

AD-A078 137 KENTUCKY UNIV LEXINGTON DEPT OF MATHEMATICS

AD-A078 137 KENTUCKY UNIV LEXINGTON DEPT OF MATHEMATICS F/8 12/1
A STOCHASTIC REALIZATION APPROACH TO THE SMOOTHING PROBLEM. REV--ETC(U)
1979 F BADAWI , A LINDQUIST , M PAVON AFOSR-78-3519

AFOSR-78-3519

AFOSR-78-3519

UNCLASSIFIED

AFOSR-TR-79-1274

AD
A078137

AD
A078137

END
DATE
FILMED
1 - 80
DDC

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-79-1274	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 3
4. TITLE (and Subtitle) A STOCHASTIC REALIZATION APPROACH TO THE SMOOTHING PROBLEM. Revised.	5. TYPE OF REPORT & PERIOD COVERED Interim Report	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Faris Badawi, Anders Lindquist, and Michele Pavon	8. CONTRACT OR GRANT NUMBER(s) AFOSR-78-3519	
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Kentucky Department of Mathematics Lexington, KY 40506	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304A1	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	12. REPORT DATE 1979	13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 1234	15. SECURITY CLASS. (of this report) UNCLASSIFIED	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Supersedes AD-A071103		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) 79 22 5 057		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The purpose of this paper is to develop a theory of smoothing for finite dimensional linear stochastic systems in the context of stochastic realization theory. The basic idea is to embed the given stochastic system in a class of similar systems all having the same output process and the same Kalman-Bucy filter. This class has a lattice structure with a smallest and a largest element; these two elements completely determine the smoothing estimates. This approach enables us to obtain stochastic interpretations of many important smoothing formulas and to explain the relationship between them.		

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A078137

DDC FILE COPY

DDC
RECEIVED
DEC 12 1979
D

To appear as a regular paper in IEEE Trans. AC-24 (Dec. 1979).
AFOSR-TR- 79 - 1274

REVISED VERSION

A STOCHASTIC REALIZATION APPROACH TO THE SMOOTHING PROBLEM*

Faris Badawi, Anders Lindquist, and Michele Pavon

Department of Mathematics
University of Kentucky
Lexington, Kentucky 40506

ABSTRACT: The purpose of this paper is to develop a theory of smoothing for finite dimensional linear stochastic systems in the context of stochastic realization theory. The basic idea is to embed the given stochastic system in a class of similar systems all having the same output process and the same Kalman-Bucy filter. This class has a lattice structure with a smallest and a largest element; these two elements completely determine the smoothing estimates. This approach enables us to obtain stochastic interpretations of many important smoothing formulas and to explain the relationship between them.

*Research sponsored by Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-78-3519.

Accession For	
NTIS GRA&I	☒
DDC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

DDC
RECEIVED
DEC 12 1979
D

Approved for public release;
distribution unlimited.

1. INTRODUCTION

Let $\{x(t); 0 \leq t \leq T\}$ and $\{y(t); 0 \leq t \leq T\}$ be two stochastic vector processes, of dimensions n and m respectively, defined as the solution of the linear system of stochastic differential equations

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

$$(1.1b)$$

where w is a vector process, of dimension $p \geq m$, with orthogonal increments such that

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

(prime denotes transposition), ξ is a centered random vector with finite covariance $\Pi := E\{\xi\xi'\}$ and uncorrelated with w , $R(t) := D(t)D(t)'$ is positive definite on $[0, T]$, and A, B, C, D , and R^{-1} are matrices of analytic functions defined on $[0, T]$. The model S is usually called a *linear stochastic system*; y is its *output process*, w is its *input process* and x its *state process*. We shall assume that the representation S is *minimal* in the sense that there is no other model of the form (1.1) with the process y as its output and with a state process x of smaller dimension than n . Clearly the matrix function $P(t) := E\{x(t)x(t)'\}$ satisfies the differential equation

$$\dot{P} = AP + PA' + BB'; \quad P(0) = \Pi \quad (1.3)$$

on $[0, T]$. We shall call P the *state covariance function* of S .

The following problem is of considerable importance in the systems sciences. For an arbitrary $t \in [0, T]$, find the linear least-squares estimate $\hat{x}(t)$ of the state vector $x(t)$ given the output $\{y(\tau); 0 \leq \tau \leq T\}$, i.e., the wide sense conditional expectation

$$\hat{x}(t) = \hat{E}\{x(t) | y(\tau); 0 \leq \tau \leq T\} \quad (1.4)$$

in the terminology of Doob [1]. This is the *smoothing problem*, and it has generated a rather extensive literature [2-17]. (See the survey paper [18] for further references.) Here we shall study this problem from a new angle employing concepts and techniques from the *stochastic realization theory* developed in [20-22] and more recently in [23-33]. The basic idea consists in embedding the model (1.1) into a class S of models S all having the same process y as its output and all having the same Kalman-Bucy filter. Such a representation is called a *stochastic realization* of y . (Note that we only consider *proper* realizations [20], i.e., models S whose outputs not merely have the same covariance properties, the only requirement in the earlier realization theory [34-38], but are equal for each t a.s.) It can be seen that, slightly extended, the class S has a lattice structure with a smallest (S_*) and a largest (S^*) element, the partial ordering being induced by the "size" of the covariance matrix $P(t)$ in the sense that $P_1 > P_2$ if $P_1 - P_2$ is positive definite. This approach will enable us to obtain stochastic interpretations of many important smoothing formulas and lay the groundwork for a *theory* of smoothing which so far has been lacking.

Our interest in the smoothing problem was caused by the Mayne-Fraser two-filter formula [5, 6], on which topic a large number of papers have been written [7-9, 12-17]. In some of these papers the authors have encountered difficulties in motivating this formula, and the many attempts to justify it stochastically have, in our opinion, been less than convincing¹. In our stochastic realization setting the two filters have a natural interpretation: they are simply the minimum and maximum variance realizations S_* and S^* 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.

At first sight some of the technical assumptions above may seem rather stringent, namely the minimality condition and the analyticity of the coefficient matrices. These conditions are introduced to insure that, for each $S \in \mathcal{S}$, the state covariance matrix $P(t)$ is invertible for each $t \in (0, T)$. It is quite probable that these assumptions can be relaxed², but our object here is to convey some basic ideas, and we do not want to obscure matters by introducing extra difficulties of a purely technical nature. On the other hand, the model (1.1) is more general than the one usually encountered in the smoothing literature in that Bdw and Ddw may be correlated. There is a reason for this too. To limit our analysis to models S for which $BD' = 0$ would render the class $\mathcal{S}$ incomplete.

The contents of the paper are as follows. Section 2 is devoted to some preliminary results. We present a strict sense version of some results on backward Markovian representations developed, for much the same purposes, in [15, 16]. The idea of proof is borrowed from [20]. In Section 3 we define the stochastic realization setting mentioned above, and in Section 4 we apply it to derive and interpret various smoothing procedures.

2. PRELIMINARIES

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 $\{z(t); 0 \leq t \leq T\}$ with components in H , define $H_t(z)$ to be the (closed) subspace spanned by the random variables $\{z_1(t), z_2(t), \dots, z_k(t)\}$, and let $H(z)$ be the closed linear hull in H of the subspaces $\{H_t(z); 0 \leq t \leq T\}$; we shall write this as $H(z) = \bigvee_{t \in [0, T]} H_t(z)$. Similarly define the *past space* $H_t^-(z) := \bigvee_{\tau \in [0, t]} H_\tau(z)$

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* [1]. 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 $\hat{E}\{u_i|K\}$. We shall need the following lemma, the proof of which can be found in most standard texts on estimation theory.

LEMMA 2.1. Let u and v be two stochastic vectors with components in H and assume that $E\{vv'\}$ is positive definite. Then

$$\hat{E}\{u|v\} = E\{uv'\}(E\{vv'\})^{-1}v. \quad (2.1)$$

The state process x defined by (1.1a) is a *wide sense Markov process* [1], i.e.,

$$\hat{E}\{x(t)|H_s^-(x)\} = \hat{E}\{x(t)|x(s)\} \quad \text{for } t \geq s. \quad (2.2)$$

To see this, merely note that $x(t)$ can be written

$$x(t) = \Phi(t, s)x(s) + \int_s^t \Phi(t, \tau)B(\tau)dw \quad (2.3)$$

and that $H_s^+(dw) \perp H_0(x) \oplus H_s^-(dw) \supset H_s^-(x)$. (The symbol $\perp$ denotes "orthogonal to.") Here, of course, Φ is the transition matrix defined by

$$\frac{\partial \Phi}{\partial t}(t, s) = A(t)\Phi(t, s); \quad \Phi(s, s) = I \quad (2.4)$$

In deriving the main results of this paper we shall need to reverse the direction of time in (1.1). The Markov property is independent of the choice of time direction and therefore we also have

$$\hat{E}\{x(s)|H_t^+(x)\} = \hat{E}\{x(s)|x(t)\} \quad \text{for } t \geq s. \quad (2.5)$$

(In the present setting this can be seen by observing that, in view of (2.3), $H_t^+(x) \ominus H_t^-(x) \subset H_t^+(dw) \perp x(s)$.) The differential equation (1.1a), however, is not symmetric with respect to time; the two terms in the right member of (2.3) are orthogonal if and only if $t \geq s$. Hence we need to define a backward version of (1.1a). This requires the inversion of the covariance matrix $P(t)$, which is the topic of the following lemma. Here and in the sequel $Q > 0$ ($Q \geq 0$) means that the symmetric matrix Q is positive (nonnegative) definite.

LEMMA 2.2. Let P be the state covariance function of the linear stochastic system S defined in §1. Then, for any $\epsilon > 0$, P^{-1} exists and is analytic on the interval $[\epsilon, T]$. If $\Pi > 0$, the same holds for the complete interval $[0, T]$.

Proof. Integrating (1.2) yields

$$P(t) = \Phi(t, 0)\Pi\Phi(t, 0)' + \int_0^t \Phi(t, \tau)B(\tau)B(\tau)'\Phi(t, \tau)'d\tau \quad (2.6)$$

which is positive definite if $\Pi > 0$; hence, since A and B are analytic on $[0, T]$, so is P^{-1} . Now assume that $\Pi \not> 0$. Since S is minimal, (A, B) must be completely controllable. In fact, were this not the case, the input-output map of S could be reduced [39], contradicting minimality. Since in addition A and B are analytic, (A, B) is totally controllable [40, 41]. Therefore since the second term in (2.6) is the controllability gramian, $P(t) > 0$ on any interval $[\epsilon, T]$. The analyticity of P^{-1} then follows in the same way as above. $\square$

As we shall see below, it is more convenient to express the backward representation in terms of the process

$$\bar{x}(t) = P(t)^{-1}x(t) \quad (2.7)$$

rather than x . In view of Lemma 2.2, $\bar{x}(t)$ is well defined with components in H on any interval $[\epsilon, T]$. Let $\bar{P}$ denote its covariance function, i.e.,

$$\bar{P}(t) = E\{\bar{x}(t)\bar{x}(t)'\}. \quad (2.8)$$

We are now in a position to formulate a backward version of the state equation (1.1a).

LEMMA 2.3. *Let x be the state process of the linear stochastic system S . Then, for any $\epsilon > 0$, the process $\bar{x}$ defined by (2.7) satisfies the backward model*

$$d\bar{x} = -A(t)'\bar{x}(t)dt + \bar{B}(t)d\bar{w}; \quad \bar{x}(T) = \bar{\xi} \quad (2.9)$$

on $[\epsilon, T]$, where $\bar{\xi} = P(T)^{-1}x(T)$, $\bar{B} = P^{-1}B$ and $\bar{w}$ is a p -dimensional orthogonal increment process satisfying (1.2) and the condition $H_t^-(d\bar{w}) \perp H_t^+(\bar{x})$ for all t . The increments of $\bar{w}$ are given by

$$d\bar{w} = dw - B(t)'P(t)^{-1}x(t)dt, \quad (2.10)$$

and the covariance function (2.8) by $\bar{P} = P^{-1}$; it satisfies the Liapunov equation

$$\dot{\bar{P}} = -A'\bar{P} - \bar{P}A - \bar{B}\bar{B}'; \quad \bar{P}(T) = \bar{\Pi}, \quad (2.11)$$

where $\bar{\Pi} = P(T)^{-1}$. If $\Pi > 0$, equations (2.9)-(2.11) are defined on the whole interval $[0, T]$.

Lemma 2.3 is a strict sense version of a similar result presented in [15, 16]. As explained in [42], an alternative justification of the wide

sense results [15, 16] can be obtained by means of the earlier work [12, 13]. The version given in all these papers is however insufficient for our purposes since it provides a deterministic rather than a probabilistic result. Moreover, we have chosen to write the backward equation in terms of $\bar{x}$ rather than x as in [15, 16]. (However, see the "adjoint" formulation in [16].) The reason for this will become evident in Section 3. Our choice will yield a backward Kalman-Bucy filter which is invariant over the class S , the one in [15, 16] will not.

The proof of Lemma 2.3 follows exactly the same lines as in [20]. It is based on the observation that, for $s \leq t$, the orthogonal decomposition

$$\bar{x}(s) = \hat{E}\{\bar{x}(s) | H_t^+(x)\} + [\bar{x}(s) - \hat{E}\{\bar{x}(s) | H_t^+(x)\}] \quad (2.12)$$

can be written in the form

$$\bar{x}(s) = \Phi(t, s)' \bar{x}(t) + \int_t^s \Phi(\tau, s)' \bar{B}(\tau) d\bar{w} \quad (2.13)$$

which is the integral form of (2.9).

Proof of Lemma 2.3. In view of Lemma 2.2, the state covariance function P is invertible on the stated interval. Clearly $\bar{P} = P^{-1}$. Then, since $\dot{\bar{P}} = -\bar{P}\dot{P}\bar{P}$, (2.11) follows from (1.3). Then Lemma 2.1 together with (2.5) and (2.7) yields

$$\hat{E}\{\bar{x}(s) | H_t^+(x)\} = \Phi(t, s)' \bar{x}(t) \quad (2.14)$$

for it follows from (2.3) that $E\{x(s)x(t)'\} = P(s)\Phi(t, s)'$ for $s \leq t$. Consequently, the process $u(t) := \Phi(t, 0)' \bar{x}(t)$ is a wide sense *backward martingale* with respect to $H_t^+(x)$, i.e.,

$$\hat{E}\{u(s) | H_t^+(x)\} = u(t) \quad \text{for } s \leq t, \quad (2.15)$$

and hence it has orthogonal increments. We shall now show that u can be normalized as follows:

$$u(s) - u(t) = \int_t^s \Phi(\tau, 0)' \bar{B}(\tau) d\bar{w}, \quad (2.16)$$

where $\bar{w}$ is defined by (2.10). To this end differentiate $u(t) = \Phi(t, 0)' \bar{P}(t)x(t)$ and use (1.1a), (2.4) and (2.11) to obtain $du = \Phi(t, 0)' \bar{B}(dw - B' \bar{P}x dt)$. It remains to show that $\bar{w}$ is an orthogonal increment process satisfying (1.2). This follows from a tedious but straight-forward calculation of the incremental covariance function. (If B were full rank, we could conclude this directly from the martingale property (2.15); this could be achieved by working with the complete system S instead.) The desired representation is then obtained by noting that

$$\bar{x}(s) = \Phi(0, s)' [u(t) + u(s) - u(t)],$$

into which we insert (2.16) to obtain (2.13). Obviously $H_t^+(\bar{x}) \perp H_t^-(d\bar{w})$, for, by construction, the two terms in (2.13) are orthogonal for all t . $\square$

3. FORWARD AND BACKWARD STOCHASTIC REALIZATIONS

Let $\{y(t); 0 \leq t \leq T\}$ be an m -dimensional vector process defined as the output of the linear stochastic system S introduced in Section 1. Any system of type (1.1) [with $\xi \in H$, w satisfying (1.2) and $\xi \perp H(dw)$] having the given process y as its output is called a *realization* of y . In particular, by assumption, S is *minimal*, i.e., there is no other realization of y with a state process of smaller dimension³, and *analytic*, i.e., its parameter matrices A , B , C , D and R^{-1} are analytic on $[0, T]$. Clearly the components of $x(t)$ and $y(t)$ belong to H for all $t \in [0, T]$, and the same holds for the increments of w .

It is well-known that the least-squares estimate

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

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

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

where $R^{1/2}(t)$ is the symmetric square root of $R(t) = D(t)D(t)'$, and the gain function B_* is given by

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

the error covariance matrix

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

being the solution of the matrix Riccati equation

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

As we shall see shortly there are other realizations which have the same Kalman-Bucy filter (3.2a). Hence we define S to be the class of all analytic realizations S of y whose Kalman-Bucy filter, determined as in (3.2), has the same coefficient functions A , C , R and B_* as in (3.2a). Then (since we only consider *proper* [20] realizations) the estimates x_* are also the same. (The error covariance Q_* , however, will of course vary over S .) Clearly all realizations in S are minimal. Moreover, it is well-known that the *innovation process* $\{w_*(t); 0 \leq t \leq T\}$, whose increments are defined by

$$dw_* = R^{-1/2}(dy - Cx_*dt), \quad (3.3)$$

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 (3.3) yield

$$(S_*) \begin{cases} dx_* = Ax_* dt + B_* dw_* & ; \quad x_*(0) = 0 \\ dy = Cx_* dt + R^{1/2} dw_*. \end{cases} \quad (3.4)$$

Since B_* is analytic on $[0, T]$, this is a realization in S whose state covariance matrix $P_*(t) := E\{x_*(t)x_*(t)'\}$ satisfies

$$\dot{P}_* = AP_* + P_*A' + B_*B_*'; \quad P_*(0) = 0. \quad (3.5)$$

(This can also be seen by subtracting (1.3) from (3.2d), noting that $Q_* = P - P_*$.) Now define the $n \times m$ matrix function

$$G = P_*C' + B_*R^{1/2}, \quad (3.6)$$

which is clearly analytic on $[0, T]$.

LEMMA 3.1. Let G be defined by (3.6). Then for any realization $S \in S$,

$$P(t)C(t)' + B(t)D(t)' = G(t) \quad (3.7)$$

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

Proof. This follows from (3.2b) and the fact that $Q_* = P - P_*$. $\square$

Consequently A , C , G and R are invariants for the class S --in fact, the covariance function of y is determined by these four matrix functions [37, 44]--whereas B , D , P , w and x will vary with different realizations S . Actually even the dimension p of the process w will vary. However, since R is full rank, we will always have $p \geq m$.

The Kalman-Bucy filter realization S_* belongs to a class of realizations for which p is minimal, i.e., $p = m$. Define S_0 to be the subclass of all $S \in S$ such that $p = m$ and $x(0) \in H(dy)$. (Note that, since $y(0) = 0$,

$H(dy) = H(y)$. We shall use the former notation as we are really only interested in the increments of y , the assumption $y(0) = 0$ being one of convenience.) Let

$$(S_0) \begin{cases} dx_0 = Ax_0 dt + B_0 dw_0 & ; \quad x_0(0) = \xi_0 \\ dy = Cx_0 dt + D_0 dw_0 \end{cases} \quad (3.8)$$

be a realization in S_0 with state covariance function P_0 . Then D_0 is invertible and therefore

$$dx_0 = Ax_0 dt + B_0 D_0^{-1} (dy - Cx_0 dt); \quad x_0(0) = \xi_0. \quad (3.9a)$$

Now let (1.1) be an arbitrary realization in S and define

$$Q_0 = P - P_0. \quad (3.9b)$$

Then Lemma 3.1 yields

$$B_0 = (Q_0 C' + B D') (D')^{-1} \quad (3.9c)$$

where Q_0 satisfies the matrix Riccati equation

$$\begin{cases} \dot{Q}_0 = A Q_0 + Q_0 A' - (Q_0 C' + B D') R^{-1} (Q_0 C' + B D')' + B B' \\ Q_0(0) = \Pi - \Pi_0 \end{cases} \quad (3.9d)$$

To see this just insert (3.9c) into the equation (1.3) corresponding to P_0 and subtract from (1.3). Formally (3.9) looks precisely like the Kalman-Bucy filtering equations (3.2). In fact, the differential equations are the same, only the initial conditions differ. However, note that, unlike Q_* , Q_0 is in general indefinite due to the definition of $Q_0(0)$. In view of the fact that $\xi_0 \in H(dy)$, (3.9a) implies that $H(x_0) \subset H(dy)$. We shall call a realization $S \in S$ satisfying the condition $H(x) \subset H(dy)$ *internal*; if

$H(x) \notin H(dy)$ we shall say that S is *external* [20]. Hence we have shown that all $S \in S_0$ are internal. In Section 4 we shall see that, if $\begin{pmatrix} B \\ D \end{pmatrix}$ is full rank, S_0 is precisely the class of all internal realizations. We shall also see that S_0 is a partially ordered set with a smallest element and that it can be slightly extended to also contain a largest element.

Our next task is to establish a backward counterpart $\bar{S}$ to each realization $S \in S$. We shall begin by restricting our attention to the subclass S_+ of all realizations $S \in S$ for which $\Pi > 0$.

LEMMA 3.2. *The class S_+ is nonempty.*

Proof. It is shown in [44] that, since y is generated by the model (1.1), for some $\epsilon > 0$ the covariance function of y can be continuously extended to the interval $[0, T + \epsilon]$ while retaining its nonnegativity property and its "lumped" character. It is not hard to modify the proof of [44, Appendix II] to show that a similar extension, which also preserves analyticity, can be made to the interval $[-\epsilon, T]$ for some $\epsilon > 0$. Hence, by the main result of [44], there is an (analytic) realization S_ϵ of y on $[-\epsilon, T]$ with state-dimension n . Since its restriction to $[0, T]$ belongs to S , it is minimal. Therefore (A, B) corresponding to S_ϵ is totally controllable [40], and consequently $P(0) > 0$ by the argument of Lemma 2.2. Hence the restriction of S_ϵ to $[0, T]$ belongs to S_+ . $\square$

Let $S \in S_+$. Then, by Lemma 2.3, $\bar{x} = P^{-1}x$ is defined on all of $[0, T]$ and satisfies (2.9) there. Inserting (2.10) into (1.1b) yields $dy = (CP + DB')\bar{x}dt + Dd\bar{w}$, so in view of Lemma 3.1 we have obtained a backward model for y on $[0, T]$, namely

$$(\bar{S}) \begin{cases} d\bar{x} = -A'\bar{x}dt + \bar{B}d\bar{w} & ; \quad \bar{x}(T) = \bar{\xi} \\ dy = G'\bar{x}dt + Dd\bar{w} \end{cases} \quad (3.10)$$

where $\bar{\xi} = P(T)^{-1}x(T) \perp H(d\bar{w})$ and $\bar{B} = P^{-1}B$. Its state covariance function $\bar{P} = P^{-1}$ satisfies (2.11). We shall call any model of type (3.10) with y as its output, $\bar{\xi} \in H$, $\bar{w}$ satisfying (1.2) and $\bar{\xi} \perp H(d\bar{w})$ a *backward realization* of y . In view of Lemma 2.2, $\bar{S}$ is also *analytic* (i.e., A , $\bar{B}$, G , D and R^{-1} are analytic). Note that S and $\bar{S}$ have the same *state space*, i.e.,

$$H_t(\bar{x}) = H_t(x), \quad (3.11)$$

for each $t \in [0, T]$.

By symmetry with the forward setting we can now see that

$$\bar{x}_*(t) = \hat{E}\{\bar{x}(t) | H_t^+(dy)\} \quad (3.12)$$

is generated by the *backward Kalman-Bucy filter*

$$d\bar{x}_* = -A'\bar{x}_*dt + \bar{B}_*R^{-1/2}(dy - G'\bar{x}_*dt) ; \quad \bar{x}_*(T) = 0, \quad (3.13a)$$

where $\bar{B}_* = -(\bar{Q}_*G - \bar{B}D')R^{-1/2}$, and the error covariance $\bar{Q}_*(t) := E\{[\bar{x}(t) - \bar{x}_*(t)][\bar{x}(t) - \bar{x}_*(t)]'\}$ satisfies

$$\begin{cases} \dot{\bar{Q}}_* = -A'\bar{Q}_* - \bar{Q}_*A + (\bar{Q}_*G - \bar{B}D')R^{-1}(\bar{Q}_*G - \bar{B}D')' - \bar{B}\bar{B}' \\ \bar{Q}_*(T) = P(T)^{-1}, \end{cases} \quad (3.13b)$$

and that the *backward innovation process* $\{\bar{w}_*(t); 0 \leq t \leq T\}$, given by

$$d\bar{w}_* = R^{-1/2}(dy - G'\bar{x}_*dt) \quad (3.14)$$

has orthogonal increments and satisfies (1.2) and $H_t^+(d\bar{w}_*) = H_t^+(dy)$ for all $t \in [0, T]$. (see [20, 45].) Hence the covariance function $\bar{P}_*(t) := E\{\bar{x}_*(t)\bar{x}_*(t)'\}$ satisfies

$$\dot{\bar{P}}_* = -A'\bar{P}_* - \bar{P}_*A - \bar{B}_*\bar{B}_*' ; \quad \bar{P}_*(T) = 0. \quad (3.15)$$

The following lemma ensures the invariance of the backward filter (3.13a).

LEMMA 3.3. The gain function $\bar{B}_*$ is uniquely determined by the four (invariant) matrix functions A , C , G , and R , i.e., $\bar{B}_*$ is invariant for S .

Proof. Since $\bar{Q}_* = \bar{P} - \bar{P}_*$, it follows from Lemma 3.1 that $\bar{B}_* = (C' - \bar{P}_*G)R^{-1/2}$, which inserted into (3.15) yields an equation for $\bar{P}_*$ which only depends on A , C , G and R . Hence the same holds for $\bar{B}_*$. $\square$

Now define $\bar{S}$ to be the class of all analytic backward realizations $\bar{S}$ having (3.13a) as its backward Kalman-Bucy filter, and let $\bar{S}_+$ be the subclass consisting of those $\bar{S} \in \bar{S}$ for which $\bar{\Pi} > 0$. In the same way as in the forward setting it is seen that the realization

$$(\bar{S}_*) \begin{cases} d\bar{x}_* = -A'\bar{x}_*dt + \bar{B}_*d\bar{w}_*; & \bar{x}_*(T) = 0 \\ dy = G'\bar{x}_*dt + R^{1/2}d\bar{w}_* \end{cases} \quad (3.16)$$

belongs to $\bar{S}$. The state covariance function $\bar{P}_*$ of $\bar{S}_*$ is given by (3.15). By Lemma 3.3 the class $\bar{S}$ is uniquely defined in terms of the invariants A , C , G and R , and therefore the backward counterpart $\bar{S}$ of any $S \in S_+$ belongs to $\bar{S}$. In particular, since $P(T)$ is finite and positive definite, $\bar{S} \in \bar{S}_+$. Obviously there is a complete symmetry between the forward and the backward settings; all results of this section have backward versions obtained by merely starting from a minimal backward realization instead. Consequently all realizations in $\bar{S}$ are minimal. (Indeed, were this not the case, we could, by reversing the procedure above, construct a forward realization of dimension less than n , contradicting our original minimality assumption.) Moreover, it is not hard to see that the two subclasses S_+ and $\bar{S}_+$ are in one-to-one correspondence.

In order to extend the one-one correspondence between forward and backward realizations beyond S_+ and $\bar{S}_+$ we shall have to enlarge the classes S and $\bar{S}$ slightly in the following way. Let $\hat{S}$ be the class of all ana-

lytic realizations (1.1) of y defined on $[0, T - \epsilon]$ for any $\epsilon > 0$ and having (3.2a) for its Kalman-Bucy filter on each such interval, and let $\hat{S}_0$ be the corresponding extension of S_0 . The classes $\hat{S}$ and $\hat{\bar{S}}_0$ are defined analogously with respect to (3.13a) and all intervals $[\epsilon, T]$. We shall call the elements of $\hat{S}$ and $\hat{\bar{S}}$ *generalized realizations*. Clearly $S \subset \hat{S}$ and $\bar{S} \subset \hat{\bar{S}}$. Then the forward-backward construction above can be redone in the light of Lemma 2.3 to yield the following theorem, which also summarizes some of the pertinent facts on this topic.

THEOREM 3.1. *To each realization (1.1) in S there corresponds a generalized backward realization (3.10) in $\hat{S}$ 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 (3.10) 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 $dw = d\bar{w} + \bar{B}'\bar{P}^{-1}\bar{x}dt$. For each such pair $(S, \bar{S})$ of forward and backward (generalized) realizations, relation (3.11) holds for each t for which both S and $\bar{S}$ are defined.*

Since $\bar{P}_*(T) = 0$, the backward filter realization $\bar{S}_*$ has a forward counterpart only in this generalized sense, and it has the form

$$(S^*) \begin{cases} dx^* = Ax^*dt + B^*dw^* & ; \quad x^*(0) = \bar{P}_*(0)^{-1}\bar{x}_*(0) \\ dy = Cx^*dt + R^{1/2}dw^* \end{cases} \quad (3.17)$$

with state covariance function $P^* = \bar{P}_*^{-1}$ satisfying

$$\dot{P}^* = AP^* + P^*A' + B^*B^{*'} \quad ; \quad P^*(0) = \bar{P}_*(0)^{-1} \quad (3.18)$$

on $[0, T)$. Obviously $P^*(t) \rightarrow \infty$ as $t \rightarrow T$. [Note that $B^*B^{*'}$ is not integrable on $(0, \infty)$.] The following lemma explains the "super star" notation.

LEMMA 3.4. Let P be the state covariance function of a realization $S \in \mathcal{S}$. Then

$$P_*(t) \leq P(t) \leq P^*(t) \quad (3.19)$$

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

Proof. Since $Q_*(t)$ is a covariance matrix, $Q_*(t) \geq 0$. But $Q_* = P - P_*$, and therefore $P(t) \geq P_*(t)$. An analogous argument in the backward setting yields $\bar{P}(t) \geq \bar{P}_*(t)$, i.e., $P(t)^{-1} \geq P^*(t)^{-1}$, from which $P(t) \leq P^*(t)$ follows. $\square$

Relation (3.19) induces a partial ordering of $\hat{\mathcal{S}}$, S_* being the smallest and S^* the largest element; the same holds for $\hat{\mathcal{S}}_0$, for both S_* and S^* belong to this subclass. (It can be shown that $\mathcal{S}$ and $\mathcal{S}_0$ have lattice structures, but this goes beyond the scope of this paper.) Since $S^* \in \hat{\mathcal{S}}_0$, x^* satisfies a Kalman-Bucy type equation

$$dx^* = Ax^*dt + B^*R^{-1/2}(dy - Cx^*dt) ; \quad x^*(0) = \xi^*, \quad (3.20a)$$

where $\xi^* = \bar{P}_*(0)^{-1}\bar{x}_*(0)$, and B^* can be determined from any other realization $S \in \mathcal{S}$ through equations (3.9c, d), setting $B^* = B_0$ and $Q_0 = \Pi - \Pi^*$. The corresponding solution Q_0 of the matrix Riccati equation (3.9d) is, in view of (3.9b), $Q_0 = P - P^*$, which is nonpositive definite (Lemma 3.4). For the smoothing problem it will be more convenient to express B^* in terms of a nonnegative definite solution of (3.9d) instead, and therefore we define $Q^* := -Q_0$, i.e.,

$$Q^* = P^* - P, \quad (3.20b)$$

in terms of which (3.9c, d) yields

$$B^* = -(Q^*C' - BD')R^{-1/2} \quad (3.20c)$$

with Q^* satisfying the matrix Riccati equation

$$\begin{cases} \dot{Q}^* = AQ^* + Q^*A' + (Q^*C' - BD')R^{-1}(Q^*C' - BD')' - BB' \\ Q^*(0) = \Pi^* - \Pi, \end{cases} \quad (3.20d)$$

where $\Pi^* = \bar{P}_*(0)^{-1}$. Clearly $Q^*(t) \rightarrow \infty$ as $t \rightarrow T$. The definition (3.20b) enables us to interpret Q^* as an error covariance function, much in analogy with the Kalman-Bucy filter. In fact,

$$Q^*(t) = E\{[x(t) - x^*(t)][x(t) - x^*(t)]'\} \quad (3.21)$$

for all $t \in [0, T]$. This is an immediate consequence of the lemma, which we shall also need in §4.

LEMMA 3.5. *Let x be the state process and P the state covariance function of any realization in S . Then*

$$E\{x(t)x_*(t)'\} = P_*(t), \quad E\{x(t)x^*(t)'\} = P(t) \quad (3.22)$$

and

$$E\{[x(t) - x_*(t)][x^*(t) - x(t)]'\} = 0. \quad (3.23)$$

Proof. In view of the definition (3.1), $H_t(x - x_*) \perp H_t(x_*)$ and therefore the first of relations (3.22) follows. The analogous relation in the backward setting reads $E\{\bar{x}(t)\bar{x}_*(t)'\} = \bar{P}_*$. Hence $E\{x(t)x^*(t)'\} = PE\{\bar{x}(t)\bar{x}_*(t)'\}P^* = P$, for $\bar{x} = P^{-1}x$ and $P^* = \bar{P}_*^{-1}$. Then (3.23) is an immediate consequence of (3.22). $\square$

In §4 we shall need to invert both $Q_*(t)$ and $Q^*(t)$ for arbitrary $t \in [0, T]$. This is possible for all realizations $S \in S$ 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 $\text{int } S$.

LEMMA 3.6. *The interior of S is nonempty.*

Proof. Let Q_* be the error covariance (3.2c) corresponding to realization $S \in S_+$. Then $Q_*(0) = \Pi > 0$. A simple reformulation of (3.2d) yields

$$\dot{Q}_* = \Gamma_* Q_* + Q_* \Gamma_*' + (B_* R^{-1/2} D - B)(B_* R^{-1/2} D - B)', \quad (3.24)$$

where Γ_* is the *feedback matrix*

$$\Gamma_* = A - B_* R^{-1/2} C \quad (3.25)$$

of the Kalman-Bucy filter (3.2). The Liapunov type equation (3.24) can be integrated to yield an expression of the same general form as (2.6). From this it is seen that $Q_*(0) > 0$ implies that $Q_*(t) > 0$ for all $t \in [0, T]$. It remains to show that $Q^*(t) > 0$ for all $t \in [0, T]$. To this end first note that the corresponding backward realization $\bar{S}$ belongs to $\bar{S}_+$; this is clear from the discussion leading to Theorem 3.1. Then we can repeat the argument above to see that $\bar{Q}_*(t) > 0$ for all $t \in [0, T]$. But $Q^* = P(\bar{P} - \bar{P}_*)P^* = P\bar{Q}_*P^*$. Since $P > 0$ and $P^* > 0$ on $[0, T]$, $Q^*(t) > 0$ for all $t \in [0, T]$. $\square$

COROLLARY 3.6.1. *Let $Q = P^* - P_*$. Then $Q(t) > 0$ for all $t \in [0, T]$.*

COROLLARY 3.6.2. *Let $S \in S_+$. Then $Q_*(t) > 0$ for all $t \in [0, T]$.*

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 $S \in S$. To this end first note that (3.1) can be written

$$\hat{E}\{H_t(x) | H_t^-(dy)\} = H_t(x_*), \quad (3.26)$$

and that (3.11) and (3.12) yield

$$\hat{E}\{H_t(x) | H_t^+(dy)\} = H_t(x^*) \quad (3.27)$$

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

$$H(dy) = N_t^- \oplus H_t^\square \oplus N_t^+ \quad (3.28)$$

where $H_t^\square$ is the frame space

$$H_t^\square = H_t(x_*) \vee H_t(x^*) \quad (3.29)$$

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

LEMMA 3.7. (cf. [27]) Let x be the state process of a realization in 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 $H_t(x) \perp N_t^+$. $\square$

4. THE SMOOTHING PROBLEM

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

$$\hat{x}(t) = \hat{E}\{x(t) | H(dy)\} \quad (4.1)$$

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.,

$$\Sigma(t) = E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\}. \quad (4.2)$$

Of course this problem is interesting only if the realization S is external. However, by not restricting our analysis to external realizations, as a by-product we shall obtain some interesting results on internal models also.

In view of Lemma 3.7, $\hat{x}(t) \in H_t^D$, and consequently there are two matrix functions K_* and K^* such that

$$\hat{x}(t) = K_*(t)x_*(t) + K^*(t)x^*(t). \quad (4.3)$$

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 3.5, the first of these relations yields $P_* = K_*P_* + K^*P_*$ and consequently

$$K_*(t) + K^*(t) = I \quad (4.4)$$

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

$$P(t) = K_*(t)P_*(t) + K^*(t)P^*(t) \quad (4.5)$$

for all $t \in [0, T]$. Then solving (4.4) and (4.5) 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 $t \in [0, T]$ (Corollary 3.6.1) and that

$$Q(t) = Q_*(t) + Q^*(t). \quad (4.6)$$

THEOREM 4.1. *Let x be the state process of a realization (1.1) of class S . Then the smoothing estimate (4.1) is given by*

$$\hat{x}(t) = [I - Q_*(t)Q(t)^{-1}]x_*(t) + Q_*(t)Q(t)^{-1}x^*(t) \quad (4.7)$$

and the error covariance function (4.2) by

$$\Sigma(t) = Q_*(t) - Q_*(t)Q(t)^{-1}Q_*(t) \quad (4.8)$$

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

Proof. Relation (4.7) was derived above for $t \in (0, T)$; for $t = 0$ (4.7) follows from (4.19) below. To prove (4.8) note that

$$x - \hat{x} = (I - Q_* Q^{-1})(x - x_*) + Q_* Q^{-1}(x - x^*). \quad (4.9)$$

By Lemma 3.5 the two terms of (4.9) are orthogonal and therefore, observing (3.2c) and (3.21),

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

which, in view of (4.6), yields (4.8). $\square$

Relation (4.5) should be compared with the decomposition in [46, Theorem 6]. Note however that $K_*(t)$ and $K^*(t)$ are projections if and only if the realization S is internal. To see this observe that $(K^*)^2 = K^*$, i.e., $Q_* Q^{-1}Q_* = Q_*$, if and only if $\Sigma = 0$ (Theorem 4.1).

Theorem 4.1 is a generalization of results given in [20-22]. Following the procedure in [22] we obtain an alternative derivation by observing that $x_*(t)$ and

$$z(t) = x^*(t) - x_*(t) \quad (4.10)$$

are orthogonal (to see this, note that $x_*(t) = \hat{E}\{x^*(t) | H_t^-(dy)\}$) and applying Lemma 2.1. In fact, since $\hat{x}(t) = \hat{E}\{x(t) | H_t^\square\}$ (Lemma 3.7) and $H_t^\square = H_t(x_*) \oplus H_t(z)$,

$$\hat{x}(t) = E\{x(t) | x_*(t)\} + E\{x(t) | z(t)\}. \quad (4.11)$$

Then using Lemmas 2.1 and 3.5 and the fact that

$$Q(t) = E\{z(t)z(t)'\} \quad (4.12)$$

we obtain

$$\hat{x}(t) = x_*(t) + Q_*(t)Q(t)^{-1}z(t), \quad (4.13)$$

which is precisely (4.7).

If, for the moment, we restrict our attention to realizations in the interior of S we obtain the following well-known result.

COROLLARY 4.1. *Let $S \in \text{int } S$, let x be the state process of S , and let $\hat{x}$ be the corresponding smoothing estimate (4.1). Then, for each $t \in [0, T)$*

$$\hat{x}(t) = \Sigma(t)[Q_*(t)^{-1}x_*(t) + Q^*(t)^{-1}x^*(t)], \quad (4.14)$$

where x_* and x^* are given by (3.2) and (3.20) respectively and the smoothing error covariance Σ by

$$\Sigma(t)^{-1} = Q_*(t)^{-1} + Q^*(t)^{-1}. \quad (4.15)$$

Proof. Since $S \in \text{int } S$, Q_* and Q^* are invertible. By writing (4.8) as $\Sigma = Q_*Q^{-1}(Q - Q_*)$ and using (4.6), it is seen that

$$\Sigma = Q_*Q^{-1}Q^*. \quad (4.16)$$

Inverting this and again using (4.6) yields (4.15). From (4.16) we also see that $Q_*Q^{-1} = \Sigma(Q^*)^{-1}$. Then $I - Q_*Q^{-1} = \Sigma[\Sigma^{-1} - (Q^*)^{-1}] = \Sigma Q_*^{-1}$. Hence (4.14) follows from (4.7). $\square$

Relations (4.14) and (4.15) together with (3.2) and (3.20) 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], its probabilistic justification has caused considerable difficulty, partly due to the fact that $Q^*(t) \rightarrow \infty$ as $t \rightarrow T$. The system (3.20) 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 (3.20) has a very

natural interpretation: It is simply the maximum-variance *forward* realization S^* . By using the identity

$$x^*(t) = \bar{P}_*(t)^{-1} \bar{x}_*(t) \quad (4.17)$$

we can instead write the smoothing formula (4.14) in terms of two Kalman-Bucy filters, one (3.2) evolving forward and the other (3.13) evolving backward in time. (Note that then (4.14) is defined on the whole interval $[0, T]$.) This fact was pointed out in [14, 15, 17], in which papers the backward estimate

$$\hat{x}_b(t) = \hat{E}\{x(t) | H_t^+(dy)\} \quad (4.18)$$

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

$$\hat{x}_b(t) = P(t)P^*(t)^{-1}x^*(t) \quad (4.19)$$

is not invariant over S and is therefore less suitable for our purposes.

It is not hard to see that

$$(Q^*)^{-1} = [(Q^*)^{-1} + P^{-1}]P(P^*)^{-1} \quad (4.20)$$

and consequently (4.14) may also be written

$$\hat{x}(t) = \Sigma(t)\{Q_*(t)^{-1}x_*(t) + [Q^*(t)^{-1} + P(t)^{-1}]\hat{x}_b(t)\}, \quad (4.21)$$

which is the formula presented in [14, 15, 17]. The partitioned smoothing formula [12, 13] also can be seen to be equivalent to (4.14), and it can be used to derive all the equations of the Mayne-Fraser procedure. In the early papers [7, 8], relation (4.14) 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. (A more satisfactory treatment has recently been presented in [48].) However, the stochastic realization theory provides a natural justification of this procedure. Indeed, (3.23) is the required orthogonality condition.

The smoothing formulas (4.7) and (4.14) are both based on the nonorthogonal decomposition (3.29), whereas (3.13) corresponds to the orthogonal decomposition

$$H_t^{\square} = H_t(x_*) \oplus H_t(z) \quad (4.22)$$

(where, in either case, Lemma 3.7 justifies the restriction to the finite dimensional frame space $H_t^{\square}$). We shall now take a closer look at representations of the latter type. It follows from (3.4) and (3.17) that z as defined by (4.10) is the solution of

$$dz = \Gamma_* z dt - QC'R^{-1/2} dw^*; \quad z(0) = x^*(0), \quad (4.23)$$

where Γ_* is the feedback matrix (3.25) of the Kalman-Bucy filter (3.2). To see this, note that the input process w^* of the maximum variance realization S^* is related to the innovation process w_* through the relation

$$dw_* = R^{-1/2} C z dt + dw^* \quad (4.24)$$

and that $B^* - B_* = -QC'R^{-1/2}$. We shall need the backward counterpart of (4.23). Observing that Q is the covariance function of z , Lemma 2.3 yields the following equation for $\bar{z} = Q^{-1}z$:

$$d\bar{z} = -\Gamma_*' \bar{z} dt - C'R^{-1/2} dw_*; \quad \bar{z}(T) = 0, \quad (4.25)$$

for, in view of (4.24), w_* is the backward counterpart of w^* with respect to (4.23). Note that $\bar{z}$ is defined on the whole interval $[0, T]$. The covariance matrix $\bar{Q} = Q^{-1}$ of $\bar{z}$ satisfies

$$\dot{\bar{Q}} = -\Gamma_*' \bar{Q} - \bar{Q} \Gamma_* - C'R^{-1}C; \quad \bar{Q}(T) = 0. \quad (4.26)$$

The estimate $\hat{x}$ is then obtained from (4.13).

THEOREM 4.2. *Let x be the state process of an arbitrary realization in S . Then the smoothing estimate $\hat{x}(t)$ satisfies*

$$\hat{x}(t) = x_*(t) + Q_*(t) \bar{z}(t) \quad (4.27)$$

for all $t \in [0, T]$, where x_* is given by (3.2) and $\bar{z}$ by (4.25) and (3.3). The process $\bar{z}$ is related to x_* and x^* through the relation

$$\bar{z}(t) = \bar{Q}(t)[x^*(t) - x_*(t)] \quad \text{for } t \in [0, T]. \quad (4.28)$$

Relation (4.27) is the smoothing formula of Bryson and Frazier [2]. (Also see [3, 4] and, in particular, [9].) What is new here is its interpretation (4.28) in terms of the minimum and maximum variance realizations S_* and S^* . Theorem 4.2 can also be regarded as a generalization of a result presented in [21], and the basic techniques used there provide an alternative approach to deriving the above result.

COROLLARY 4.2. *The smoothing estimate (4.27) satisfies the stochastic differential equation*

$$d\hat{x} = A\hat{x}dt + B(I - D'R^{-1}D)B'\bar{z}dt + BD'R^{-1}(dy - C\hat{x}dt). \quad (4.29)$$

with initial condition $\hat{x}(T) = x_*(T)$. If $S \in S_+$, $\bar{z}$ can be replaced by $Q_*^{-1}(\hat{x} - x_*)$ in (4.29).

Proof. Inserting (3.2a), (3.2d) and (4.25) into

$$d\hat{x} = dx_* + Q_*d\bar{z} + \dot{Q}_*\bar{z}dt$$

and using (3.2b) yields (4.29). If $S \in S_+$, Q_*^{-1} exists (Corollary 3.6.2), and (4.27) can be solved for $\bar{z}$. $\square$

We shall now study two different special cases of (4.29). First, let $BD' = 0$; this is a standard assumption in the smoothing literature. Then $\hat{x}$ is differentiable, and (4.29) reduces to

$$\frac{d\hat{x}}{dt} = A\hat{x} + BB'\bar{z}; \quad \hat{x}(T) = x_*(T). \quad (4.30)$$

For realizations $S \in S_+$ (4.30) reduces to the smoothing formula of Rauch, Tung and Striebel [3]

$$\frac{d\hat{x}}{dt} = A\hat{x} + BB'Q^{-1}(\hat{x} - x_*) ; \quad \hat{x}(T) = x_*(T). \quad (4.31)$$

Secondly, assume that D is square. Then D is full rank and $D'R^{-1}D = I$.

Hence

$$d\hat{x} = A\hat{x}dt + BD^{-1}(dy - C\hat{x}dt) ; \quad \hat{x}(T) = x_*(T), \quad (4.32)$$

which defines a realization in S_0 . Note that the original realization S need not be internal; it may have an initial condition $x(0) \notin H(dy)$.

The problem of smoothing can be regarded as that of finding the "internal part" of the state process. Given a realization $S \in S_+$, we shall next look at the structure of the "external part," i.e., the smoothing error $\tilde{x} := x - \hat{x}$. To this end, first 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

$$\begin{bmatrix} B(t) \\ D(t) \end{bmatrix} = \begin{bmatrix} B_1(t) & B_2(t) \\ R^{1/2}(t) & 0 \end{bmatrix} V(t), \quad (4.33a)$$

where B_1 is $n \times m$ and B_2 is $n \times (p - m)$. Next let

$$\begin{bmatrix} du \\ dv \end{bmatrix} = V dw \quad (4.33b)$$

define a pair of orthogonal increment processes u and v , of dimensions m and $p - m$ respectively. Obviously (4.33b) satisfies (1.2).

THEOREM 4.3. Let x be the state process of a realization $S \in S_+$ and let B_2 and v be defined by (4.33). Then the smoothing error $\tilde{x}$ is given by

$$\tilde{x}(t) = Q_*(t)\eta(t) \quad (4.34a)$$

$$\begin{cases} d\eta = -\Gamma_*'\eta dt + Q_*^{-1}B_2 dv \\ \eta(T) = \eta_T \end{cases} \quad (4.34b)$$

where $\eta_T = Q_*^{-1}(T)[x(T) - x_*(T)]$ and ζ is a $(p-m)$ -dimensional orthogonal increment process of type (1.2) such that $H(d\zeta) \perp H(dy)$. The representation (4.34) is a backward realization in the sense that $\eta_T \perp H(d\zeta)$ and the increments of ζ are given by

$$d\zeta = dv - B_2' Q_*^{-1}(x - x_*) dt. \quad (4.35)$$

Proof. Define $z_* := x - x_*$. Replacing $B dw$ and $D dw$ in (1.1) by $B_1 du + B_2 dv$ and $R^{1/2} du$ respectively and noting that the innovation process w_* in (3.4) is given by

$$dw_* = du + R^{-1/2} C z_* dt \quad (4.36)$$

and that $B_1 - B_* = -Q_* C' R^{-1/2}$ (Lemma 3.1), it is just a matter of simple calculations to see that z_* satisfies

$$dz_* = \Gamma_* z_* dt - Q_*^{-1} C' R^{-1/2} du + B_2 dv; \quad z_*(0) = \xi.$$

for, since $S \in S_+$, $Q_*(t)^{-1}$ exists for all $t \in [0, T]$ (Corollary 3.6.2). By Lemma 2.5 and (4.36), $\bar{z}_* = Q_*^{-1} z_*$ satisfies the backward Markovian representation

$$d\bar{z}_* = -\Gamma_*' \bar{z}_* dt - C' R^{-1/2} dw_* + Q_*^{-1} B_2 d\zeta; \quad \bar{z}_*(T) = \eta_T \quad (4.37)$$

where ζ is given by (4.35). Since $H(d\zeta) \perp H(dw_*)$ (by construction) and $H(dw_*) = H(dy)$, $H(d\zeta) \perp H(dy)$ as required. Now, in view of (4.27), $\tilde{x} = z_* - Q_* \bar{z}_*$, i.e., (4.34a) holds with $\eta := \bar{z}_* - \bar{z}$. Then (4.34b) follows from (4.25) and (4.37). $\square$

As a corollary we see that the state process of any realization $S \in S_+$ can be decomposed into three orthogonal terms

$$x(t) = x_*(t) + Q_*(t) \bar{z}(t) + Q_*(t) \eta(t), \quad (4.38)$$

each of which is the output of a stochastic system whose dynamical behavior is determined by the function Γ_* . This is seen from (4.25), (4.34) and the fact that (3.2a) can be written

$$dx_* = \Gamma_* x_* dt + B_* R^{-1/2} dy \quad ; \quad x_*(0) = 0. \quad (4.39)$$

[Note that both (4.25) and (4.34b) are backward representations. If we transform to the forward setting the systems matrices will be Γ_* rather than $-\Gamma'_*$.]

The internal realizations play an interesting role in the theory of smoothing. These are precisely the representations (1.1) for which the smoothing problem is trivial, i.e., $\tilde{x} \equiv 0$. The next theorem shows that (subject to a mild regularity condition) $S \in \mathcal{S}$ is internal if and only if $B_2 \equiv 0$ and the components of the initial conditions of (1.1) belong to $H(dy)$. In view of Theorem 4.3, this is to be expected.

THEOREM 4.4. *A realization $S \in \mathcal{S}$ such that $\begin{pmatrix} B \\ D \end{pmatrix}$ has full rank is internal if and only if $S \in \mathcal{S}_0$.*

Proof. We only need to prove the "only if" part; the "if" part was proved in Section 3. Let S be internal. Since the condition $x(0) \in H(dy)$ holds trivially, it only remains to show that B_2 , as defined by (4.33a), is identically zero. In view of the fact that $x = \hat{x}$, comparing (1.1) and (4.29) shows that the identity

$$B(I - D'R^{-1}D)B'\tilde{z}dt + BD'R^{-1}Ddw = Bdw \quad (4.40)$$

must hold. It is not hard to see that $BD'R^{-1}D = (B_1, 0)V$ and $B(I - D'R^{-1}D)B' = B_2B_2'$, and therefore (4.40) can be written

$$B_2B_2'\tilde{z}dt = B_2dv$$

which cannot hold unless $B_2 = 0$. Then the full rank condition implies that $p = m$. $\square$

REFERENCES

1. J. L. Doob, *Stochastic Processes*, Wiley, 1953.
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 likelihood 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), 1137-1148.
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. 1re 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 Appliquées, Ecole Polytechnique, Internal Report No.41, Nov. 1978.

34. B. D. O. Anderson, The inverse problem of stationary covariance generation, *J. Statistical Physics*, 1(1969), 133-147.
35. P. Faurre, Réalisations 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, Algorithmes 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. Automatic Control* 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. Automatic Control* 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. Control*, Vol.12, (1974), 747-754.
46. J. C. Willems, Least squares stationary optimal control and the algebraic Riccati equation, *IEEE Trans. Automatic Control* AC-16(1971), 621-684.
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).

FOOTNOTES

1. Some of these shortcomings have been pointed out in a recent thesis by Wall [48], brought to our attention after the submission of this paper.
2. e.g., the Moore-Penrose pseudo-inverse can be used.
3. It is not hard to see that the concept of minimality used here is equivalent to assuming both that (i) the input-output map of (1.1a) is minimal *and* that (ii) the family of state spaces $\{H_t(x); t \in [0, T]\}$ is minimal in the sense of the geometric state space theory outlined in [27].