1. *To each realization (1.1) in* S *there corresponds a generalized backward realization (3.9) in* $\hat{\bar{S}}$ *such that* $\bar{P} = P^{-1}$, $\bar{B} = P^{-1}B$, $\bar{x} = P^{-1}x$ *and* $d\bar{w} = dw - B'P^{-1}xdt$. *Likewise to each backward realization (3.9) in* $\bar{S}$ *there is a generalized realization (1.1) in* $\hat{S}$ *such that* $P = \bar{P}^{-1}$, $B = \bar{P}^{-1}\bar{B}$, $x = \bar{P}^{-1}\bar{x}$ *and* $dw = d\bar{w} + \bar{B}'\bar{P}^{-1}\bar{x}dt$.

5. THE MINIMUM- AND MAXIMUM- VARIANCE REALIZATIONS

It is well-known that the *innovation process* $\{w_*(t); 0 \le t \le T\}$, whose increments are defined by

$$dw_* = R^{-1/2}(dy - Cx_* dt), \qquad (5.1)$$

is a process with orthogonal increments satisfying (1.2) and $H_t^-(dw_*) = H_t^-(dy)$ for all $t \in [0,T]$ (see e.g. [43]). Then (3.2a) and (5.1) yield

$$(S_*)\begin{cases} dx_* = Ax_* dt + B_* dw_* \; ; \; x_*(0) = 0 \\ dy = Cx_* dt + R^{1/2} dw_*, \end{cases} \qquad (5.2)$$

which is a realization in S, for B_* is clearly analytic.

Likewise, the backward Kalman-Bucy filter (3.12) can be written

$$(\bar{S}_*)\begin{cases} d\bar{x}_* = -A'\bar{x}_* dt + \bar{B}_* d\bar{w}_*; \; \bar{x}_*(T) = 0 \\ dy = G'\bar{x}_* dt + R^{1/2} d\bar{w}_*, \end{cases} \qquad (5.3)$$

where $\{\bar{w}_*(t); 0 \le t \le T\}$ is the *backward innovation process*

$$d\bar{w}_* = R^{-1/2}(dy - G'\bar{x}_* dt), \qquad (5.4)$$

which has orthogonal increments and satisfies (1.2) and the condition $H_t^+(d\bar{w}_*) = H_t^+(dy)$ for all $t \in [0,T]$. (See [20,45].) Clearly, $\bar{S}_* \in \bar{S}$. Now let

$$(S^*)\begin{cases} dx^* = Ax^* dt + B^* dw^*; \; x^*(0) = \xi^*, \\ dy = Cx^* dt + R^{1/2} dw^* \end{cases} \qquad (5.5)$$

be the forward counterpart of $\bar{S}_*$ as defined by Theorem 4.1, and let P^* be the corresponding state covariance function. Since $\bar{S}_* \notin \bar{S}_+$, S^* exists only as a generalized realization, and obviously $P^*(t) \to \infty$ as $t \to T$.

LEMMA 5.1. *Let* P_* *and* P^* *be the state covariance functions of* S_* *and* S^* *respectively and define* $Q := P^* - P_*$. *Then the state covariance function* P *of an arbitrary realization* $S \in S$ *satisifies*

$$P_*(t) \le P(t) \le P^*(t) \qquad (5.6)$$

for all $t \in [0,T)$. *Moreover,* $Q > 0$.

Consequently, we shall call S_* the *minimum-* and S^* the *maximum-variance realization*. By eliminating dw^* in (5.5), it is immediately seen that x^* satisfies the Kalman-Bucy type equation

$$dx^* = Ax^* dt + B^* R^{-1/2}(dy - Cx^* dt); \; x^*(0) = \xi^* \qquad (5.7a)$$

Let S be an arbitrary realization of class S. Then, defining $Q^* := P^* - P$, it is not hard to see from (3.4) and (3.6) that

$$B^* = -(Q^*C' - BD')R^{-1/2} \qquad (5.7b)$$

with Q^* satisfying the matrix Riccati equation

$$\begin{cases} \dot{Q}^* = AQ^* + Q^*A' + (Q^*C' - BD')R^{-1}(Q^*C' - BD')' - BB' \\ Q^*(0) = \Pi^* - \Pi, \end{cases} \qquad (5.7c)$$

where $\Pi^* = \bar{P}_*(0)^{-1}$. Clearly $Q^*(t) \to \infty$ as $t \to T$. The filter (5.7) is precisely the mysterious "backward filter" of the Mayne-Fraser two-filter formula; as we have seen above, it is actually a *forward* realization. Since $Q^* = P^* - P$, we can interpret Q^* as an error covariance function, much in analogy with the Kalman-Bucy filter. In fact,

$$Q^*(t) = E\{[x(t) - x^*(t)][x(t) - x^*(t)]'\} \qquad (5.8)$$

for all $t \in [0,T)$. This is an immediate consequence of the following lemma, which we shall need again in the next section.

LEMMA 5.2. *Let* x *be the state process and* P *the state covariance function of any realization in* $\hat{S}$. *Then*

$$E\{x(t)x_*(t)'\} = P_*(t), \quad E\{x(t)x^*(t)'\} = P(t) \qquad (5.9)$$

and

$$E\{[x(t) - x_*(t)][x^*(t) - x(t)]'\} = 0 \qquad (5.10)$$

on any interval on which these quantities are defined.

We shall now demonstrate that the two processes x_* and x^* together contain all the relevant information on y needed in estimating the state process x of an arbitrary realization $S \in \hat{S}$. To this end first note that (3.1) can be written

$$\hat{E}\{H_t(x)|H_t^-(dy)\} = H_t(x_*), \qquad (5.11)$$

and that, since obviously $H_t(x^*) = H_t(\bar{x}_*)$, (3.11) yields

$$\hat{E}\{H_t(x)|H_t^+(dy)\} = H_t(x^*) \qquad (5.12)$$

for all $t \in [0,T)$. Now define the orthogonal complements $N_t^- := H_t^-(dy) \ominus H_t(x_*)$ and $N_t^+ := H_t^+(dy) \ominus H_t(x^*)$ respectively. Then we obtain the orthogonal decomposition

$$H(dy) = N_t^- \oplus H_t^\square \oplus N_t^+ \qquad (5.13)$$

where $H_t^\square$ is the *frame space*

$$H_t^\square = H_t(x_*) \vee H_t(x^*) \qquad (5.14)$$

(where $A \vee B$ denotes the closed linear hull in H of A and B.) (Cf. [22,24,26].)

LEMMA 5.3. (cf. [27]) *Let x be the state process of a realization in $\hat{S}$. Then, for* $t \in [0,T)$,

$$H_t(x) \subset H_t^\square \oplus [H(dy)]^\perp$$

where $[H(dy)]^\perp$ *is the orthogonal complement of* H(dy) *in* H.

Proof. Clearly $H_t(x) \perp N_t^-$. To see this note that the components of $x(t) - x_*(t)$ are orthogonal to $H_t^-(dy) \supset N_t^-$ and that the components of $x_*(t)$ belong to $H_t(x_*) \perp N_t^-$. In the same way we show that $H_t(x) \perp N_t^+$. □

6. THE SMOOTHING PROBLEM

Consider an arbitrary realization (1.1) in the class S. The basic problem before us is to determine the smoothing estimate

$$\hat{x}(t) = \hat{E}\{x(t)|H(dy)\} \qquad (6.1)$$

for each $t \in [0,T)$ and to interpret it in terms of stochastic realizations. Let Σ denote the corresponding estimation error covariance, i.e.,

$$\Sigma(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}. \qquad (6.2)$$

In view of Lemma 5.3, $\hat{x}(t) \in H_t^\square$, and consequently there are two matrix functions K_* and K^* such that

$$\hat{x}(t) = K_*(t)x_*(t) + K^*(t)x^*(t). \qquad (6.3)$$

The components of the estimation error $x(t) - \hat{x}(t)$ are clearly orthogonal to H(dy), and hence, in particular, to the components of $x_*(t)$ and $x^*(t)$. Therefore, $E\{x(t)x_*(t)'\} = E\{\hat{x}(t)x_*(t)'\}$ and $E\{x(t)x^*(t)'\} = E\{\hat{x}(t)x^*(t)'\}$. By Lemma 5.2, the first of these relations yields $P_* = K_*P_* + K^*P_*$ and consequently

$$K_*(t) + K^*(t) = I \qquad (6.4)$$

for all $t \in (0,T)$, because $P_*(t)$ is nonsingular on this interval. The second relation yields

$$P(t) = K_*(t)P_*(t) + K^*(t)P^*(t) \qquad (6.5)$$

for all $t \in [0,T)$. Then solving (6.4) and (6.5) for K_* and K^* we obtain $K_* = Q^*Q^{-1}$ and $K^* = Q_*Q^{-1}$, where as before $Q_* = P - P_*$, $Q^* = P^* - P$ and $Q = P^* - P_*$. Note that Q(t) is nonsingular for all $t \in [0,T)$ (Lemma 5.1) and that

$$Q(t) = Q_*(t) + Q^*(t). \qquad (6.6)$$

THEOREM 6.1. *Let x be the state process of a realization (1.1) of class $\hat{S}$. Then the smoothing estimate* (6.1) *is given by*

$$\hat{x}(t) = [I - Q_*(t)Q(t)^{-1}]x_*(t) + Q_*(t)Q(t)^{-1}x^*(t) \qquad (6.7)$$

and the error covariance function (6.2) *by*

$$\Sigma(t) = Q_*(t) - Q_*(t)Q(t)^{-1}Q_*(t) \qquad (6.8)$$

for all $t \in [0,T)$.

Proof. Relation (6.7) was derived above for $t \in (0,T)$; for $t = 0$, (6.7) follows from (7.6) below. To prove (6.8) note that

$$x - \hat{x} = (I - Q_*Q^{-1})(x - x_*) + Q_*Q^{-1}(x - x^*). \qquad (6.9)$$

By Lemma 5.2 the two terms of (6.9) are orthogonal and therefore, observing (3.2c) and (5.8),

$$\Sigma = (I - Q_*Q^{-1})Q_*(I - Q^{-1}Q_*) + Q_*Q^{-1}Q^*Q^{-1}Q_*,$$

which, in view of (6.6), yields (6.8). □

7. THE MAYNE-FRASER SMOOTHING FORMULA

We shall now restrict our attention to realizations for which both Q_* and Q^* are invertible for arbitrary $t \in [0,T)$. Since $Q_* = P - P_*$ and $Q^* = P^* - P$, this is possible for all $S \in S$ such that $P_*(t) < P(t) < P^*(t)$ for all t on this interval. We shall call the class of all such S the *interior* of S and denote it int S. It can be shown that int S is indeed nonempty [1; Lemma 3.6].

THEOREM 7.1. *Let* $S \in$ int *S, let x be the state process of S, and let $\hat{x}$ be the corresponding smoothing estimate* (6.1). *Then, for each* $t \in [0,T)$

$$\hat{x}(t) = \Sigma(t)[Q_*(t)^{-1}x_*(t) + Q^*(t)^{-1}x^*(t)], \qquad (7.1)$$

where x_* and x^* are given by (3.2) *and* (5.7) *respectively and the smoothing error covariance Σ by*

$$\Sigma(t)^{-1} = Q_*(t)^{-1} + Q^*(t)^{-1}. \qquad (7.2)$$

Proof. Since $S \in$ int S, Q_* and Q^* are invertible. By writing (6.8) as $\Sigma = Q_*Q^{-1}(Q - Q_*)$ and using (6.6), it is seen that

$$\Sigma = Q_*Q^{-1}Q^*. \qquad (7.3)$$

Inverting this and again using (6.6) yields (7.2). From (7.3) we also see that $Q_*Q^{-1} = \Sigma(Q^*)^{-1}$. Then $I - Q_*Q^{-1} = \Sigma[\Sigma^{-1} - (Q^*)^{-1}] = \Sigma Q_*^{-1}$. Hence (7.1) follows from (6.7). □

Relations (7.1) and (7.2) together with (3.2) and (5.7) is the *Mayne-Fraser two-filter formula* [5,6], which has received considerable attention in the literature [7-9,13-17]. Although this algorithm is easy to derive formally [9,12,13], its probabilistic justification has caused considerable difficulty, partly due to the fact that $Q^*(t) \to \infty$ as $t \to T$. The system (5.7) has usually been interpreted as a backward filter, and in [14,17] it is presented as the limit of such a filter as a certain covariance matrix function tends to infinity. However, in our stochastic realization setting (5.7) has a very natural interpretation: It is simply the maximum-variance *forward* realization S^*. By using the identity

$$x^*(t) = \bar{P}_*(t)^{-1}\bar{x}_*(t) \tag{7.4}$$

we can instead write the smoothing formula (7.1) in terms of two Kalman-Bucy filters, one (3.2) evolving forward and the other (3.12) evolving backward in time. (Note that then (7.1) is defined on the whole interval [0,T].) This fact was pointed out in [14,15,17], in which papers the backward estimate

$$\hat{x}_b(t) = \hat{E}\{x(t)|H_t^+(dy)\} \tag{7.5}$$

was used in place of $\bar{x}_*$, a choice that may at first sight seem more natural. The reader should however note that

$$\hat{x}_b(t) = P(t)P^*(t)^{-1}x^*(t) \tag{7.6}$$

is not invariant over S and is therefore less suitable for our purposes. It is not hard to see that

$$(Q^*)^{-1} = [(Q^*)^{-1} + P^{-1}]P(P^*)^{-1} \tag{7.7}$$

and consequently (7.1) may also be written

$$\hat{x}(t) = \Sigma(t)\{Q_*(t)^{-1}x_*(t) + [Q^*(t)^{-1} + P(t)^{-1}]\hat{x}_b(t)\}, \tag{7.8}$$

which is the formula presented in [14,15,17]. In the early papers [7,8], relation (7.1) was introduced via a formula [47] for optimal weighting of two estimates with orthogonal errors. No justification of this orthogonality was given in [8], and the argument in [7] is incomplete due to problems with the end point condition. (Cf. [48].) However, the stochastic realization theory provides a natural justification of this procedure. Indeed, (5.10) is the required orthogonally condition.

8. INTERNAL AND EXTERNAL REALIZATIONS

Since $R := DD'$ is assumed to be full rank, the dimension of w is always greater than or equal to the dimension of y, i.e., $m \le p$. We shall now consider realizations for which $m = p$. Then D is invertible, and w can be eliminated from (1.1) to yield the Kalman-Bucy type equation

$$dx = Axdt + BD^{-1}(dy - Cxdt);\ x(0) = \xi. \tag{8.1}$$

If, in addition, we assume that $\xi \in H(dy)$, it is immediately clear that

$$H(x) \subset H(dy). \tag{8.2}$$

We shall call any generalized realization of $\{y(t);\ 0 \le t \le T\}$ for which (8.2) holds an *internal* realization; all other $S \in \hat{S}$ will be named *external* [20]. Obviously, the internal realizations are precisely those for which the smoothing problem is trivial, the estimate being exact. In particular, S_* and S^* are internal.

Let $\hat{S}_0$ be the class of all $S \in \hat{S}$ such that $p = m$ and $\xi \in H(dy)$. Then we have just shown that all realizations in $\hat{S}_0$ are internal. The following theorem, the proof of which is given in [1], states that, under some mild regularity conditions, the converse is also true.

THEOREM 8.1. *A realization* $S \in \hat{S}$ *such that* $\begin{pmatrix} B \\ D \end{pmatrix}$ *has full rank is internal if and only if* $S \in \hat{S}_0$.

In view of Theorem 6.1, the state process x of any realization S of class S can be written

$$x(t) = [I - \Pi(t)]x_*(t) + \Pi(t)x^*(t) + \tilde{x}(t), \tag{8.3}$$

where $\Pi := Q_*Q^{-1}$ and $\tilde{x} := x - \hat{x}$. The smoothing error $\tilde{x}$ is identically zero if and only if S is internal. To obtain a complete characterization of the external realizations in S_+, we shall provide a representation for $\tilde{x}$ also. To this end, note that, given a realization (1.1), there exists an orthogonal $p \times p$-matrix $V(t)$ for each $t \in [0,T]$ such that

$$\begin{bmatrix} \bar{B}(t) \\ D(t) \end{bmatrix} = \begin{bmatrix} B_1(t) & B_2(t) \\ R^{1/2}(t) & 0 \end{bmatrix} V(t), \tag{8.4a}$$

where B_1 is $n \times m$ and B_2 is $n \times (p - m)$, and that

$$\begin{bmatrix} du \\ dv \end{bmatrix} = V\,dw \tag{8.4b}$$

defines a pair of orthogonal increment processes u and v, of dimensions m and p - m respectively. Obviously (8.4b) satisfies (1.2).

THEOREM 8.2. *Let x be the state process of a realization* $S \in S_+$ *and let* B_2 *and v be defined by* (8.4). *Then the smoothing error* $\tilde{x}$ *is given by*

$$\begin{cases} \tilde{x}(t) = Q_*(t)\eta(t) & (8.5a) \\ d\eta = -\Gamma_*'\eta dt + Q_*^{-1}B_2\, d\zeta;\ \eta(T) = \eta_T & (8.5b) \end{cases}$$

where $\Gamma_* = A - B_*R^{-1/2}C$, $\eta_T = Q_*^{-1}(T)[x(T) - x_*(T)]$ *and* ζ *is a* (p - m)*-dimensional orthogonal increment process of type (1.2) such that* $H(d\zeta) \perp H(dy)$. *Moreover,* $\eta_T \perp H(d\zeta)$, *i.e., (8.5a) is a backward Markovian representation (3.8a), and the increments of* ζ *are given by*

$$d\zeta = dv - B_2'Q_*^{-1}(x - x_*)dt. \tag{8.6}$$

Together with (8.5), (8.3) constitutes a generalization of the stationary internal state representation in Theorem 5.5 of [20]. For external realizations, however, Π is not a projection. In fact, Π is a projection if and only if S is internal. To see this observe that $\Pi^2 = \Pi$, i.e., $Q_*Q^{-1}Q_* = Q_*$, if and only if $\Sigma = 0$ (Theorem 6.1).

For internal realizations we have the following stronger result, which illustrates the important role played by the feedback matrix Γ_* defined in Theorem 8.2. It is a generalization of a result found in [22; pp.75-79].

THEOREM 8.3. *Let* Ψ *be the transition function of* Γ_*, *i.e.,*

$$\frac{\partial}{\partial t}\Psi(t,s) = \Gamma_*(t)\Psi(t,s);\ \Psi(s,s) = I \tag{8.7}$$

Then x is a state process of an internal realization of class $\hat{S}$ *if and only if there is a family* $\{M_t;\ t \in [0,T)\}$ *of subspaces of* R^n, *satisfying the condition*

$$\Psi(t,s)M_s \subset M_t \quad \textit{for all } s \le t, \tag{8.8}$$

such that, for each $t \in [0,T)$,

$$x(t) = [I - \Pi(t)]x_*(t) + \Pi(t)x^*(t), \tag{8.9}$$

where $\Pi(t)$ *is a projection onto* M_t *along* $Q(t)M_t^\perp$, $Q(t)$ *being the covariance matrix of* $x^*(t) - x_*(t)$. *Then* Π *is given by*

$$\Pi = Q_*Q^{-1} \tag{8.10}$$

where Q_* *is defined by (3.2c).*

Proof. We shall use the same idea of proof as in [22]. (only if): Let x be the state process of an internal realization of class $\hat{S}$. Then, by Theorem 6.1, x satisfies (8.9) with Π given by (8.10). We just proved that Π is a projection (onto some subspace M_t), and the fact that $\Pi(t)Q(t)$ is symmetric implies that $\Pi(t)$ projects along $Q(t)M_t^{\perp}$. It just remains to show that the family $\{M_t; t \in [0,T]\}$ of subspaces satisfies (8.8), or, which is equivalent,

$$\Pi(t)\Psi(t,s)\Pi(s) = \Psi(t,s)\Pi(s) \text{ for all } t \geq s. \quad (8.11)$$

To this end, first note that (8.9) can be written

$$z_* = \Pi z \quad (8.12)$$

where $z_* := x - x_*$ and $z := x^* - x_*$. From the differential equations for z_* and z it is immediately seen that $\hat{E}\{z(t)|z(s)\} = \Psi(t,s)z(s)$ and $\hat{E}\{z_*(t)|z_*(s)\} = \Psi(t,s)z_*(s)$. Projecting the first of these relations over $H_s(z_*)$ and premultiplying by $\Pi(t)$, one obtains

$$\hat{E}\{z_*(t)|z_*(s)\} = \Pi(t)\Psi(t,s)\Pi(s)z(s), \quad (8.13)$$

for, in view of (8.12), $H_s(z_*) \subset H_s(z)$ and, by Lemma 5.2 and the usual projection formula [1; Lemma 2.1], $\hat{E}\{z(s)|z_*(s)\} = Q_*(s)Q(s)^{-1}z(s)$. Then comparing (8.13) with the second of the formulas in the text above (8.13), (8.11) follows by noting (8.12) and the fact that $Q(s) = E\{z(s)z(s)'\} > 0$ (Lemma 5.1). (if): Let $\{\Pi(t); t \in [0,T]\}$ be a family of projections satisfying the conditions of the theorem. Since $Q := E\{z(t)z(t)'\}$ and $Q_* := E\{z_*(t)z_*(t)'\}$, it is immediately seen that Π is given by (8.10). In fact, by (8.12), $Q_* = \Pi Q \Pi' = \Pi^2 Q = \Pi Q$, for, since $\Pi(t)$ projects along $Q(t)M_t^{\perp}$,

$$Q\Pi' = \Pi Q. \quad (8.14)$$

Moreover, since

$$H_t(z_*) \subset H_t(z) \perp H_t^-(dy) \quad (8.15)$$

[see (8.12)], (3.1) holds. Hence, it only remains to prove that x is actually the state process of a realization S with the prescribed values of A, C and R; then B_* will have the right value also, and, in view of (8.9), S must be internal. To this end, first note that, in view of (5.2) and (5.5), z satisfies the differential equation

$$dz = \Gamma_* z dt - QC'R^{-1/2}dw^*; \; z(0) = \xi^*. \quad (8.16)$$

We need to prove that

$$\hat{E}\{z(t)|H_s^-(z_*)\} = \Psi(t,s)z_*(s) \text{ for } t \geq s. \quad (8.17)$$

But (8.17) is equivalent to $E\{z(t)z_*(\tau)'\} = \Psi(t,s)E\{z_*(s)z_*(\tau)'\}$ for all $\tau \leq s$, and, by (8.12) and (8.16), this is the same as

$$\Psi(t,\tau)Q(\tau)\Pi(\tau)' = \Psi(t,s)\Pi(s)\Psi(s,\tau)Q(\tau)\Pi(\tau)', \quad (8.18)$$

which is an immediate consequence of (8.11) and (8.14). Then, premultiplying (8.17) by $\Pi(t)$ and using (8.11) and (8.12), we have

$$\hat{E}\{z_*(t)|H_s^-(z_*)\} = \Psi(t,s)z_*(s) \text{ for } t \geq s. \quad (8.19)$$

Now, inserting

$$dw_* = R^{-1/2}Czdt + dw^* \quad (8.20)$$

into the state equation of S_* we see that

$$\hat{E}\{x_*(t)|H_s^-(z_*)\} = \Phi(t,s)\hat{E}\{x_*(s)|H_s^-(z_*)\} + \int_s^t \Phi(t,\tau)B_*(\tau)R^{-1/2}(\tau)C(\tau)\Psi(\tau,s)d\tau z_*(s), \quad (8.21)$$

where we have used (8.17) to obtain the last term. Adding (8.19) and (8.21) yields

$$\hat{E}\{x(t)|H_s^-(z_*)\} = \Phi(t,s)\hat{E}\{x(s)|H_s^-(z_*)\} \text{ for } t \geq s, \quad (8.22)$$

for the last term of (8.21) can be written $[\Phi(t,s) - \Psi(t,s)]z_*(s)$ [39; p.117]. Moreover,

$$\hat{E}\{x(t)|H_s^-(dy)\} = \Phi(t,s)\hat{E}\{x(s)|H_s^-(dy)\} \text{ for } t \geq s. \quad (8.23)$$

To see this, replace x by $x_* + z_*$ and remember that $H_t(z_*) \subset H_t(z) \perp H_s^-(dy)$. Now, (8.22) and (8.23) together yield

$$\hat{E}\{x(t)|H_s^-\} = \Phi(t,s)x(s) \text{ for } t \geq s, \quad (8.24)$$

where $H_s^- := H_s^-(dy) \vee H_s^-(z_*)$. Next define a process u with increments $du = dw_* - R^{-1/2}Cz_*dt$, which can be seen to be an orthogonal increment process of type (1.2) such that $H(du) \subset H_T^-$. It is not hard to see that $H_t^+(du) \perp H_t^-$. In fact, in view of (8.20), $du = R^{-1/2}C(z - z_*)dt + dw^*$, and, since $H_t^-(z_*) \subset H_t^-(z) \perp H_t^+(dw^*)$ [see (8.16)], (8.17) and (8.19) imply that $H_t^+(du) \perp H_t^-(z_*)$; also, because of (8.15), $H_t^+(du) \perp H_t^-(dy)$ holds trivially. But

$$dy = Cxdt + R^{1/2}du, \quad (8.25)$$

and therefore

$$\hat{E}\{y(t)|H_s^-\} = [\int_s^t C(\tau)\Phi(\tau,s)d\tau]x(s) \quad (8.26)$$

follows from (8.24) and the fact that $H_s^+(du) \perp H_s^-$. Now, since $H_s^-(dy) \vee H_s^-(x) \subset H_s^-$, (8.24) and (8.26) imply that $(x',y')'$ is a Markov process, and consequently it has a representation (1.1). Relations (8.24) - (8.26) insure that A, C and R have the required values. □

Finally, to further stress the importance of the feedback matrix Γ_*, let us point out that (8.3) can be written

$$x(t) = x_*(t) + Q_*(t)Q(t)^{-1}z(t) + Q_*(t)\eta(t) \quad (8.27)$$

where x_* satisfies

$$dx_* = \Gamma_* x_* dt + B_* R^{-1/2}dy; \; x_*(0) = 0, \quad (8.28)$$

and z and η are given by (8.16) and (8.5b).

REFERENCES

1. F. Badawi, A. Lindquist, and M. Pavon, A stochastic realization approach to the smoothing problem, *IEEE Trans. Automatic Control*, AC-24(Dec. 1979).

2. A. E. Bryson and M. Frazier, Smoothing for linear and nonlinear dynamic systems, TDR 63-119, Aero. Sys. Div., (1963), Wright-Patterson Air Force Base, Ohio, 353-364.

3. H.E. Rauch, F. Tung, and C.T. Striebel, Maximum liklihood estimates of linear dynamic systems, *AIAA J.*, Vol.3, No.8, August 1965, 1445-1450.

4. J.S. Meditch, On optimal fixed point smoothing, *Int. J. Control*, Vol.6, No.2(1967), 189-199.

5. D.Q. Mayne, A solution of the smoothing problem for linear dynamic systems, *Automatica*, Vol.4(1966), 73-92.

6. D.C. Fraser, A new technique for optimal smoothing of data, Sc. D. Dissertation, MIT, Cambridge, MA(1967).

7. R.K. Mehra, On optimal and suboptimal linear smoothing, *Proc. Nat. Electr. Conf.*, XXIV(1968), 119-124.

8. D.C. Fraser and J.E. Potter, The optimum linear smoother as a combination of two optimum linear filters, *IEEE Trans. Automatic Control*, AC-14(1969), 387-390.

9. T. Kailath and P.A. Frost, An innovations approach to least-squares estimation-Part II: linear smoothing in additive noise, *IEEE Trans. Automatic Control*, AC-13 (1968), 655-660.

10. L.E. Zachrisson, On optimal smoothing of continuous-time Kalman processes, *Information Sciences,* 1(1969) 143-172.

11. A. Lindquist, A theorem on duality between estimation and control for linear stochastic systems with time delay, *J. Mathematical Analysis and Applications,* Vol.37, No.2(1972), 516-536.

12. D.G. Lainiotis, Optimal linear smoothing: continuous data case, *Int. J. Control,* Vol.14, No.6(1971), 1137-340.

13. ____________, Partitioned estimation algorithms, II: linear estimation, *Information Sciences,* 7(1974), 317-340.

14. L. Ljung and T. Kailath, A unified approach to smoothing formulas, *Automatica,* Vol.12(1976), 147-157.

15. T. Kailath and L. Ljung, Backwards Markovian models for second order stochastic process, *IEEE Trans. Information Theory,* IT-22(1976), 488-491.

16. G.S. Sidhu and U.B. Desai, New smoothing algorithms based on reversed-time lumped models, *IEEE Trans. Automatic Control* AC-21(1976), 538-541.

17. ______________________, Reversed-time lumped models, smoothing algorithms and Fredholm resolvents, *Proc. of the JHU Conf. on Information Science and Systems,* Johns Hopkins Univ., Baltimore, MD, March-April, 1976.

18. J.S. Meditch, A survey of data smoothing for linear and non-linear dynamic systems, *Automatica,* Vol.9 (1973), 151-162.

19. G. Picci, Stochastic realization of Gaussian processes, *Proc. IEEE,* 64(1976), 112-122.

20. A. Lindquist and G. Picci, On the stochastic realization problem, *SIAM J. Control and Optimization,* 17(May,1979), 365-389.

21. M. Pavon, Stochastic realization and invariant directions of the matrix Riccati equation, *SIAM J. Control and Optimization* (to appear).

22. G. Ruckebusch, Représentations Markoviennes de processus Gaussiens stationnaires, Ph.D. Thesis, Univ. of Paris, VI, 1975.

23. A. Lindquist and G. Picci, On the structure of minimal splitting subspaces in stochastic realization theory, *Proc. 1977 Conf. Decision and Control,* New Orleans, 1977, 42-48.

24. ______________________, A state-space theory for stationary stochastic processes, *Proc. 21st Midwest Symposium on Circuits and Systems,* Ames, Iowa, August, 1978.

25. ______________________, A Hardy space approach to the stochastic realization problem, *Proc. 1978 Conf. Decision and Control,* San Diego, January, 1979.

26. ______________________, Realization theory for multivariate stationary Gaussian process I: State space construction, 4th International Symposium on the Mathematical Theory of Networks and Systems, July, 1979, Delft, Holland.

27. A. Lindquist, G. Picci and G. Ruckebusch, On minimal splitting subspaces and Markovian representations, *Mathematical Systems Theory,* 12(1979), 271-279.

28. M. Pavon, Constant directions and singular stochastic realization: the scalar case, *Proc. Ire Colloque AFCET-SMF,* Sept. 1978, Palaiseau, France.

29. Y.A. Rozanov, On two selected topics connected with stochastic systems theory, *Applied Mathematics and Optimization,* 3(1976), 73-80.

30. G. Ruckebusch, Représentations Markoviennes de processus Gaussiens stationnaires, *C.R. Acad. Sc. Paris,* Series A, 282(1976), 649-651.

31. ___________, Représentations Markoviennes de processus Gaussien stationnaires et applications statistiques, *Journées de Statistiques des Processus Stochastiques,* Grenoble, June 1977, Springer-Verlag Lecture Notes in Mathematics 636.

32. ___________, A state space approach to the stochastic realization problem, *Proc. 1978 Intern. Symp. Circuits and Systems,* New York.

33. ___________, A geometric approach to the stochastic realization problem Center de Mathématiques Appliqués, Ecole Polytechnique, Internal Report No.41, Nov. 1978.

34. B.D.O. Anderson, The inverse problem of stationary covariance generation, *J. Stat. Phys.,* 1(1969)133-147.

35. P. Faurre, Réalizations Markoviennes de processus stationnaires, Research Report No.13, March 1973, IRIA(LABORIA), Le Chesnay, France.

36. J.C. Willems, Dissipative dynamical systems, Part II: linear systems with quadratic supply rates, *Archive for Rational Mechanics and Analysis,* 45(1972),352-393.

37. M. Clerget, Systèmes linéaires positifs non stationnaires, Research Report No.65, April 1974, IRIA (LABORIA), Le Chesnay, France.

38. F. Germain, Algoritmes de calcul de réalisations stationnaires, Research Report No.66, April 1974, IRIA (LABORIA), Le Chesnay, France.

39. R.W. Brockett, *Finite Dimensional Linear Systems,* Wiley, 1970.

40. L.M. Silverman and H.E. Meadows, Controllability and observability in time-variable linear systems, *SIAM J. Control,* Vol.5, No.1(1967), 64-73.

41. D.M. Wiberg, *State Space and Linear Systems,* Shaum's Outline Series, McGraw-Hill Book Company, 1971.

42. D.G. Lainiotis, General backward Markov models, *IEEE Trans. Auto. Cont.* AC-21 (August 1976), 595-598.

43. T. Kailath, A note on least squares estimation by the innovations method, *SIAM J. Control,* Vol.10, No.3 (1972), 477-486.

44. B.D. Anderson and P.J. Moylan, Spectral factorization of a finite-dimensional non-stationary matrix covariance, *IEEE Trans. Auto. Cont.* AC-19, No.6 (Dec.1974), 680-692.

45. A. Lindquist, Optimal filtering of continuous-time stationary processes by means of the backward innovation process, *SIAM J. Cont.,* Vol.12,(1974), 747-754.

46. J.C. Willems, Least squares stationary optimal control and the algebraic Riccati equation, *IEEE Trans. Auto. Cont.* AC-16(1971), 621-634.

47. R. Deutsch, *Estimation Theory,* Prentice Hall, Inc., Englewood Cliffs, N.J., 1965.

48. J. Wall, Jr., Control and estimation for large-scale systems having spatial symmetry, Ph.D. Thesis, MIT, Cambridge, MA, (1978).