AD 694482

FRANK J. SEILER RESEARCH LABORATORY

SRL 69—0010

JUNE 1969

APPLICATIONS OF THE CANONICAL REPRESENTATION TO ESTIMATION AND DETECTION IN COLORED NOISE

MAJOR ROGER A. GEESEY
PROFESSOR THOMAS KAILATH

THIS DOCUMENT HAS BEEN APPROVED FOR PUBLIC RELEASE AND SALE; ITS DISTRIBUTION IS UNLIMITED.

D D C
OCT 1 0 1969

OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
PROJECT 7904

CLEARINGHOUSE

43

SRL 69—0010

JUNE 1969

APPLICATIONS OF THE CANONICAL REPRESENTATION TO ESTIMATION AND DETECTION IN COLORED NOISE

MAJOR ROGER A. GEESEY
PROFESSOR THOMAS KAILATH

PROJECT 7904

APPLICATIONS OF THE CANONICAL REPRESENTATION TO ESTIMATION

AND DETECTION IN COLORED NOISE

(SHORT TITLE - CANONICAL REPRESENTATIONS)

by

R. A. Geesey
F. J. Seiler Research Laboratory, USAF Academy, Colo., 80840

T. Kailath
Department of Electrical Engineering, Stanford University,
Stanford, California, 94305

ABSTRACT

The representation of a random process as the output of a causal and causally invertible linear system driven by white noise is called canonical and specifies, quite simply, the whitening filter for the process. Whitening filter techniques replace the observation process, without loss of information, by a white noise process and allow simple formulation of the solutions of estimation and detection problems in terms of the equivalent process obtained by the whitening. Constructive methods based on the solution of a matrix Riccati equation are given for determining the canonical representation of differentiable observation processes which consist of a linear combination of the component processes of a finite dimensional Markov process. Implementation of filtering solutions and likelihood ratios for detection are then obtained in a common formulation for a variety of signal with colored noise situations. The approach emphasizes the canonical representation of the observation process while requiring a minimum of attention to models for signal and noise components of the observation. Finite time interval problems for differentiable processes require attention to "initial condition" random variables and the solutions discussed account for their contribution in a natural way.

Presented at the Symposium on Computer Processing in Communications
Polytechnic Institute of Brooklyn, April 8, 9, 10, 1969

1. Introduction

The three problems of detection, estimation, and covariance factorization are intimately related and this relationship has become clear for processes which are the sum of a smooth signal process and a white noise process such that the cross-covariance function of the signal and noise is of a particular "one-sided" form [1], [2], [3]. A unified formulation for the three problems in such a situation arises from the occurrence of a common Wiener-Hopf integral equation involving the covariance function of the observed process. When the kernel function of the covariance has a separable form, the solution of the Wiener-Hopf integral equation may be obtained in terms of an associated matrix Riccati equation. This leads to effective digital computational schemes for solving these three important problems in the analysis of stochastic processes.

In this paper we will show how the results obtained for the additive white noise situation may be easily extended to stochastic problems concerned with observed processes which do not contain an additive white noise but which have certain differentiability properties. An innovations approach for differentiable processes is formulated which extends the innovations approach to least-squares filtering developed by Kailath [1] for the additive white noise problem. The innovations approach is to first convert the observation process to a white noise, called the innovation process, by means of a causal and causally invertible linear transformation and then treat the simplified estimation or detection problem based on (derived) white noise observations. This whitening filter approach was applied by Kailath to extend the technique used by Bode and Shannon in the solution of the Wiener filtering problem for stationary processes over a semi-infinite interval. In reference [1], the nonstationary continuous time process over a finite time interval is handled and the Kalman-Bucy recursive filtering formulas are simply derived.

The innovations approach to estimation and detection reflects on the covariance factorization problem which seeks a causal linear filter that when driven by white noise yields an output process with the given covariance (see [4] and [5] and the references cited therein). In general there may be no such representation, but if one exists there is an essentially unique causally invertible representation called the canonical representation (CR). In a previous paper [4], the authors give several examples of these facts and point out that the CR is just the inverse of the whitening filter giving the innovation process. Conversely, then, the implementation of the whitening filter approach to estimation and detection is readily carried out if the canonical representation is obtained as the solution of the covariance factorization problem.

In this paper, we will determine the CR for a class of differentiable observation processes and show solutions for estimation and detection in terms of this representation.

2. The Additive White Noise Case

In this section we will summarize results for an observation process containing additive white noise and show the relation of solutions for estimation, detection and covariance factorization. The details are presented in references [1], [2], and [3].

Consider a process of the form

$$y(t) = z(t) + v(t),\ 0 \le t \le T;\ E[v(t)v(s)] = \delta(t - s) \quad (1)$$

where the cross-covariance function of the signal and noise is of the particular one-sided form

$$E[z(t)v(s)] \equiv 0, \quad s > t \tag{2}$$

Assume that $E[z(t)v(s)]$ is continuous on the triangle $0 \leq s \leq t \leq T$ and that $E[z(t)z(s)]$ is continuous on $0 \leq s,t \leq T$. Let

$$K(t,s) = E[z(t)z(s)] + E[z(t)v(s)] + E[v(t)z(s)] \tag{3}$$

Then the symmetric kernel function $K(t,s)$ is continuous on $0 \leq s,t \leq T$. We will use an obvious integral operator notation in which the covariance function R_y of the process (1) has the form $R_y = I + K$ where I is the identity operator whose kernel function is $\delta(t - s)$ and K is the operator corresponding to the kernel function (3).

The significance of the assumption (2) is that the covariance function $R_y = I + K$ is positive definite and hence that -1 is not an eigenvalue of the integral operator K. This result is established by Kailath in reference [3] where the converse result is also shown. Namely, for every process y_t having covariance function R_y of the form

$$R_y = I + K, \quad -1 \text{ not an eigenvalue of } K, \tag{4}$$

with $K(t,s)$ continuous in t and s, there exists a signal process z_t and white noise process v_t such that the process y_t has the form (1) with properties (2) and (3). In fact, for a canonical representation of the observation process we seek the unique sum decomposition of the form (1) such that the white noise appearing in the representation can be recovered by a causal linear operation on the observation process. We show the solution below.

The solutions for the three problems of estimation, detection and covariance factorization will now be shown in terms of the solution of the Wiener-Hopf integral equation which may be associated with a covariance function of the form (4). Let h denote the Volterra operator whose causal kernel function $h(t,s)$ $[h(t,s) = 0,\ s > t]$ is the solution of

$$h(t,s) + \int_0^t h(t,\tau)K(\tau,s)d\tau = K(t,s)\ ,\ 0 \leq s \leq t \leq T \qquad (5)$$

In operator notation, this W-H equation can be written

$$h + \{hK\}_+ = \{K\}_+ \qquad (6)$$

where $\{A\}_+$ denotes the causal part of an operator A whose kernel function is determined by

$$\{A(t,s)\}_+ = \begin{cases} A(t,s)\ , & s \leq t \\ 0 & t < s \end{cases}$$

Through an identity obtained by Kailath [6] for the Fredholm resolvent of the integral operator K, with continuous kernel function, it may be shown that a necessary and sufficient condition for the W-H equation (5) and (6) to have a solution h with continuous kernel $h(t,s)$ on $0 \leq s \leq t \leq T$ is that -1 is not an eigenvalue of K. This resolvent identity shows that the solution h satisfies

$$I - h - h^* + h^*h = (I - h^*)(I - h) = (I + K)^{-1} \qquad (7)$$

where h^* is the adjoint of h with kernel function

$$h^*(t,s) = \begin{cases} 0 & s \leq t \\ h(s,t) & t < s \end{cases}$$

The <u>estimation problem</u> which consists of determining the linear, least-squares estimate of process z_t given observations $\{y_s, s \le t\}$, where y_t has the form (1) under condition (2), is obtained by an application of the projection theorem (see the discussion by Kailath in [1]). This estimate, denoted by $\hat{z}(t)$ is a causal linear operation on the observation and has the general form

$$\hat{z}(t) = \int_0^t h(t,\tau)y(\tau)d\tau \tag{8}$$

By the projection theorem $(E[(z_t - \hat{z}_t)y_s] = 0 \ , \ s \le t)$, the causal kernel $h(t,\tau)$ satisfies the W-H equation (5) where $K(t,s)$ is given by (3).

It is well known in the theory of integral operators that for h a Volterra operator having continuous kernel function, then

$$(I - h)^{-1} = (I + k) \tag{9}$$

for a uniquely determined Volterra operator k having continuous, causal kernel function $k(t,s)$. Thus, the resolvent identity (7) shows that <u>covariance factorization</u> is achieved through the solution of the W-H equation (6) as

$$R_y = (I + K) = (I - h)^{-1}(I - h^*)^{-1} = (I + k)(I + k^*) \tag{10}$$

But here, the causal factorization of R_y in (10) is indeed causally invertible by the relation (9).

The <u>canonical representation</u> for the process whose covariance function is of the form (4) is then

$$y_t = (I + k)\nu_t = \nu_t + \int_0^t k(t,\tau)\nu_\tau d\tau \tag{11}$$

where ν_t is a white noise process with $E[\nu_t \nu_s] = \delta(t - s)$ since the covariance of the representation (11) is given by (10). The complete association of the CR with the solution of the estimation problem is made by writing (11) as

$$y_t = (I - h)^{-1} \nu_t \tag{12}$$

Then

$$\nu_t = (I - h) y_t = y_t - \hat{z}_t \tag{13}$$

where the causal estimate $\hat{z}_t$ is identified from (8). This relation (13) is the expression of the innovations theorem given by Kailath in [1]. The innovations theorem states that for the observation process (1), under the assumption (2), the process $y_t - \hat{z}_t$, called the innovation process, is a white noise process with the same covariance as v_t process. Thus the CR in (11) determines the innovation process in (13) through the relation (9) and hence determines the solution of the estimation problem as

$$\hat{z}_t = y_t - \nu_t \tag{14}$$

Conversely, we have shown that the causal filter h which gives the solution of the estimation problem determines the CR in (11) through the relation (9).

The Gaussian signal in white Gaussian noise detection problem to be considered is that of observing a Gaussian process y_t , $0 \leq t \leq T$ (of zero mean) and choosing between the hypotheses

$$H_1 : R_y^{(1)} = I + K \text{ , -1 not an eigenvalue of K,}$$
$$K(t,s) \text{ continuous on } 0 \leq t,s \leq T$$
$$H_2 : R_y^{(2)} = I \tag{15}$$

This detection problem is nonsingular as is shown by Shepp (Theorem 1 of [7]). However, under H_1 we can obtain the CR of the observation process and regard this detection problem as being one of signal-in-noise with one-sided cross-correlation of the signal and noise. For the CR given in (11) under H_1, consider

$$z_t = k\nu_t = \int_0^t k(t,\tau)\nu_\tau d\tau$$

Then the detection problem (15) becomes

$$\begin{aligned} H_1 &: y_t = z_t + \nu_t \ , \quad E[\nu_t\nu_s] = \delta(t-s) \\ & \qquad\qquad\qquad\qquad E[z_t\nu_s] \equiv 0 \ , \quad s > t \\ H_2 &: y_t = \nu_t \end{aligned} \tag{16}$$

The likelihood ratio for this signal-in-noise detection problem with one-sided cross-correlation has been obtained by Kailath [3], [8] in an "estimator-correlator" form in terms of the causal linear estimate $\hat{z}_t$ of the signal given the observations under H_1. This formula for the likelihood ratio is

$$\text{L.R.} = \exp\left\{ \int_0^T \hat{z}_t y_t dt - \tfrac{1}{2}\int_0^T \hat{z}_t^2 dt - \tfrac{1}{2}\int_0^T h(t,t)dt \right\} \tag{17}$$

where $h(t,\tau)$ is the kernel function of the causal filter for $\hat{z}_t$. The stochastic integral in the first term of the exponent of (17) is of the Stratonovich form and is well defined for this detection problem with continuous kernel K (see the discussions in [3] and [8]).

But we showed above that such a causal estimate $\hat{z}_t$ is given by the solution h of the W-H equation (6) with kernel function K(t,s) given in the statement (15). Thus the solution (17) for the detection problem (15) is also specified by the solution of the W-H equation (6).

This section has shown that the three problems of estimation, detection and covariance factorization for an observation process with additive white noise are solved in terms of the solution of a common Wiener-Hopf integral equation involving the kernel function K which is the continuous part of the covariance function of the observation process. In the following section we will show that for kernel K of "separable" form, the form of solutions given in this section may be readily implemented in terms of the solution of a matrix Riccati differential equation which may be effectively solved by digital computation.

3. Separable Kernel Function

In engineering applications, the signal process z_t in (1) is typically the output of a finite dimensional linear system driven by white noise which we will call a lumped process. When a particular dynamical model is known for such a lumped process, the Kalman-Bucy [9] recursive solution for the estimation problem is computationally effective and has received widespread application. The treatment in [9] shows that the solution of the Wiener-Hopf integral equation can be obtained by the solution of a matrix Riccati equation formed in terms of the coefficients of the model of the process. However, for some estimation problems and many detection problems, such a model for the signal process is not known and only the covariance function is available. The determination of a lumped model for the process from a covariance function is non-trivial and is the essence of the covariance factorization problem itself.

But Anderson [10] has shown that the covariance function for an observation process (1) with the signal process having a lumped model (which guarantees that condition (2) applies) is of the form

$$R_y(t,s) = \delta(t-s) + a^{\dagger}(t \vee s)\ b(t \wedge s) \tag{18}$$

(where $t \vee s$ denotes maximum of t and s and $t \wedge s$ denotes minimum of t and s) for n-vector functions a(t) and b(t). In the operator notation of Section 2, the form of the covariance (18) is $R_y = I + K$ with "separable" kernel function

$$K(t,s) = a^{\dagger}(t \vee s)\ b(t \wedge s) \tag{19}$$

Further, Anderson shows that covariance factorization for (18) is achieved in terms of the solution of a matrix Riccati equation with coefficients given by the n-vector functions a(t) and b(t). This solution given by Anderson is

$$I + K = (I + k)\ (I + k^{*}) \tag{20a}$$

$$k(t,\tau) = a^{\dagger}(t)\ \psi(\tau), \quad \tau \leq t \tag{20b}$$

$$\psi(t) = b(t) - P(t)\ a(t) \tag{20c}$$

for n x n matrix function P determined by the solution of

$$\dot{P}(t) = [P(t)\ a(t) - b(t)]\ [P(t)\ a(t) - b(t)]^{\dagger}, \quad P(o) = o \tag{21}$$

A more physically revealing account of the occurrence of the Riccati equation (21) is made in [2] where the relation of (21) to the Riccati equation arising in the Kalman-Bucy filter is explored and references are made to other work relating integral equations to Riccati differential equations.

The factorization given in (20a, b, c) is, however, causally invertible (see the discussion of equation (9)) so that the kernel function $k(t,\tau)$ given by (20b) determines the CR for the process y_t. The discussions in Section 2 then show that the solution h of the W-H equation (5), for the separable kernel (19), is determined through the solution of the Riccati equation (21) by

$$h = k(I + k)^{-1} \tag{22}$$

Although we have not simply determined the kernel function $h(t,\tau)$, a realization for the filter having impulse response $h(t,\tau)$ is easily obtained. Let (F, G, H) denote the linear system

$$\dot{x} = Fx + Gu$$
$$y = H^{\dagger}x$$

(with u and y denoting input and output respectively). Then the kernel function (20b) has the realization (o, ψ, a) and a realization for h determined in (22) is

$$h: \quad (-\psi a^{\dagger}, \ \psi, \ a) \tag{23}$$

This reveals our motive for exploring the relationship of the canonical representation to solutions for estimation and detection in Section 2. Through the determination of the CR, we have obtained a realization of the filter h (which determines the solutions for the estimation and detection problems) in terms of the solution P of the Riccati equation (21). Note that from (23)

$h(t,t) = a^{\dagger}(t)\,\psi(t) = k(t,t)$ so that the bias term in the solution (17) for the detection problem is also readily formed.

We point out that if the separable kernel K in (19) does not have eigenvalue -1 and $a(t)$ and $b(t)$ are continuous, the nonlinear differential equation (21) has a well defined solution. This is seen by using the solution h of W-H equation (5), which has a continuous solution for such K, to form

$$\overline{P}(t) = \int_0^t \overline{\psi}(\tau)\overline{\psi}^{\dagger}(\tau)d\tau, \quad \overline{\psi}(t) = b(t) - \int_0^t h(t,\tau)b(\tau)d\tau \tag{24}$$

Then direct calculation, using the W-H equation for h, will show that $\overline{P}(t)$ satisfies the Riccati equation (21). Furthermore, continuity of a, b gives a Lipschitz constant on the solution of (21) so that (24) is the unique solution.

4. Differentiable Processes

In this section we will consider observation processes which are differentiable such that the α^{th} derivative process contains additive white noise. The technique for obtaining solutions for estimation and detection problems by differentiating an observation process a sufficient number of times to recover a derivative process which contains white noise and then treating the underlying additive white noise problem is well used. In particular, Bryson and Johansen [11] use the technique for colored noise filtering problems in order to apply the Kalman-Bucy filter. As in the Kalman-Bucy problem, the Bryson-Johansen procedure is applied to processes for which a lumped model is known. Here, by extension of the results of the previous sections, we will obtain solutions for estimation, detection and covariance factorization which do not require a priori knowledge of a lumped model. It is shown by example in [5] that the form of solution obtained here for filtering may be expressed in terms of the system matrices of a given model to give the solution in the form obtained by the Bryson-Johansen procedure.

In order to perform estimation and detection by utilizing a derivative process it is necessary to properly incorporate the "initial condition" random variables which relate the derivative process to the given observation. Our approach is to perform a decomposition of the observation into a sum of uncorrelated processes such that an innovation process may be separately determined for each. The appropriate decomposition to be used below is shown by Shepp [7].

We will consider an α times differentiable process $y(t)$; $0 \leq t \leq T$ (denoting the i^{th} derivative process by $y_i(t)$) such that

$$R_\alpha(t,s) = E[y_\alpha(t)y_\alpha(s)] = \delta(t - s) + K(t,s) \tag{25}$$

with $K(t,s)$ continuous in (t,s), and such that the covariance matrix $R_Y = E[YY^\dagger]$ of the α vector of random variables

$$Y \triangleq [y(0), y_1(0), \ldots y_{\alpha-1}(0)]^\dagger \tag{26}$$

is nonsingular. A slightly more general form would allow

$$R_\alpha(t,s) = r^2(t)\delta(t - s) + K(t,s)$$

But for continuous $r^2(t) > 0$, $0 \leq t \leq T$, the following innovations approach can be followed by replacing process $y_\alpha(t)$ with the process $\{y_\alpha(t)/r(t)\}$.

The desired uncorrelated sum decomposition of process y_t is obtained by forming the orthogonal projection of random variables y_t onto the family of random variables which comprise the vector Y. Let

$$\hat{y}(t) = E[y(t)/Y] \tag{27a}$$

$$= C^{\dagger}(t)R_Y^{-1}Y \tag{27b}$$

where the α-vector function

$$C(t) = E[Yy(t)] \tag{27c}$$

Then

$$y(t) = \hat{y}(t) + \tilde{y}(t) \tag{28}$$

for $\tilde{y}(t) = y(t) - \hat{y}(t)$, and $E[\hat{y}(t)\tilde{y}(s)] = 0$ for all t,s by the projection theorem.

The component $\tilde{y}(t)$ in the decomposition (28) is said to be "pinned at zero" by Shepp since the derivative processes $\tilde{y}_i(t)$ satisfy

$$\tilde{y}_i(0) = 0 \ , \ i = 0, 1, \cdots (\alpha - 1)$$

with probability one. Then

$$\tilde{y}_{\alpha}(t) = d^{\alpha}\tilde{y}(t)/dt^{\alpha}$$

and

$$\tilde{y}(t) = \int_0^t \frac{1}{(\alpha - 1)!} (t - \tau)^{(\alpha - 1)}\tilde{y}_{\alpha}(\tau)d\tau$$

showing that the processes $\tilde{y}(t)$ and $\tilde{y}_{\alpha}(t)$ can be obtained from each other by causal linear operations. Thus $\tilde{y}(t)$ is equivalent to $\tilde{y}_{\alpha}(t)$ in that they contain the same statistical information as far as linear operations are concerned and implies that the innovation process for $\tilde{y}_\alpha(t)$ will be the innovation process for $\tilde{y}(t)$. Notice that $y(t)$ and $y_\alpha(t)$ are not equivalent since $y(t)$ is not recoverable from $y_\alpha(t)$ alone due to the presence of non-degenerate initial condition random variables for $y(t)$ process.

The covariance function for $\tilde{y}_\alpha(t)$ process is

$$\begin{aligned}\tilde{R}_\alpha(t,s) &= E[\tilde{y}_\alpha(t)\tilde{y}_\alpha(s)] = R_\alpha(t,s) - \hat{R}_\alpha(t,s) \\ &= \delta(t - s) + K(t,s) - C_\alpha^\dagger(t)R_Y^{-1}C_\alpha(s) \\ &= \delta(t - s) + \tilde{K}(t,s) \end{aligned} \tag{29}$$

Then if -1 is not an eigenvalue of $\tilde{K}$, we may obtain the innovation process for $\tilde{y}_\alpha(t)$, and hence for $\tilde{y}(t)$, by applying the results of Section 2, using the kernel function $\tilde{K}(t,s)$ in the associated Wiener-Hopf equation.

Returning to the decomposition (28), we will interpret an innovation process for the degenerate component $\hat{y}(t)$ (a deterministic process) to be a vector of independent, standard normal random variables V. Then the complete innovation process for $y(t)$ is formed by "augmenting" the white noise innovation process, $\nu(t)$, of the component $\tilde{y}(t)$ by the vector of random variables V, independent of $\nu(t)$ process and considered to occur at $t = 0$.

The canonical representation of process $y(t)$ in terms of the augmented innovation process $\{V,\nu(t)\}$ is carried out in two steps. At first, we relate the process $y(t)$ to the augmented process $\{Y,\tilde{y}_\alpha(t)\}$ by the causal and causally invertible linear operation

$$y(t) = C^\dagger(t)R_Y^{-1}Y + \int_0^t \frac{1}{(\alpha - 1)!}(t - \tau)^{(\alpha - 1)}\tilde{y}_\alpha(\tau)d\tau \tag{30a}$$

This linear operation is causally invertible since the components of Y vector are (theoretically) generated at $t = 0$. Then, the augmented process $\{Y, \tilde{y}_\alpha(t)\}$ is represented in terms of the augmented innovation process $\{V, \nu(t)\}$ by

$$\tilde{y}_\alpha(t) = \nu(t) + \int_0^t \tilde{k}(t,\tau)\nu(\tau)d\tau \tag{30b}$$

$$Y = R_Y^{\frac{1}{2}} V \tag{30c}$$

where $R_Y^{\frac{1}{2}}$ is an arbitrary non-singular square root of the positive definite covariance matrix R_Y and the kernel $\tilde{k}(t,\tau)$ solves

$$(I + \tilde{k})\ (I + \tilde{k}^*) = I + \tilde{K} \tag{31}$$

for $\tilde{K}$ given in (29). The causal representation (30b, c) is causally invertible with

$$\nu_t = (I - \tilde{h})\tilde{y}_\alpha(t) \tag{32a}$$

$$V = R_Y^{-\frac{1}{2}} Y \tag{32b}$$

for $(I - \tilde{h}) = (I + \tilde{k})^{-1}$.

We will now construct the CR of (30a, b, c) for a differentiable process y_t with separable covariance function having the form

$$R(t,s) = A^\dagger(t \vee s)\ B(t \wedge s),\ 0 \le t,\ s \le T \tag{33}$$

where $A(t)$ and $B(t)$ are n-vector functions. Such a covariance function corresponds to a process with additive white noise component in the α^{th} derivative process y_α ($\alpha \le n$), whose covariance function R_α is of the form (25), if and only if

(i) $A(t)$ and $B(t)$ are continuously α-differentiable [with i^{th} derivatives indicated by $A_i(t)$ and $B_i(t)$]

(ii) $A_{i-1}^\dagger(t)\ B_i(t) - B_{i-1}^\dagger(t)\ A_i(t) \equiv 0,\ 0 \le t \le T$ (34a)

$$i = 1, 2, \ldots, \alpha - 2 \text{ if } \alpha > 1$$

$$A^{\dagger}_{\alpha-1}(t)\ B_{\alpha}(t) - B^{\dagger}_{\alpha-1}(t)\ A_{\alpha}(t) \equiv 1,\ 0 \leq t \leq T \tag{34b}$$

for arbitrary α

It is shown by Brandenburg and Meadows [12], that (34a) is necessary and sufficient for mean-square differentiability of order $\alpha - 1$ and that the left hand side of (34b) is nonnegative. So, by repeated differentiation of the conditions (34a), the covariances for the derivative processes are obtained as

$$R_{ij}(t,s) = E[y_i(t)y_j(s)] = \begin{cases} A^{\dagger}_i(t)\ B_j(s),\ s \leq t \\ B^{\dagger}_i(t)\ A_j(s),\ t < s \end{cases} \tag{35}$$

for i, j $\leq \alpha - 1$. Then application of (34b) gives

$$\begin{aligned} R_{\alpha}(t,s) &= \frac{\partial}{\partial t}\frac{\partial}{\partial s} R_{\alpha-1,\ \alpha-1}(t,s) \\ &= \delta(t-s) + A^{\dagger}_{\alpha}(t \vee s) B_{\alpha}(t \wedge s) \end{aligned} \tag{36}$$

But for Y vector defined as in (26), the covariance matrix R_Y is determined by (35) to be

$$R_Y = Q^{\dagger}_A(0)\ Q_B(0) \tag{37}$$

where the $n \times \alpha$ matrix functions Q_A and Q_B are composed of columns as

$$Q_A(t) = [A(t),\ \ldots\ A_{\alpha-1}(t)] \tag{38a}$$

$$Q_B(t) = [B(t),\ \ldots\ B_{\alpha-1}(t)] \tag{38b}$$

It is also shown in [12] that the $\alpha \times \alpha$ matrix

$$Q^{\dagger}_A(t)\ Q_B(t) \text{ is nonsingular for all } t \tag{39}$$

so that R_Y is nonsingular, permitting the decomposition described in (27), (28) to be carried out.

The covariance functions for the component processes $\hat{y}$ and $\tilde{y}$ of the decomposition are easily determined from (35) to be

$$\hat{R}(t,s) = E[\hat{y}(t)\ \hat{y}(s)] = A^{\dagger}(t)P_o\ A(s) \tag{40}$$

for nxn matrix P_o formed as

$$P_o = Q_B(o)\ [Q_A^{\dagger}(o)\ Q_B(o)]^{-1}\ Q_B^{\dagger}(o) \tag{41}$$

and

$$\begin{aligned}\tilde{R}(t,s) &= R(t,s) - \hat{R}(t,s)\\ &= A^{\dagger}(t\vee s)\ [B(t\wedge s) - P_o\ A(t\wedge s)]\end{aligned} \tag{42}$$

Then from (36) and (40),

$$\begin{aligned}\tilde{R}_{\alpha}(t,s) &= E[\tilde{y}_{\alpha}(t)\ \tilde{y}_{\alpha}(s)] = R_{\alpha}(t,s) - \hat{R}_{\alpha}(t,s)\\ &= \delta(t-s) + A_{\alpha}^{\dagger}(t\vee s)\ [B_{\alpha}(t\wedge s) - P_o\ A_{\alpha}(t\wedge s)]\\ &= \delta(t-s) + \tilde{K}(t,s)\end{aligned} \tag{43}$$

If -1 is not an eigenvalue of $\tilde{K}$ in (43), the results of Section 3 may be applied to determine the CR for the derivative process $\tilde{y}_{\alpha}$. The covariance factorization (20a, b, c) is applied to the separable kernel function

$$\tilde{K}(t,s) = A_{\alpha}^{\dagger}(t\vee s)\ \tilde{B}_{\alpha}(t\wedge s)$$

where $\tilde{B}_{\alpha}(t)$ is identified in (43) to be

$$\tilde{B}_{\alpha}(t) = B_{\alpha}(t) - P_o A_{\alpha}(t)$$

and the associated matrix Riccati equation is

$$\dot{\tilde{P}} = (\tilde{P}A_{\alpha} - \tilde{B}_{\alpha})\ (\tilde{P}A_{\alpha} - \tilde{B}_{\alpha})^{\dagger},\ \tilde{P}(0) = 0 \tag{44a}$$

Then the canonical representation of process $\tilde{y}_{\alpha}$ is obtained in terms of the white noise innovation process ν_t, whose covariance is $\delta(t-s)$, as

$$\tilde{y}_{\alpha} = (I+\tilde{k})\nu \tag{45a}$$

$$\tilde{k}(t,\tau) = A_{\alpha}^{\dagger}(t)\ \psi(\tau),\ \tau \leq t \tag{45b}$$

$$\psi(t) = \tilde{B}_{\alpha}(t) - \tilde{P}(t)\ A_{\alpha}(t) \tag{45c}$$

We may change the Riccati variable in (44a) by

$$P(t) = \tilde{P}(t) - P_o$$

giving the more convenient equation

$$\dot{P} = (PA_\alpha - B_\alpha)\ (PA_\alpha - B_\alpha)^\dagger,\ P(0) = P_o \tag{44b}$$

so then

$$\psi(t) = B_\alpha(t) - P(t)\ A_\alpha(t) \tag{45d}$$

This Riccati equation (44b) will have a well defined solution for the initial condition matrix P_o given in (41) whenever the solution of equation (44a) is well defined, which follows when -1 is not an eigenvalue of $\tilde{K}$ in (43) as will be discussed in more detail at the conclusion of this section.

To complete the canonical representation of the observation process, we will evaluate (30a) in terms of the representation (45a) for $\tilde{y}_\alpha$ process. Denote the kernel function which relates $\tilde{y}$ to $\tilde{y}_\alpha$ by

$$S(t,\tau) = \frac{1}{(\alpha-1)!}\ (t - \tau)^{\alpha-1},\ \tau \leq t \tag{46}$$

Then in operator notation

$$\tilde{y} = S\tilde{y}_\alpha = S(I + \tilde{k})\nu \tag{47}$$

for kernel $\tilde{k}(t,\tau)$ given in (45b). It is shown in [5] and [12] that a solution of Riccati equation (44b), whose coefficients $A_\alpha(t)$ and $B_\alpha(t)$ are related to the separable covariance function satisfying (34a, b), must necessarily satisfy

$$P(t)\ Q_A(t) = Q_B(t) \text{ for all } t \tag{48}$$

So then

$$A_i^\dagger(t)\ \psi(t) = A_i^\dagger(t)\ [B_\alpha(t) - P(t)\ A_\alpha(t)]$$

$$= A_i^\dagger(t)\ B_\alpha(t) - B_i^\dagger(t)\ A_\alpha(t)$$

and it also follows from the properties (34a) that the right hand side is identically zero for $i = 0, 1, \ldots, \alpha - 2$. With (34b), this gives

$$A_i^{\dagger}(t)\ \psi(t) \equiv 0, \text{ for all } t,\ i = 0, 1, \ldots, \alpha - 2 \tag{49a}$$

$$A_{\alpha-1}^{\dagger}(t)\ \psi(t) \equiv 1, \text{ for all } t \tag{49b}$$

The relations (49a, b) allow making an explicit evaluation of the kernel function for $S(I + \tilde{k})$ by performing repeated integration by parts to get

$$S(t,\tau) + \int_{\tau}^{t} S(t,\sigma)\ A_{\alpha}^{\dagger}(\sigma)\ \psi(\tau)d\sigma$$
$$= A^{\dagger}(t)\ \psi(\tau),\ \tau \leq t$$

Thus from (47),

$$\tilde{y}(t) = \int_{0}^{t} A^{\dagger}(t)\psi(\tau)\nu(\tau)d\tau \tag{50a}$$

The representation of component process $\hat{y}$ in (30a) is obtained by using the relations (35) to evaluate (27) giving

$$\hat{y}(t) = A^{\dagger}(t)\ Q_B(0)\ R_Y^{-\frac{1}{2}}V$$

where V is the α-vector of standard normal random variables forming the augmentation of the innovation process. But in this situation of separable covariance function, it is convenient to denote an n-vector of random variables by

$$\chi_o = Q_B(0)\ R_Y^{-\frac{1}{2}}V,\ E[\chi_o\chi_o^{\dagger}] = P_o$$

where P_o is given in (41). So then

$$\hat{y}(t) = A^{\dagger}(t)\ \chi_o \tag{50b}$$

and the inverse relation for the χ_o vector in this representation is obtained by applying (32b) to get

$$\chi_o = Q_B(0)\ R_Y^{-1}Y \tag{51}$$

Now equations (50a) and (50b) give the desired canonical representation for a differentiable process with separable covariance function and a realization is obtained by noting that the component $\hat{y}$ in (50b) is generated by adding initial conditions to the obvious realization for (50a). This combined realization for the CR is

$$\dot{\chi}_t = \psi(t)\nu_t, \quad \chi(0) = \chi_o, \quad E[\nu_t \nu_s] = \delta(t-s) \tag{52a}$$

$$y_t = A^\dagger(t)\chi_t, \qquad E[\chi_o \chi_o^\dagger] = P_o \tag{52b}$$

where

$$P_o = Q_B(0)\ [Q_A(0)\ Q_B(0)]^{-1}\ Q_B^\dagger(0), \quad \psi(t) = B_\alpha(t) - P(t)A_\alpha(t) \tag{52c}$$

and $P(t)$ solves the Riccati equation (44b).

By the above discussion of equations (30a, b, c), this construction of the realization (52a, b, c) is indeed causally invertible and we may evaluate the inverse relations (32a, b) for this separable case. First, the inverse of (30a) is

$$\begin{aligned}\tilde{y}_\alpha(t) &= y_\alpha(t) - \hat{y}_\alpha(t) \\ &= y_\alpha(t) - A_\alpha^\dagger(t)\ Q_B(0)\ R_Y^{-1} Y\end{aligned} \tag{53}$$

by using (51). Then a realization for $\tilde{h} = \tilde{k}(I + \tilde{k})^{-1}$, where $\tilde{k}(t,s)$ is given in (45b), is obtained by applying (23) to get

$$\tilde{h}: (-\psi A_\alpha^\dagger,\ \psi,\ A_\alpha)$$

The complete inverse operation which results by applying (32a, b) to (53) (using 51) is then realized by the linear system

$$\dot{w}_t = [-\psi(t)\ A_\alpha^\dagger(t)]w_t + \psi(t)\ y_\alpha(t), \quad w(0) = \chi_o \tag{54a}$$

$$\nu_t = -A_\alpha^\dagger(t)w_t + y_\alpha(t) \tag{54b}$$

$$\chi_o = Q_B(0)\ R_Y^{-1} Y \tag{54c}$$

This system (54a, b, c) is then the whitening filter which yields the augmented innovation process $\{\chi_o, \nu_t\}$ for the observation process y_t having separable covariance function. Notice that the differentiability properties of the process y_t imply constraints on the representation (52a, b) such that the complete initial condition random vector χ_0 (an n-vector) is determined by the α components of $Y = [y(0), y_1(0), \ldots y_{\alpha-1}(0)]^{\dagger}$ as shown by (54c). Since the CR and the whitening filter have the identical initial condition vector, the whitening filter in (54a, b, c) is actually an inverse system to the CR which satisfies $w_t = \chi_t$ for all t (properties of inverses for linear systems are given by L. Silverman [13] and discussed in [5] in relation to the canonical property for representations of stochastic processes). Applications of this explicit realization for the whitening filter will be made in Sections 5 and 6.

The CR in (52a, b, c) is expressed in terms of the solution P of the Riccati equation (44b) and it was pointed out in Section 3 that this matrix differential equation has a well defined solution if and only if -1 is not an eigenvalue of the kernel function $\tilde{K}(t,\tau)$ in (43). For applications in many situations, a sufficient condition for this spectral property of $\tilde{K}$ is that the given differentiable process be a lumped process. However, we will not require a priori that a model be given explicitly, but will show that existence of such guarantees that the canonical representations (45a, b) and (52a, b) can be obtained through solution of the Riccati equation. Thus the applications given in the following sections, which exploit the causal invertibility of the CR, will be well defined by knowledge that the separable covariance function for an observation process is in fact the covariance of a lumped process.

As is shown by Kalman [14], the structure of a lumped process implies the existence of a nonnegative, increasing matrix function W(t), ($W_t \geq 0$,

$\dot{W}_t \geq 0$ for all t) such that the separable covariance function of the lumped process is of the form

$$R(t,s) = A^{\dagger}(t \vee s)\ B(t \wedge s) = A^{\dagger}(t)\ W(t \wedge s)\ A(s) \tag{55a}$$

A realization for a process y_t with covariance (55a) is given by

$$\dot{x}_t = u_t,\ x(0) = x_o \qquad E[u_t u_s^{\dagger}] = \dot{W}(t)\ \delta(t-s) \tag{55b}$$

$$y_t = A^{\dagger}(t) x_t \qquad E[x_o x_o^{\dagger}] = W(0) \tag{55c}$$

But such a realization does not correspond to a CR for process y_t since in general the vector white noise process u_t and the initial condition random vector x_o are not uniquely recoverable from the observation (in particular, the rank of $\dot{W}(t)$ may be greater than unity so that the vector white noise process u_t is not equivalent to the innovation process ν_t).

From (55a), the matrix function W satisfies

$$W(t)\ A(t) = B(t) \text{ for all } t$$

and in consideration of the differentiability of the lumped process, the relations (34a, b) imply that

$$B_{\alpha}(t) = W(t)\ A_{\alpha}(t) + \dot{W}(t)\ A_{\alpha-1}(t) \text{ for all } t \tag{56a}$$

$$A^{\dagger}_{\alpha-1}(t)\ \dot{W}(t)\ A_{\alpha-1}(t) \equiv 1 \tag{56b}$$

(see [5], Lemma 2.6 for detailed calculations). Thus the covariance $\tilde{R}_{\alpha}$ in (43) becomes

$$\begin{aligned}\tilde{R}_{\alpha}(t,s) &= \delta(t-s) + \tilde{K}(t,s) \\ &= \delta(t-s) + A^{\dagger}_{\alpha}(t)\ [W(t \wedge s) - P_o]\ A_{\alpha}(s) \\ &\quad + A^{\dagger}_{\alpha}(t \vee s)\ \dot{W}(t \wedge s)\ A_{\alpha-1}(t \wedge s)\end{aligned} \tag{57}$$

and a realization of process $\tilde{y}_{\alpha}$ is obtained as

$$\dot{\tilde{x}}_t = u_t \quad , \tilde{x}(0) = \tilde{x}_o \quad , E[\tilde{x}_o \tilde{x}_o^\dagger] = W(0) - P_o \tag{58a}$$

$$\tilde{z}_t \triangleq A_\alpha^\dagger(t)\, \tilde{x}_t \quad , \tilde{v}_t \triangleq A_{\alpha-1}^\dagger(t)\, u_t \; , E[u_t u_s^\dagger] = \dot{W}(t)\delta(t-s) \tag{58b}$$

$$\tilde{y}_\alpha(t) = \tilde{z}_t + \tilde{v}_t \, , \qquad E[\tilde{x}_o u_t^\dagger] \equiv 0 \tag{58c}$$

A direct calculation of the covariance of the process modeled in (58a, b, c) will give the covariance function (57) since

$$E[\tilde{v}_t \tilde{v}_s] = A_{\alpha-1}^\dagger(t)\, \dot{W}(t)\, A_{\alpha-1}(s)\; \delta(t-s) = \delta(t-s)$$

by (56b) and

$$E[\tilde{z}_t \tilde{z}_s] = A_\alpha^\dagger(t)\, [W(t\wedge s) - P_o]\, A_\alpha(s)$$

$$E[\tilde{z}_t \tilde{v}_s] = \begin{cases} A_\alpha^\dagger(t)\, \dot{W}(s)\, A_{\alpha-1}(s) & , s \leq t \\ 0 & , t < s \end{cases} \tag{59}$$

But then $\tilde{y}_\alpha$ is the sum of a signal process $\tilde{z}_t$ and white noise $\tilde{v}_t$ with <u>one-sided</u> cross-covariance function, which is the situation discussed in Section 2 for the general covariance function having the form (4). Thus, the desired conclusion that -1 is not an eigenvalue of the continuous kernel $\tilde{K}$ of the covariance function $\tilde{R}_\alpha$ follows, in this situation of lumped observation process, from the <u>existence</u> of $\tilde{z}_t$ and $\tilde{v}_t$ in (58b) satisfying (59). We point out again that the (in general non-canonical) realization (55b, c) which results from the assumption of a lumped observation process need not be explicitly known to perform the canonical representation of (52a, b, c); but the existence of the realization (58a,b, c) implies existence of a well behaved solution for the matrix Riccati equation (44b) which specifies the CR in (52a, b, c).

5. <u>Estimation</u>

In this section we will discuss the linear least-squares estimation of a random process z_t given observation of a related random process $\{y_\tau, \tau \leq t\}$ by the whitening filter technique. For an observation process y_t having a separable covariance function in the form discussed in Section 4, we will show that the solution of this estimation problem, denoted by $\hat{z}(t/t)$, can be implemented in terms of the whitening filter of the observation process. In the whitening filter technique, the only data required in addition to the whitening filter for the observation is the cross-covariance function

$$R_{zy}(t,\tau) = E[z_t y_\tau], \ \tau \leq t$$

which determines an appropriate output gain vector function to be applied to the state variable of the whitening filter. The goal here is to divorce the solution of the filtering problem from any particular coordinate system for models of z_t and y_t processes so that solutions may be obtained in situations where models for the processes are not given a priori.

Consider an observation process y_t with separable covariance function of the form (33),

$$R_y(t,s) = A^\dagger(t \vee s)\ B(t \wedge s), \ 0 \leq t, \ s \leq T \tag{33}$$

such that the differentiability properties discussed in Section 4 apply. Let z_t be an arbitrary random process for which there exists an n-vector function $C(t)$ such that

$$R_{zy}(t,\tau) = E[z_t y_\tau] = C^\dagger(t)\ B(\tau), \ \tau \leq t \tag{60}$$

where vector function $B(t)$ appears in the separation of covariance function R_y. Then the canonical representation of the observation process with realization as in (52a, b, c) simultaneously models the estimate process $\hat{z}(t/t)$ (but not z_t itself) as is shown by an application of the projection theorem in the following.

The projection theorem (see the discussion in [1]) gives the linear least-squares estimate $\hat{z}(t/t)$ as the solution of

$$E[\{z(t) - \hat{z}(t/t)\} y(\tau)] = 0, \ \tau \leq t \tag{61}$$

and we will consider the process $C^{\dagger}(t)\ \chi_t$ where χ_t is the state vector in the canonical representation (52a, b, c) of the observation process y_t. Then since $y_t = A^{\dagger}(t)\ \chi_t$ from (32b),

$$\begin{aligned} E[C^{\dagger}(t)\ \chi_t\ y_\tau] &= C^{\dagger}(t)\ E[\chi(t)\ \chi^{\dagger}(\tau)]\ A(\tau) \\ &= C^{\dagger}(t)\ \{\int_0^t \psi(\sigma)\ \psi^{\dagger}(\sigma)\ d\sigma + P_0\}\ A(\tau) \end{aligned} \tag{62}$$

for ψ given in (52c) and P_0 the covariance matrix of the initial condition χ_0 in (52a). The bracketed term in (62) is immediately seen to be the solution P(t) of the Riccati equation (44b); but from (48)

$$P(t)\ A(t) = B(t)$$

so that

$$\begin{aligned} E[C^{\dagger}(t)\ \chi_t\ y_\tau] &= C^{\dagger}(t)\ P(t \wedge \tau)\ A(\tau) \\ &= C^{\dagger}(t)\ B(\tau), \ \tau \leq t \\ &= E[z(t)\ y(\tau)], \ \tau \leq t \end{aligned} \tag{63}$$

from the assumption (60). This verifies that the solution of (61) is the process $C^{\dagger}(t)\ \chi_t$.

Now we use the canonical property (causal invertibility) of the CR for the observation process to causally reconstruct the state vector χ_t from the observation process itself. This reconstruction is provided by the state vector w_t of the whitening filter (54a, b, c) since the whitening filter is an inverse system to the CR as discussed in Section 4. Thus the solution for the general estimation problem described above is simply obtained by applying vector function C(t) as an output vector on the state equation (54a) of the whitening filter giving the realization of the filter for estimation as

$$\dot{w}_t = [-\psi(t)\ A_\alpha^\dagger(t)]w_t + \psi(t)y_\alpha(t),\ w(0) = Q_B(0)\ R_y^{-1}\ Y \quad (64a)$$

$$\hat{z}(t/t) = C^\dagger(t)\ w_t \quad (64b)$$

The separable form (60) of the cross-covariance function is easily determined in a variety of situations where z_t and y_t are lumped processes. First consider the (noncanonical) model (55b, c) for the lumped observation process with covariance function (55a) and consider $z_i(t)$ to be the i^{th} component of the state vector x_t in (55b) [i.e., $z_i(t) = e_i^\dagger\ x(t)$ where e_i is the unit vector obtained from the i^{th} column of the identity matrix]. Then

$$\begin{aligned} E[z_i(t)\ y(\tau)] &= e_i^\dagger\ E[x_t x_\tau^\dagger]\ A(\tau) \\ &= e_i^\dagger\ W(t\wedge\tau)\ A(\tau) \\ &= e_i^\dagger\ B(\tau),\ \tau \leq t \end{aligned}$$

showing that $C(t)$ in the separation of the cross-covariance is given by the unit vector e_i. Thus the solution of the state estimation problem for a lumped observation process is simply given by the state of the whitening filter, i.e.

$$\hat{x}(t/t) = w_t \text{ (an n-vector estimate)} \quad (65)$$

If a lumped model for the observed process is given a priori, the given coordinates may have the more common feedback form of state equation

$$\dot{x}'(t) = F(t)\ x'(t) + G(t)\ u(t) \quad (66a)$$

$$y_t = H^\dagger(t)\ x'(t) \quad (66b)$$

However, the fundamental solution matrix for the system equation (66a) determines a state transformation

$$x'(t) = T(t)\, x(t)$$

which relates the state $x(t)$ of the model (55b, c) to the state $x'(t)$ of the given model. Then, by the linearity of the estimate,

$$\hat{x}'(t/t) = T(t)\, \hat{x}(t/t) = T(t)\, w_t \tag{67}$$

giving the solution of the state estimate problem in arbitrary coordinates in terms of the realization (64a) for the whitening filter. But, we may also apply transformation T to the state equation (64a) to implement an equivalent realization of the whitening filter which has state variable (equivalent to) $\hat{x}'(t/t)$ directly. The resulting realization for the whitening filter may be determined without explicitly evaluating the transformation T as is shown by example in [5].

A common problem in filtering consists of observing the sum of signal process z_t and a noise process v_t (colored noise) where both z_t and v_t are lumped processes. Suppose $x_z(t)$ and $x_v(t)$ are state vectors of models (noncanonical) for processes z_t and v_t respectively, and

$$z_t = H_z^{\dagger}(t)\, x_z(t), \quad v_t = H_v^{\dagger}(t)\, x_v(t) \tag{68a}$$

Then $y_t = z_t + v_t$ has a lumped model which is formed by augmenting the state vector $x_z(t)$ by the state vector $x_v(t)$ of the noise model. Denoting this augmented state vector by $x'(t)$ gives

$$y_t = [H_z^{\dagger}(t) \,\vdots\, H_v^{\dagger}(t)]\, x'(t) \tag{68b}$$

The solution for the estimate $\hat{z}(t/t)$ of the signal process z_t from observation of y_t is then obtained by forming

$$\hat{z}(t/t) = [H_z^{\dagger}(t) \,\vdots\, 0]\, \hat{x}'(t/t)$$

where $\hat{x}'(t/t)$ is the state estimate for the augmented model of the observation process. As discussed above, the solution for this state estimation problem is obtained by realizing the whitening filter for the observation process in the appropriate coordinate w'(t) [corresponding to x'(t) in (68b)] which gives

$$\hat{z}(t/t) = [H_z^{\dagger}(t) \vdots 0] \, w'(t) \tag{69}$$

Formulas (64b) and (69) demonstrate the nature of solutions for estimation which are obtained by the whitening filter approach. Namely, it is the whitening filter for the observed process which contains the structure needed to perform filtering. Given the whitening filter, the solution for estimation in particular situations then requires only the determination of an appropriate coordinate system for realizing the whitening filter and the determination of the output vector to be applied to the realization of the whitening filter - such determinations involving only algebraic analysis of the data (covariance functions or lumped models).

The whitening filter approach to estimation reveals that in situations which require stochastic modeling as well as filter design, the overall computational effort is minimized by achieving the canonical property in the modeling phase. Although other (noncanonical) solutions for stochastic modeling are possible, filter design for process models not possessing the canonical property is indeed nontrivial. But by directly determining the canonical representation for the observed process as in Section 4, which quite simply provides the whitening filter as a result of the causal invertibility of the CR, a minimum of detail and computation is subsequently required to obtain solutions for filtering.

6. Detection

Discrimination between two Gaussian processes of the type discussed in Section 4 is basic to many detection problems of interest. However, it is more convenient to discuss the simpler problem of discrimination between a differentiable process and an integrated white noise since the chain rule for likelihood ratios may be used in the more general situation. Consider H_1 to be the hypothesis that the observed process y_t is some Gaussian process (of zero mean) such that the α^{th} derivative process has covariance function of the form $I + K_1$ for continuous kernel function $K_1(t,s)$ and such that the initial condition random variables of the derivative processes through order $\alpha - 1$ are linearly independent with nonsingular covariance matrix J [see (25) and (26)]. And, consider the observed process under the testing hypothesis H_2 to be the α-fold integration of Gaussian white noise ν_t $(E[\nu_t \nu_s] = \delta[t-s])$ as given by

$$y_t = \sum_{i=0}^{\alpha-1} \frac{t^i}{i!} \nu_i + \int_0^t \frac{1}{(\alpha - 1)!} (t - \tau)^{(\alpha - 1)} \nu_\tau d\tau \tag{70}$$

where $\nu_0, \dots \nu_{\alpha-1}$ are independent, standard normal variables. As in Section 4, the covariance function of the α^{th} derivative process $y_\alpha(t)$ is denoted by $R_\alpha(t,s)$ and the covariance matrix of the α-vector of initial condition random variables Y is denoted by R_Y.

Then, under each of the hypotheses H_1 and H_2, the observation process y_t is equivalent to the augmented process $\{Y, y_\alpha(t)\}$ since y_t is related to $\{Y, y_\alpha(t)\}$ through an invertible linear operation (the differentiability is the same under each hypothesis). This equivalence implies that the detection problem which tests H_1 versus H_2 can be replaced (through a change of variable in the likelihood ratio), without affecting the probability of error, by the detection problem for augmented observations $\{Y, y_\alpha(t)\}$ having hypotheses

$$\bar{H}_1 : \begin{cases} R_\alpha(t,s) = \delta(t - s) + K_1(t,s) \\ R_Y = J \end{cases}$$

$$\tag{71}$$

$$\bar{H}_2 : \begin{cases} R_\alpha(t,s) = \delta(t - s) \\ R_Y = I \end{cases}$$

Under $\bar{H}_1$, we can obtain a representation for $\{Y, y_\alpha(t)\}$ by applying the methods of Section 4 [see (29) and (30a, b, c)] which are based on performing covariance factorization for

$$\tilde{K}_1(t,s) = K_1(t,s) - C_\alpha^\dagger(t)J^{-1}C_\alpha(s) \tag{72}$$

where

$$C_\alpha(t) = E[Yy_\alpha(t)/H_1] \tag{73}$$

If

$$-1 \text{ is not an eigenvalue of } \tilde{K}_1(t,s) \tag{74}$$

covariance factorization gives kernel $\tilde{k}_1(t,\tau)$ which solves

$$(I + \tilde{k}_1)(I + \check{\tilde{k}}_1^*) = I + \tilde{K}_1 \tag{75}$$

and results in a representation for $\{Y, y_\alpha(t)\}$ in terms of augmented innovation process $\{V_1, \nu_1(t)\}$ $(E[V_1V_1^\dagger] = I,\ E[\nu_1(t)\nu_1(s)] = \delta[t - s])$. This representation is given by

$$\bar{H}_1 : \begin{cases} y_\alpha(t) = C_\alpha^\dagger(t)J^{-\frac{1}{2}}V + (I + \tilde{k}_1)\nu_1(t) & (76a) \\ Y = J^{\frac{1}{2}}V & (76b) \end{cases}$$

The representation (76a) is of the form of signal plus white noise with a signal process

$$z_1(t) = C_\alpha^\dagger(t)J^{-\frac{1}{2}}V + (\tilde{k}_1)\nu_1(t) \tag{76c}$$

which is dependent on both components of the augmented innovation process. Under $\bar{H}_2$, $y_\alpha(t)$ is white noise and is independent of Y so that

$$\bar{H}_2 : \begin{cases} y_\alpha(t) = \nu_2(t) & \text{(77a)} \\ Y = V_2 & \text{(77b)} \end{cases}$$

where also $E[V_2 V_2^\dagger] = I$ and $E[\nu_2(t)\nu_2(s)] = \delta(t - s)$.

For this detection problem which discriminates between $\bar{H}_1$ and $\bar{H}_2$ based on the augmented observation process $\{Y, y_\alpha(t)\}$, the likelihood ratio is formed as a product of likelihood ratios. The first factor is the likelihood ratio for discrimination between $R_Y = J$ and $R_Y = I$ and is the ratio of probability densities for α-vectors of Gaussian random variables given by

$$|J|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}Y^\dagger(J^{-1} - I)Y\} \tag{78}$$

where $|J|$ denotes the matrix determinant of J. Recall that J was assumed nonsingular for the differentiable process observed under H_1. The second factor is the likelihood ratio for the conditional hypothesis test

$$\bar{H}_1^Y : y_\alpha = z_1^Y(t) + \nu_1(t)$$
$$\bar{H}_2^Y : y_\alpha = \nu_2(t) \tag{79}$$

where $z_1^Y(t)$ is the signal process (76c) conditioned on Y and given by

$$z_1^Y(t) = C_\alpha^\dagger(t)J^{-1}Y + (\tilde{k}_1)\nu_1(t) \tag{80}$$

Problem (79) is a test for signal in additive white noise like the problem (15) discussed in Section _ except that here the signal process $z_1^Y(t)$ is of nonzero (conditional) mean. However, the form of likelihood ratio (17) for the detection problem (15) and the conditions for non-singular detection are unaffected by the presence of nonzero, square integrable mean as discussed by Kailath in [3] and [8]. Since kernel $K_1(t,s)$ was assumed continuous, $E[y_\alpha(t)/Y,H_1]$ is square integrable so that the condition for nonsingular detection for the problem (79), and hence for testing H_1 versus H_2, is the same spectral condition (74) which yields the representation for $\{Y, y_\alpha(t)\}$ under H_1. This condition has been given by Shepp [7, Theorem 8]. To apply the likelihood ratio formula (17) to the problem (79), we identify $\hat{z}_1^Y(t)$ to be the causal estimate of the signal $z_1^Y(t)$ under $\bar{H}_1^Y$. But since the representation (76a, b), given Y, is causally invertible, we obtain

$$\hat{z}_1^Y(t) = y_\alpha(t) - \nu_1(t) \tag{81a}$$
$$= (\tilde{h}_1)y_\alpha(t) + (I - \tilde{h}_1)C_\alpha^\dagger(t)J^{-1}Y \tag{81b}$$

for a causal filter $\tilde{h}_1$ determined by

$$\tilde{h}_1 = \tilde{k}_1(I + \tilde{k}_1)^{-1} \tag{82}$$

Thus the likelihood ratio for the conditional hypothesis test (79) is

$$\exp\left\{ \int_0^T \hat{z}_1^Y(t)y_\alpha(t)dt - \frac{1}{2}\int_0^T[\hat{z}_1^Y(t)]^2dt - \frac{1}{2}\int_0^T\tilde{h}_1(t,t)dt \right\} \tag{83}$$

The complete likelihood ratio for the original problem testing H_1 versus H_2 is obtained, through the sequence of arguments above, as the product of (78) and (83) giving (in terms of the derived observations $\{Y, y_\alpha(t)\}$)

$$\text{L. R.} = |J|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}Y^\dagger(J^{-1} - I)Y + \int_0^T\hat{z}_1^Y(t)y_\alpha(t)dt - \frac{1}{2}\int_0^T[\hat{z}_1^Y(t)]^2dt - \frac{1}{2}\int_0^T\tilde{h}_1(t,t)dt \right\} \tag{84}$$

where $\hat{z}_1^Y(t)$ is given by (81b) in terms of $\tilde{h}_1$. Thus the solution for detection problems with differentiable observations is essentially reduced to determining the causal filter $\tilde{h}_1$. But from the results given in Section 2, the causal filter $\tilde{h}_1$ determined in (82) by the covariance factorization (75) will also solve a Wiener-Hopf equation of the form (6) which under H_1 becomes

$$\tilde{h}_1 + \{\tilde{h}_1\tilde{K}_1\}_+ = \{\tilde{K}_1\}_+$$

for $\tilde{K}_1$ given in (72). This gives an alternative to performing covariance factorization in evaluating (84).

In the situation that the differentiable process under H_1 has the separable covariance function

$$R(t,s) = A^{\dagger}(t \vee s)B(t \wedge s)$$

satisfying the differentiability properties (34a, b), the results of Section 4 for such processes are readily applied to determine the causal filter $\tilde{h}_1$. In Section 4 we first performed covariance factorization giving $\tilde{k}_1$ [see (45b, d)] as

$$\tilde{k}_1(t,\tau) = A_{\alpha}^{\dagger}(t)[B_{\alpha}(\tau)-P(\tau)A_{\alpha}(\tau)] = A_{\alpha}^{\dagger}(t)\psi(\tau), \tau \leq t$$

where $P(t)$ is the solution of matrix Riccati equation (44b). Then the whitening filter (54a, b, c) is obtained by applying the relations (22) and (23) to the realization for $\tilde{k}_1$. But the relation (54b) for the whitening filter gives

$$y_{\alpha}(t) - \nu_1(t) = A_{\alpha}^{\dagger}(t)w_t$$

so that the evaluation of (81a) for $\hat{z}_1^{Y}(t)$ is immediately obtained in terms of the state variable w_t in the whitening filter. The covariance matrix J for the separable covariance function is evaluated in (37) as

$$J = Q_A^{\dagger}(0)\; Q_B(0)$$

where matrices Q_A and Q_B are given by (38a, b).

The likelihood ratio for detecting a differentiable process with separable covariance function (with testing process (70)) is then the explicit evaluation of (84) given by

$$L.\ R. = |Q_A^{\dagger}(0)\ Q_B(0)|^{-\frac{1}{2}} \exp\Big\{ -\tfrac{1}{2}Y^{\dagger}([Q_A^{\dagger}(0)\ Q_B(0)]^{-1} - I)Y$$

$$+ \int_0^T A_{\alpha}^{\dagger}(t)w(t)y_{\alpha}(t)dt - \tfrac{1}{2}\int_0^T [A_{\alpha}^{\dagger}(t)w(t)]^2 dt - \tfrac{1}{2}\int_0^T A_{\alpha}^{\dagger}(t)\psi(t)dt \Big\} \tag{85}$$

where $w(t)$ solves

$$\dot{w}_t = [-\psi(t)\ A_{\alpha}^{\dagger}(t)]w_t + \psi(t)y_{\alpha}(t) \tag{86a}$$

$$w(0) = Q_B(0)[Q_A^{\dagger}(0)\ Q_B(0)]^{-1}\ Y \tag{86b}$$

and

$$\psi(t) = B_{\alpha}(t) - P(t)\ A_{\alpha}(t) \tag{86c}$$

for $P(t)$ the solution of matrix Riccati equation (44b) with initial condition matrix (41).

As pointed out in the previous section discussing estimation, it is the whitening filter for the observed process which contains the structure for formulating solutions for problems of statistical inference. This point is further demonstrated here by the likelihood ratio formula (84) based on the quantities J, $\hat{z}_1^Y(t)$, and $\tilde{h}_1(t,t)$, all of which appear in the structure of the whitening filter for the observation process under hypothesis H_1. The explicit likelihood ratio formula (85) for discrimination of a process with separable covariance function shows how the state variable in the lumped whitening filter (86a, b) is incorporated. A change of state variable to realize the whitening filter in other coordinates may be performed as described for the estimation problem in

Section 5, but the likelihood ratio formula is only changed by evaluating $\hat{z}_1^Y(t)$ by

$$\hat{z}_1^Y(t) = [A_\alpha^+(t)\ T^{-1}(t)]\ w'(t)$$

for the state transformation $T(t)$ which determines

$$w'(t) = T(t)\ w(t)$$

7. Conclusion

This paper has developed solutions for estimation and detection with differentiable observations. Such observation processes occur for many problems with colored noise, but the methods discussed here emphasize the analysis of the observed process itself without paying particular attention to signal and noise components of the observation. The techniques used have been based on the whitening filter for the observation process and the causally invertible representation (CR) which readily determines the whitening filter. Quite explicit solutions have been given for estimation and detection problems involving processes with separable covariance functions. These solutions, given in Sections 5 and 6, are based on the solution of a matrix Riccati equation and are in a form to which effective computational methods may be applied.

References

[1] T. Kailath, "An Innovations Approach to Least-Squares Estimation, Part I: Linear Filtering in Additive White Noise", IEEE Trans. on Automatic Control, Vol. AC-13, pp. 646-655, Dec. 1968.

[2] T. Kailath, "Fredholm Resolvents, Wiener-Hopf Equations and Riccati Differential Equations", submitted to IEEE Trans. on Infm. Theory, 1969.

[3] T. Kailath, "Gaussian Signals in Gaussian Noise - A New Look at Some Old Problems", submitted to IEEE Trans. on Infm. Theory, 1969.

[4] T. Kailath and R. Geesey, "Covariance Factorization - An Explication via Examples", Proc. Second Asilomar Conference on System Sciences, Nov. 1968.

[5] R. A. Geesey, "Canonical Representations of Second-Order Processes With Applications", Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, Stanford, Calif., Dec. 1968.

[6] T. Kailath, "Application of a Resolvent Identity to a Linear Smoothing Problem", to appear in J. SIAM on Control, 1969.

[7] L. A. Shepp, "Radon-Nikodym Derivatives of Gaussian Measures", Ann. Math. Stat., Vol. 37, pp. 321-354, Apr. 1966.

[8] T. Kailath, "A General Likelihood-Ratio Formula for Random Signals", to appear in IEEE Trans. on Infm. Theory, 1969.

[9] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory", J. Basic Engr., Trans. ASME, Ser. D, Vol. 83, pp. 95-107, Dec. 1961.

[10] B. D. O. Anderson, "Time Varying Spectral Factorization", Tech. Rept. 6560-8, Stanford Electronics Laboratories, Stanford, Calif., Oct. 1966.

[11] A. E. Bryson, Jr. and D. E. Johansen, "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise", IEEE Trans. Automatic Control, Vol. AC-10, pp. 4-10, Jan. 1965.

[12] L. H. Brandenburg and H. E. Meadows, "Shaping Filter Representation of Nonstationary Colored Noise", Systems Research Group Technical Report No. 106, Department of Electrical Engineering, Columbia University, New York, N.Y., Jul. 1968.

[13] L. M. Silverman, "Properties and Applications of Inverse Systems", IEEE Trans. on Automatic Control, Vol. AC-13, pp. 436-437, Aug. 1968.

[14] R. E. Kalman, "Linear Stochastic Filtering Theory - Reappraisal and Outlook", Proc. Symposium on Systems Theory, pp. 197-205, Polytechnic Institute of Brooklyn, New York, 1965.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Frank J. Seiler Research Laboratory (OAR) USAF Academy, Colorado 80840	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

Applications on the Canonical Representation to Estimation and Detection in Colored Noise

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Scientific

5. AUTHOR(S) *(First name, middle initial, last name)*

Roger A. Geesey, Major, USAF
Thomas Kailath (supported at Stanford University by the Applied Mathematics Division of the AF Office of Scientific Research, OAR, under contract AF 49(838)1517)

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1969	39	4

8a. CONTRACT OR GRANT NO. In-House research b. PROJECT NO. 7904-00-25 c. 61102F d. 681304	9a. ORIGINATOR'S REPORT NUMBER(S) SRL 69-0010 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* AD-

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES Presented at the Symposium on Computer Processing in Communications, Polytechnic Institute of Brooklyn, 9 April 1969.	12. SPONSORING MILITARY ACTIVITY Frank J. Seiler Research Lab (OAR) USAF Academy, Colorado 80840

13. ABSTRACT

The representation of a random process as the output of a causal and causally invertible linear system driven by white noise is called canonical and specifies, quite simply, the whitening filter for the process. Whitening filter techniques replace the observation process, without loss of information, by a white noise process and allow simple formulation of the solutions of estimation and detection problems in terms of the equivalent process obtained by the whitening. Constructive methods based on the solution of a matrix Riccati equation are given for determining the canonical representation of differentiable observation processes which consist of a linear combination of the component process of a finite dimensional Markov process. Implementation of filtering solutions and likelihood ratios for detection are then obtained in a common formulation for a variety of signal with colored noise situations. The approach emphasizes the canonical representation of the observation process while requiring a minimum of attention to models for signal and noise components of the observation. Finite time interval problems for differentiable processes require attention to "initial condition" random variables and the solutions discussed account for their contribution in a natural way.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Canonical Representations						
Estimation						
Filtering						
Likelihood Ratio						
Detection						
Whitening Filters						
Continuous Time Processes						
Nonstationary Processes						
Colored Noise						
Differentiable Processes						