MATHEMATICAL CONTROL THEORY
BANACH CENTER PUBLICATIONS, VOLUME 14
PWN – POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

ESTIMATION FOR LINEAR PURE DELAY TIME SYSTEMS

TYRONE E. DUNCAN

Department of Mathematics, University of Kansas, Lawrence, Kan., U.S.A.

Introduction

In recent years various notions from geometry have been successfully applied to problems in system theory. Hermann–Martin [22] have described a transfer function geometrically as well as important system theoretic concepts associated with a transfer function. Brockett [3] and Byrnes–Duncan [8], [9] have studied the topology of families of transfer functions using geometric methods. Byrnes–Hurt [7] have investigated the orbit space structure of the state space realizations of linear systems. These results only allude to the larger body of recent results in system theory using geometric methods.

The geometric methods that are important here are those that are used to analyze linear pure delay time systems. Such systems have been described as linear systems over a ring of polynomials [27], [43] and this description has been viewed geometrically [5]. State space realization problems have been solved [5], [44]. One method to solve the realization problem is to do a pointwise construction over the prime spectrum of the ring and to show that this construction is locally algebraic. The globalization of this method follows from the positive solution of the Serre problem. This technique of a pointwise construction and its globalization will play a basic role in the methods that are used in this paper. The algebraic condition of reachability has been shown to be a generic property of linear pure delay time systems if the transcendence degree of the polynomial ring is less than the dimension of the input space [32]. Reachability and its dual condition, observability, will be important in the subsequent presentation.

In many filtering problems the data that is given is a transfer function over a ring of polynomials. Such descriptions appear in distributed electrical networks that contain delay lines. Thus a state space realization is required which requires the algebraic conditions of reachability and

observability. From this state space realization a filtering solution is obtained.

The finite time filtering problem for linear pure delay time stochastic systems will be solved pointwise over the prime spectrum of the ring. This pointwise solution will follow from the filtering solution for linear stochastic systems without delays. The solution will be suitably locally algebraic and it will be globalized. Similar geometric methods had been used in [18] to solve a filtering problem for a special class of linear pure delay time stochastic systems.

The infinite time filtering solution for linear pure delay time stochastic systems will be obtained with the algebraic conditions of reachability and observability. This solution will be obtained by solving a dual optimal control problem. A control with finite cost will be constructed by geometric methods.

Briefly to outline this paper some preliminaries will be initially reviewed. A linear pure delay time differential equation can be formulated mathematically in a number of ways. Various functional analytic descriptions as well as a ring theoretic description can be used. Describing the solution in terms of a transfer function over a ring of polynomials will be particularly important here. Some geometric notions are reviewed that enable a geometric description of a linear pure delay time system. The Lagrangian Grassmannian is introduced which plays an important role in optimization problems.

Since the filtering solution of some time delay equations requires some smoothing estimates, the problem of smoothing for linear systems without delays is investigated. Ljung–Kailath [35] have used a result from scattering theory [39] to describe the various smoothing solutions. However, their approach does not clarify the geometric methods that implicitly appear. In this paper the aforementioned scattering result is described geometrically and the various smoothing solutions are shown to arise directly from natural geometric ideas.

The finite time filtering problem for linear pure delay time stochastic systems is formulated and solved by geometric techniques. Using these geometric methods this solution is induced naturally from the filtering solution for linear systems without delays.

The duality between filtering and control for linear pure delay time systems is shown to follow from the duality for linear systems without delays. The infinite time filtering problem is solved by solving the dual infinite time optimal control problem with the algebraic conditions of reachability and observability. Finally some remarks are made on how to apply these methods to linear stochastic optimal control.

1. Preliminaries

A *linear pure delay time system* with input u and output y is a family of equations of the form

$$\frac{dx}{dt} = \sum_{i=1}^{k} \left(A_i x(t-\theta_i) + B_i u(t-\theta_i)\right), \tag{1.1}$$

$$y(t) = \sum_{i=1}^{k} C_i x(t-\theta_i) \tag{1.2}$$

where $A_i \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^n)$, $B_i \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^m, \boldsymbol{R}^n)$, $C_i \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^p)$, $\theta_i \geqslant 0$, $i = 1, 2, \ldots, k$. The differential equation (1.1) will be called *time invariant* if $(A_i, B_i)_{i=1,\ldots,k}$ are constant homomorphisms. Similarly the system (1.1)–(1.2) will be called *time invariant* if $(A_i, B_i, C_i)_{i=1,\ldots,k}$ are constant. The space of initial conditions is the Cartesian product of the Euclidean space $\boldsymbol{R}^n$ and the function space $L^p([-\tau, 0], \boldsymbol{R}^n)$ where $\tau = \max \theta_i$. For our purposes p will be 1 or 2. This space of initial conditions was introduced by Delfour–Mitter [13], [14] who demonstrated the existence and the uniqueness of a locally integrable solution of (1.1) when $(A_i, B_i)_{i=1,\ldots,k}$ are locally integrable. Alternatively, a solution in the space of continuous functions can be established given an initial condition in this space [21]. However, the former initial conditions are important in some applications and will be used in this paper. Let δ_a be the Dirac distribution at a. Since $(\delta_a * x)(t) = x(t-a)$, the equation (1.1) can be described in the convolution ring of Schwartz distributions. This approach can be viewed functional analytically using distribution theory or algebraically using the ring structure. The equation (1.1) is viewed as a linear functional differential equation over the subring of the convolution ring of Schwartz distributions formed from the finite sums of Dirac distributions and locally integrable functions.

The equation (1.1) can also be described as an equation over a ring of polynomials. Given the delays $\{\theta_1, \ldots, \theta_k\}$ it is possible to obtain a collection $\{\tilde{\theta}_1, \ldots, \tilde{\theta}_l\}$ such that

$$\sum_{i=1}^{l} n_i \tilde{\theta}_i = 0$$

implies that $n_i = 0$ for all $i = 1, \ldots, l$, $\tilde{\theta}_i > 0$, $i = 1, \ldots, l$ and

$$\theta_j = \sum_{i=1}^{l} n_i(j)\, \tilde{\theta}\,, \qquad n_i(j) \geqslant 0, \qquad j = 1, \ldots, k,\ i = 1, \ldots, l.$$

Define the linear operator σ_i as

$$(\sigma_i x)(t) = x(t-\tilde{\theta}_i), \qquad i = 1, \ldots, l.$$

Then the equation (1.1) can be described as an equation over the ring $\boldsymbol{R}[\sigma_1, \ldots, \sigma_l]$ and, of course, the transcendence degree of this ring is l. Proceeding more algebraically, a ring of polynomials could have been defined in indeterminates that correspond to the delays $(\theta_1, \ldots, \theta_k)$ and then the transcendence degree of this ring could have been computed to specify independent indeterminates.

The solution of the equation (1.1) can be obtained by many methods. The formulations in the space $\boldsymbol{R}^n \times L^p([-\tau, 0], \boldsymbol{R}^n)$ or the space of continuous functions can be used and the solution can be constructed using the methods of functional analysis. However, for our purposes the formulation of the equation over the ring of polynomials and its relation to the formulation over the subring of the convolution ring of Schwartz distributions will be particularly important. For a time invariant system (1.1)–(1.2) this approach will provide a transfer function method for solution of the equations. This method was developed by Kamen [26], [27]. Kamen (Prop. 3, [27]) showed that the ring homomorphism

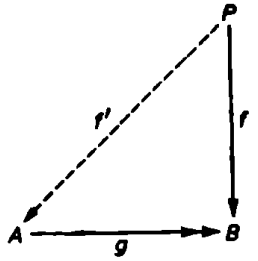
$$\varrho\colon S[w] \to S[p]$$

given by $\sum a_i w^i \mapsto \sum a_i * p^i$ where w is an indeterminate, p is the distributional derivative, and S is the subring of Schwartz distributions is an isomorphism for characteristic polynomials. Thus the transfer function over the ring of polynomials converts the problem of finding the solution of (1.1)–(1.2) from analysis to algebra. This transfer function is solved over the fraction field and the solution is shown to be locally integrable. The solution is given as

$$x = (pI - A)^{-1} * (B * u_+ + x_0 \delta_0 + z)$$

where x_0 and z are the initial conditions and u_+ is the input for t positive. The equivalence between the algebraic description of the solution of (1.1)–(1.2) in terms of a transfer function over a ring of polynomials and the functional analytic description of the solution of (1.1)–(1.2) in the convolution ring of Schwartz distributions will be especially important in the subsequent work.

To describe geometrically the system (1.1)–(1.2) viewed as a system over a ring of polynomials it is useful to review some terminology. Let R be a ring. An R-module P is called *R-projective* if for any R-epimorphism $g \in \mathrm{Hom}_R(A, B)$ and any $f \in \mathrm{Hom}_R(P, B)$, there exists an $f' \in \mathrm{Hom}_R(P, A)$ such that $g \circ f' = f$. This property is described by the diagram



For the case where $P = B$ we have the geometric notion of projection. This definition is equivalent to the property that any R-epimorphism $A \to P$ splits or that P is a direct summand of a free R-module. Let p be a prime ideal of R. The localized module P_p of the R-module P is the module induced from P by the localization R_p of the ring R. An R-module P is projective if and only if P_p is free for each prime ideal p. This description of a projective module alludes strongly to its geometric description. Let Spec(R) be the prime spectrum of R, that is the set of all prime ideals of R with the Zariski topology. There is a bijection between algebraic vector bundles over Spec(R) and R-projective modules [41]. This bijection preserves such operations as direct sum, tensor product and exterior powers. The rank of the bundle can be computed locally on the residue field and it is easy to see that it is constant on connected components.

For the ring $R = \boldsymbol{R}[\sigma_1, \ldots, \sigma_l]$ it is well known that Spec(R) is affine space $\boldsymbol{A}^l$. The linear system (1.1)–(1.2) is said to be *reachable* if

$$\text{rk}[B, AB, \ldots, A^{n-1}B](\sigma) = n \quad \text{for all } \sigma \in \boldsymbol{A}^l \tag{1.3}$$

and it is said to be *observable* if

$$\text{rk}[C, CA, \ldots, CA^{n-1}](\sigma) = n \quad \text{for all } \sigma \in \boldsymbol{A}^l. \tag{1.4}$$

If the linear system over the ring corresponding to (1.1)–(1.2) is reachable and observable then the canonical state module is a finitely generated, projective R-module of rank n. Given a transfer function over R, if the rank of the associated Hankel matrix is a constant n then a realization (A, B, C) can be obtained where the canonical state module is a finitely generated, projective module of rank n [5]. Byrnes [5] constructed locally a minimal realization using the rank condition of the Hankel matrix and then globalized this construction by the positive solution of the Serre problem [38], [45]. The realization can also be done using the notion of split system [44]. On the canonical state module there is a GL$(n; R)$ action. Kappel [28] has shown that the natural analytic identification exists between the solution of (1.1)–(1.2) and the solution of the equations obtained by a coordinate transformation using an element of GL$(n; R)$. Using the geometric approach many problems of linear pure delay time systems can be solved pointwise over Spec(R). These solutions are shown to be described locally algebraically so that a projective module is obtained. Finally, it is necessary to globalize the solution, which requires showing that the algebraic vector bundle is globally trivial. In the realization problem described above the positive solution of the Serre problem is required. Often local constructions exhibit certain symmetries which can be described as reductions of the structure group GL$(n; R)$. For example, the internally symmetric realization of a symmetric transfer function over R requires that a symmetric form is preserved [8]. The associated algebraic problem is usually called the quadratic Serre problem.

In optimization or filtering problems a symplectic algebraic vector bundle arises which implies that a skew form is preserved or equivalently that $\mathrm{GL}(n; R)$ is reduced to $\mathrm{Sp}(n; R)$. It is known that symplectic projective modules are always free [2].

The Lagrangian Grassmannian is an important geometrical object that arises in optimization problems. Let $\mathrm{Gr}(n, 2n)$ be the real Grassmannian of n-planes in $2n$ space. The Lagrangian Grassmannian, $\mathrm{LG}(n)$, is a subset of $\mathrm{Gr}(n, 2n)$ such that if $\Pi \in \mathrm{LG}(n)$ then $J\Pi \perp \Pi$ where J is the standard complex structure on $\boldsymbol{R}^{2n}$. The Lagrangian Grassmannian $\mathrm{LG}(n)$ can also be described as the homogeneous space $U(n)/O(n)$ [1]. It can also be identified with a compactification of symmetric matrices because a plane Π is Lagrangian if and only if there is a generating function $\frac{1}{2}\langle Sq, q\rangle$ such that $p = Sq$ where (q, p) are suitable coordinates in $\boldsymbol{R}^{2n}$ [10]. In optimization problems Hamiltonian equations arise as necessary conditions for an extremum. For the optimization problems with linear equations and quadratic costs the linear Hamiltonian equations describe a flow in the symplectic group. Computing the Hamilton–Jacobi equation as a local solution of the optimization problem gives a Riccati equation. Riccati equations appear as vector fields in the Lagrangian Grassmannian because the technique of going from the Hamiltonian equations to the Hamilton–Jacobi equation is a projection from $2n$-space to Lagrangian n-planes [17]. The Hamiltonian equations lose their sufficiency for an extremum when the curve in the Lagrangian Grassmannian intersects the Maslov cycle and a caustic appears.

For the optimization or estimation problem for linear pure delay time systems an application of the results for systems without delays at each point of $\mathrm{Spec}(R)$ will piece together algebraically to define a symplectic algebraic vector bundle. Thus it should not be surprising that a Riccati equation plays a basic role in such optimization problems.

To gain some perspective of systems over rings a few examples will be briefly described. Given a transfer function over a ring of polynomials a nonprojective module will arise if the conditions of reachability and observability (1.3)–(1.4) are not satisfied or more directly if the local rank of the Hankel matrix is not constant.

It is easy to construct projective modules that are not free. Consider the coordinate ring of the real n-sphere $R_n = \boldsymbol{R}[t_0, \ldots, t_n]/(t_0^2 + \ldots + t_n^2 - 1)$. Let $x_0, \ldots, x_n$ be the images of $t_0, \ldots, t_n$ in R_n and let $\tau^{(n)}$ be the (unimodular) row $(x_0, \ldots, x_n)$. The solution space $P^{(n)}$ of $\tau^{(n)}$ defines a finitely generated (stably free) R_n-module of rank n. Specifically a local computation shows that there are n linearly independent vectors that can be adjoined to $\tau^{(n)}$ to make an $(n+1) \times (n+1)$ invertible matrix. This module is the algebraic tangent bundle of the real n-sphere. This module is free if and only if the algebraic tangent bundle is globally trivial. This

occurs only if $n = 1, 3, 7$. For example, for $n = 2$ suppose that (x_0, x_1, x_2) can be completed as

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \\ z_0 & z_1 & z_2 \end{bmatrix}$$

with determinant a unit in R_2. Since

$$e = z_0(x_1y_2 - x_2y_1) - z_1(x_0y_2 - x_2y_0) + z_2(x_0y_1 - x_1y_0)$$

is assumed to be nonzero on S^2, the vector $(y_0(v), y_1(v), y_2(v))$ is nowhere colinear with $v \in S^2$. The orthogonal projection of y on TS^2 is a nowhere zero continuous vector field which is well known to be impossible.

It is elementary to construct examples of linear pure delay time systems that are reachable and observable as defined by (1.3)–(1.4). A simple example is

$$\frac{dx_1}{dt} = x_2(t) + u(t),$$

$$\frac{dx_2}{dt} = x_1(t) + x_2(t-1) + u(t-1),$$

$$y(t) = -x_1(t-1) + x_2(t).$$

On the other hand it is easy to construct linear pure delay time differential equations that satisfy an R^n reachability property [47] but are not reachable in the ring theoretic sense. However, often the data that is available for delay time system problems is not a state space description but a transfer function description. In many physical problems the transfer function is given over the ring of polynomials in the delays. For example, this can occur in distributed parameter electrical networks that contain delay lines.

2. Smoothing for linear stochastic systems without delays

Recursive algorithms for estimation were initially obtained for the solutions of the problems of filtering and prediction [25]. The smoothing problem presented somewhat more difficulty because if a linear operation is performed on the observations for the smoothing estimate, there is a correlation in the observations which complicates the recursive procedure. However, a number of different algorithms have been given for the solution of the smoothing problem, e.g. [20], [31], [36]. One approach [35] has been given that relates these different algorithms by the use of scattering theory techniques. However, these scattering methods are

unfamiliar to most people who work in estimation theory and these methods, as they were used, do not provide a geometric or even intuitive understanding of the various smoothing formulae. The star product of Redheffer [39] that was used in [35] to obtain various smoothing algorithms is not given any geometrical description so that it can seem as merely a strange rule for multiplication. In addition, to obtain some smoothing formulae some ad hoc techniques were used because certain linear transformations are not invertible. More recently the notion of minimal splitting subspaces [34] has been applied to the smoothing problem.

Some of the elementary ideas of scattering theory will be briefly described as a prelude to the geometric view of it in the Lagrangian Grassmannian. Scattering theory can loosely be described as the mathematical study or description of the effect of an obstacle (in three space) on the propagation of waves. Schematically scattering can be described as an obstacle from which waves are scattered:

where v_1 and v_4 are the incident waves and v_2 and v_3 are the reflected waves. These waves are related by the following transformation

$$\begin{bmatrix} v_3 \\ v_2 \end{bmatrix} = \begin{bmatrix} t & \varrho \\ r & \tau \end{bmatrix} \begin{bmatrix} v_1 \\ v_4 \end{bmatrix}. \tag{2.1}$$

The quantities t and r are the transmission and the reflection coefficients, respectively, on the left side of the above diagram and ϱ and τ are the reflection and the transmission coefficients, respectively, on the right side of the diagram. The equation (2.1) can also be expressed as

$$-\begin{bmatrix} -1 & \varrho \\ 0 & \tau \end{bmatrix} \begin{bmatrix} v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} t & 0 \\ r & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \tag{2.2}$$

This equation will be important for the geometric interpretation of scattering. The star product of Redheffer [39] is a rule for the composition of two obstacles, that is the series connection of two boxes such as the one given in the above diagram. This product is given by the following equation:

$$\begin{bmatrix} t_1 & \varrho_1 \\ r_1 & \tau_1 \end{bmatrix} * \begin{bmatrix} t_2 & \varrho_2 \\ r_2 & \tau_2 \end{bmatrix} = \begin{bmatrix} t_2(1-\varrho_1 r_2)^{-1} t_1 & \varrho_2 + t_2 \varrho_1 (1 - r_2 \varrho_1)^{-1} \tau_2 \\ r_1 + \tau_1 r_2 (1 - \varrho_1 r_2)^{-1} t_1 & \tau_1 (1 - r_2 \varrho_1)^{-1} \tau_2 \end{bmatrix}. \tag{2.3}$$

For $j = 1, 2$ let

$$A_j = \begin{bmatrix} t_j & 0 \\ r_j & -1 \end{bmatrix}, \tag{2.4}$$

$$B_j = -\begin{bmatrix} -1 & \varrho_j \\ 0 & \tau_j \end{bmatrix}. \tag{2.5}$$

For a composition of two obstacles each described by equations in the form of (2.2) the waves on the left side of the two obstacles can be related to the waves on the right side by

$$A_1^{-1}B_1A_2^{-1}B_2.$$

This product must be expressed as $A_3^{-1}B_3$ where A_3 and B_3 have the form of (2.4) and (2.5), respectively. Since the upper or the lower triangular invertible matrices form a subgroup of the general linear group, it is only necessary to reexpress the product $B_1A_2^{-1}$ to obtain the expression $A_3^{-1}B_3$. The following elementary lemma will effect the desired change.

LEMMA 2.1. *Let* $a, b, c, d \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^n)$ *be such that* $c, (1-ad) \in \mathrm{GL}(n; \boldsymbol{R})$. *The following equation is satisfied where the factorization on the right-hand side is unique given the left-hand side.*

$$\begin{bmatrix} 1 & -a \\ 0 & -b \end{bmatrix}\begin{bmatrix} c & 0 \\ d & -1 \end{bmatrix}^{-1} = \begin{bmatrix} c(1-ad)^{-1} & 0 \\ -bd(1-ad)^{-1} & -1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & c(1-ad)^{-1}a \\ 0 & -[b+bd(1-ad)^{-1}a] \end{bmatrix}. \tag{2.6}$$

Applying this lemma to the equation

$$A_1^{-1}B_1A_2^{-1}B_2 = A_3^{-1}B_3$$

gives

$$A_3 = \begin{bmatrix} t_2(1-\varrho_1r_2)^{-1}t_1 & 0 \\ r_1+\tau_1r_2(1-\varrho_1r_2)^{-1}t_1 & -1 \end{bmatrix}, \tag{2.7}$$

$$B_3 = -\begin{bmatrix} -1 & \varrho_2+t_2\varrho_1(1-r_2\varrho_1)^{-1}\tau_2 \\ 0 & \tau_1(1-r_2\varrho_1)^{-1}\tau_2 \end{bmatrix}. \tag{2.8}$$

Clearly the nontrivial terms that appear in A_3 and B_3 are the ones that appear in the star product (2.3).

For a Lagrangian n-plane, a point in the Lagrangian Grassmannian $\mathrm{LG}(n)$, coordinates can be chosen so that $\xi = Px$ where $P \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^n)$ is symmetric. Expressing the outgoing waves $y = \begin{bmatrix} v_3 \\ v_2 \end{bmatrix}$ as a linear transformation of the incoming waves $z = \begin{bmatrix} v_4 \\ v_1 \end{bmatrix}$ we obtain

$$y = Pz \tag{2.9}$$

where

$$P = \begin{bmatrix} \varrho & t \\ \tau & r \end{bmatrix}. \tag{2.10}$$

Clearly this grouping of variables is natural physically and system theoretically where the incoming waves are viewed as inputs and the outgoing waves as outputs.

For the application of scattering theory to the smoothing problem ϱ and r will be symmetric and the transpose of τ is t.

To describe the smoothing problem consider the stochastic processes described by the stochastic differential equations

$$dX(t) = FX(t)dt + GdB(t), \tag{2.11}$$

$$dY(t) = HX(t)dt + d\tilde{B}(t) \tag{2.12}$$

where $F \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^n)$, $G \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^m, \boldsymbol{R}^n)$, $H \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^p)$, $(\tilde{B}(t))$ and $(B(t))$ are independent standard m- and p-dimensional Brownian motions, respectively, $Y(0) \equiv 0$ and $X(0) = X_0$ is a zero mean Gaussian random variable with covariance P_0 that is independent of the two Brownian motions.

Let $\tau < t$ and suppose that the conditional mean, $E[X_\tau | Y_s, 0 \leqslant s \leqslant t]$, and the corresponding error covariance are desired in terms of recursive algorithms. Treating X_τ as an unknown parameter this problem can be solved from the filtering solution with parameters. Apparently this approach was first suggested for such questions by Zachrisson [50]. The infinitesimal data for this augmented problem are

$$\mathscr{F} = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathscr{G} = \begin{bmatrix} G \\ 0 \end{bmatrix}, \quad \mathscr{H} = [H \quad 0]$$

with the stochastic equations

$$d\mathscr{X}(s) = \mathscr{F}\mathscr{X}(s)ds + \mathscr{G}dB(s), \tag{2.13}$$

$$d\mathscr{Y}(s) = \mathscr{H}\mathscr{X}(s)ds + d\tilde{B}(s) \tag{2.14}$$

where τ is fixed, $s \in [\tau, t]$ and

$$\mathscr{X}(s) = \begin{bmatrix} X(s) \\ X(\tau) \end{bmatrix}. \tag{2.15}$$

The equations for the augmented filtering solution are

$$d\hat{\mathscr{X}}(s) = [\mathscr{F} - \mathscr{P}\mathscr{H}^T\mathscr{H}]\hat{\mathscr{X}}(s)ds + \mathscr{P}\mathscr{H}^T dY(s), \tag{2.16}$$

$$\frac{d\mathscr{P}}{ds} = \mathscr{F}\mathscr{P} + \mathscr{P}\mathscr{F}^T + \mathscr{G}\mathscr{G}^T - \mathscr{P}\mathscr{H}^T\mathscr{H}\mathscr{P} \tag{2.17}$$

where

$$\hat{\mathscr{X}}(\tau) = \begin{bmatrix} \hat{X}(\tau) \\ \hat{X}(\tau) \end{bmatrix},$$

$$\mathscr{P}(\tau) = \begin{bmatrix} P(\tau) & P(\tau) \\ P(\tau) & P(\tau) \end{bmatrix},$$

$$\hat{X}(\tau) = E[X(\tau)|Y(\sigma),\ 0 \leqslant \sigma \leqslant \tau],$$

$$P(\tau) = E[\tilde{X}(\tau)\tilde{X}^T(\tau)] \quad \text{and} \quad \tilde{X}(\tau) = X(\tau) - \hat{X}(\tau).$$

Once the error covariance has been identified for the filtering problem, it is elementary to obtain the optimal estimate. Similarly, this will be the case for the smoothing problem.

Probably the most direct approach to the solution of the smoothing problem, which is sometimes called the forward innovations approach, is to solve the augmented Riccati equation (2.17) in forward time with the correct initial conditions. This forward innovations approach gives the equations

$$\frac{d}{ds}P_{12} = [F - P_{11}H^TH]P_{12}, \tag{2.18}$$

$$P_{12}(\tau) = P(\tau),$$

$$\frac{d}{ds}P_{11} = FP_{11} + P_{11}F^T + GG^T - P_{11}H^THP_{11}, \tag{2.19}$$

$$P_{11}(\tau) = P(\tau),$$

$$\frac{d}{ds}P_{22} = -P_{21}H^THP_{12}, \tag{2.20}$$

$$P_{22}(\tau) = P(\tau)$$

where

$$\mathscr{P} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

is partitioned into $n \times n$ blocks and $P_{22}(t)$ is the error covariance for the conditional mean of $X(\tau)$ given the observations until time t, $\hat{X}(\tau|t)$. From P_{22} it follows directly that

$$\hat{X}(\tau|t) = \hat{X}(\tau) + P(\tau)\lambda(\tau, t) \tag{2.21}$$

where

$$\lambda(\tau, t) = \int_{\tau}^{t} \varphi^T(s, \tau)H^T[dY(s) - H\hat{X}(s)ds] \tag{2.22}$$

and φ is the fundamental solution of (2.18) or equivalently the fundamental solution of the optimal estimate equation.

Another solution to the smoothing problem has been given by Lainiotis [31] via the so-called partitioned formulae. In this case a special inital condition is used in the solution of the augmented Riccati equation in forward time, that is,

$$\mathscr{P}(\tau) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.23}$$

where 1 is the identity transformation. This is a natural initial condition because the off diagonal terms in $\mathscr{P}$, P_{12} and P_{21}, are the optimal system and its transpose. In addition, it is a natural initial condition for the scattering theory description. Recall the linear transformations (2.2) that describe the scattering theory picture. Evaluating this equation for $\varrho = 0$, $r = 0$, $t = 1$ and $\tau = 1$ gives

$$\begin{bmatrix} v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Then changing the initial condition to (2.23) merely requires the composition of two linear relations which has been computed in (2.7)–(2.8). Since the primary interest is in the smoothed error covariance $P_s(\tau|t)$, only it will be given:

$$\begin{aligned} P_s(\tau|t) &= P(\tau) - P(\tau)P_{22}^0(t, \tau)\,[1 + P(\tau)P_{22}^0(t, \tau)]^{-1}P(\tau) \\ &= [P_{22}^0(t, \tau) + P^{-1}(\tau)]^{-1} \end{aligned} \tag{2.24}$$

where

$$P_{22}^0(t, \tau) = \int_\tau^t \varphi_0^T(s, \tau)H^T H\varphi_0(s, \tau)ds$$

and P_{22}^0 is obtained from the equation (2.17) with the initial condition

$$\mathscr{P}(\tau) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and φ_0 is the fundamental solution for the optimal filter. The optimal estimate is

$$\hat{X}(\tau|t) = P_s(\tau|t)\,[\lambda_0(\tau, t) + P^{-1}(\tau)\hat{X}(\tau)] \tag{2.25}$$

where

$$\lambda_0(\tau, t) = \int_\tau^t \varphi_0^T(s, \tau)H^T[dY(s) - H\hat{X}_0(s)ds] \tag{2.26}$$

and $\hat{X}_0$ is the solution of the optimal filter where the symmetric initial condition

$$\mathscr{P}(\tau) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is used.

While the optimal estimate $\hat{X}(\tau|t)$ in (2.25) can be obtained from the error covariance P_s, it seems worthwhile to describe the simple geometric ideas that give this estimate. The term λ_0 in (2.26) as well as the term λ in (2.22) are elements in the ξ-plane where the canonical variables in the Hamiltonian equations or equivalently the Lagrangian Grassmannian are (x, ξ). From the forward innovations solution it follows that $\hat{X}(\tau)$ and λ_0 must be combined. Since $\hat{X}(\tau)$ is in the x-plane, it is mapped to the ξ-plane by $P^{-1}(\tau)$. The two terms are added, $\lambda_0 + P^{-1}(\tau)\hat{X}(\tau)$ and then this sum must be mapped back to the x-plane. The symmetric form that relates x and ξ for the smoothing solution is the error covariance for the smoothing problem, $P_s(\tau|t)$, which gives (2.25).

A natural dual approach to that used to obtain the partitioned formulae is to perform a solution in the ξ-plane in reverse time with the given data at the final time t. This approach to the smoothing problem was given independently by Mayne [36] and Fraser [20] and it is often called the two filter approach. In this approach the interplay between the x and the ξ variables is more apparent. Recall that a Lagrangian Grassmannian is a natural compactification of symmetric matrices. This approach is sometimes formally described as using a forward and a backward optimal filter where the covariance for the backward solution has the initial value infinity. Clearly to make this description precise the backward solution should use the ξ-variable instead of the x-variable.

To obtain the partitioned formulae without any computations a comparison will be made between the Hamiltonian equations for an estimation problem and an optimal control problem. Recall the Hamiltonian equations for the estimation problem described by (2.11)–(2.12):

$$\frac{dx}{dt} = -F^T x + H^T H \xi, \tag{2.27}$$

$$\frac{d\xi}{dt} = GG^T x + F\xi. \tag{2.28}$$

For the augmented Riccati equation the given data is

$$\mathscr{P}^0(t) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

With this given data for the scattering problem, the infinitesimal generator for the backward solution is the same as for the forward solution [39].

This fact is also clear for this data from the geometric description of the problem because the notion of the positivity of time plays no intrinsic role in the description. Thus the vector fields in the Hamiltonian equations (2.27)–(2.28) are the same for this backward solution where the "state" is considered as the ξ variable because these equations are locally equivalent to the Riccati equation. Solving the Hamiltonian equations in reverse time is merely solving an optimal control problem where the ξ-variable is the state. Recall the Hamiltonian equations for the standard optimal control problem represented in state space by (A, B, C):

$$\frac{dx}{dt} = Ax - BB^T\xi, \tag{2.29}$$

$$\frac{d\xi}{dt} = -C^TCx - A^T\xi. \tag{2.30}$$

Now it is only necessary to make the natural identifications between these two families of Hamiltonian equations to obtain the augmented Riccati equation $\mathscr{P}^0$ and the optimal smoothing estimate:

$$\frac{-d}{d\sigma}P_{11}^0(\sigma) = -H^TH + F^TP_{11}^0 + P_{11}^0F + P_{11}^0GG^TP_{11}^0, \tag{2.31}$$

$$P_{11}^0(t) = 0,$$

$$\frac{-d}{d\sigma}P_{22}^0 = P_{21}^0GG^TP_{12}^0, \tag{2.32}$$

$$P_{22}^0(t) = 0,$$

$$\frac{-d}{d\sigma}P_{12}^0(\sigma) = [F^T + P_{11}^0GG^T]P_{12}^0, \tag{2.33}$$

$$P_{12}(t) = 1,$$

$$P_s(\tau|t) = [-P_{11}^0(t) + P^{-1}(\tau)]^{-1}, \tag{2.34}$$

$$\hat{X}(\tau|t) = P_s(\tau|t)\,[Z(\tau) + P^{-1}(\tau)\hat{X}(\tau)], \tag{2.35}$$

$$dZ(\sigma) = -[F^T + P_{11}^0GG^T]Z(\sigma)\,d\sigma - H^T\,dY(\sigma), \tag{2.36}$$

$$Z(t) = 0.$$

Since the two filter approach can be viewed as a dual to the partitioned formulae, it is natural to find a dual to the forward innovations approach. This approach is often called the backward innovations approach. This approach can be seen most clearly in two steps. Initially, a realization must be obtained for the state stochastic process in reverse time. Since this stochastic process is a zero mean Gaussian process, it is uniquely deter-

mined in probability law by its covariance function. To transport the covariance for $(X(s))$ from τ to t it suffices to solve a filtering problem without observations. This problem can be solved by the Hamiltonian equations

$$\frac{dx}{dt} = -F^T x, \tag{2.37}$$

$$\frac{d\xi}{dt} = GG^T x + F\xi \tag{2.38}$$

or equivalently by the Riccati equation

$$\frac{d\Pi}{ds} = F\Pi + \Pi F^T + GG^T, \tag{2.39}$$

$$\Pi(\tau) = P(\tau).$$

It is claimed that the state transition matrix, F_B, for the backward realization of this Gauss–Markov process is

$$F_B = -(F + GG^T \Pi^{-1}). \tag{2.40}$$

Two verifications of this result will be described. First, the state transition matrix can be seen from the Hamiltonian equations. The dual variable ξ is used because of the natural pairing in optimization problems between the state with an initial condition and the dual variable with a final condition. The negative sign arises because the system is run in reverse time and the equation for ξ in the Hamiltonian equations is in the forward direction. The equivalence of the two realizations can also be verified directly from the Riccati equations. For the system

$$-dX_B(\sigma) = F_B X_B d\sigma + G dB_B(\sigma) \tag{2.41}$$

there is the Riccati equation

$$\frac{-d\tilde{P}}{d\sigma} = GG^T - [F + GG^T \Pi^{-1}]\tilde{P} - \tilde{P}[F + GG^T \Pi^{-1}]^T, \tag{2.42}$$

$$\tilde{P}(t) = \Pi(t).$$

Since $\Pi(t) = \tilde{P}(t)$ and

$$\left.\frac{d\Pi}{ds}\right|_{s=t} = \left.\frac{d\tilde{P}}{d\sigma}\right|_{\sigma=t}, \tag{2.43}$$

it follows by the uniqueness of the solutions of the Riccati equations that $\Pi(s) = \tilde{P}(s)$ for $s \in [\tau, t]$. The vector field $\frac{d\tilde{P}}{d\sigma}$ instead of $\frac{-d\tilde{P}}{d\sigma}$ is used

in (2.43) because the negative sign was introduced to compute the solution in reverse time.

The Hamiltonian equations for the backward estimation problem are the same as those for the forward estimation problem except that the direction of time has been reversed. The Hamiltonian equations can be expressed as

$$\frac{dx}{da} = -F_B^T x + H^T H \xi, \tag{2.44}$$

$$\frac{d\xi}{da} = GG^T x + F_B \xi. \tag{2.45}$$

Recalling that these Hamiltonian equations solve an estimation problem with the notion of positive time reversed, the augmented Riccati equation in the initial notion of positive time is

$$\frac{-d}{d\sigma} P_{12} = -[F + GG^T \Pi^{-1} + P_{11} H^T H] P_{12}, \tag{2.46}$$

$$P_{12}(t) = \Pi(t),$$

$$\frac{-d}{d\sigma} P_{11} = GG^T - [F + GG^T \Pi^{-1}] P_{11} - P_{11}[F + GG^T \Pi^{-1}]^T - P_{11} H^T H P_{11}, \tag{2.47}$$

$$P_{11}(t) = \Pi(t),$$

$$\frac{-d}{d\sigma} P_{22} = -P_{21} H^T H P_{12}, \tag{2.48}$$

$$P_{22}(t) = \Pi(t).$$

The term P_{11} gives the filtered error covariance and P_{22} gives the smoothed error covariance. The optimal smoothed estimate is

$$\hat{X}(\tau|t) = \hat{X}(\tau) + \Pi(t) \lambda_B(t, \tau) \tag{2.49}$$

where

$$\lambda_B(t, \tau) = \int_t^\tau \varphi^T(\sigma, t) H^T [dY_B(\sigma) - H \hat{X}_T(\sigma) d\sigma], \tag{2.50}$$

$$-dX_B(\sigma) = F_B X_B d\sigma + G dB_B(\sigma),$$

$$dY_B(\sigma) = H X_B(\sigma) d\sigma + d\tilde{B}_B(\sigma).$$

It should be noted that equation (2.48) differs in sign from that given by Ljung–Kailath (equation (84), [35]). Integrating (2.48) it is seen that

$$P_{22}(\tau)-P_{22}(t) = -\int_{\tau}^{t} P_{21}H^{T}HP_{12},$$

$$P_{22}(\tau) = \Pi(t)-\int_{\tau}^{t} P_{21}H^{T}HP_{12}. \tag{2.51}$$

Thus the error covariance for the smoothing solution is less than the error covariance for the filtering solution. This error covariance (2.51) is also consistent with the smoothed estimate (2.20).

3. Finite time filtering for linear pure delay time stochastic systems

Consider the linear pure delay time stochastic equations

$$dX(t) = \sum_{i=1}^{k} F_i X(t-\theta_i)\,dt + G\,dB(t), \tag{3.1}$$

$$dY(t) = \sum_{i=1}^{k} H_i X(t-\theta_i)\,dt + d\tilde{B}(t) \tag{3.2}$$

where $F_i \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^n)$, $H_i \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^n, \boldsymbol{R}^p)$, $i = 1, 2, \ldots, k$, $G \in$ $\in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^m, \boldsymbol{R}^n)$, $X(0)$ is a zero mean Gaussian random variable with covariance P_0, $X(s) \equiv 0$ for $s < 0$, $Y(s) \equiv 0$, $s \leqslant 0$, and $\{B(t)\}$ and $\{\tilde{B}(t)\}$ are independent standard $\boldsymbol{R}^m$- and $\boldsymbol{R}^p$-dimensional Brownian motions. These linear delay time stochastic equations can be described as a stochastic system over a ring of polynomials using the previous methods for linear delay time equations. Specifically the system will be denoted

$$dX(t) = FX\,dt + G\,dB(t), \tag{3.3}$$

$$dY(t) = HX\,dt + dB(t) \tag{3.4}$$

where $F \in \mathrm{Hom}_{\boldsymbol{R}}(M, M)$, $G \in \mathrm{Hom}_{\boldsymbol{R}}(\boldsymbol{R}^m, M)$, $H \in \mathrm{Hom}_{\boldsymbol{R}}(M, \boldsymbol{R}^p)$, R is the ring of polynomials $\boldsymbol{R}[\sigma_1, \ldots, \sigma_l]$ with transcendence degree l and M is the state module. Unless stated to the contrary it is always assumed that (F, G) is reachable (1.3) and (F, H) is observable (1.4).

It is elementary to verify that the equations (3.1)–(3.2) have one and only one solution. One method to accomplish this is to recall that the transfer function method of Kamen [27] can be used for locally square integrable functions. Using the canonical normal distribution in Hilbert space to construct Brownian motion, a sequence of solutions to locally square integrable inputs are obtained which can be shown to converge to the desired solution. The uniqueness of the solution will also follow

from this construction. The Gaussian initial conditions will imply the square integrability of the solution and the fact that it is also Gaussian.

Some filtering problems of the form (3.1)–(3.2) have been solved in [29], [30], [37].

A description will be given of the geometric formulation of the filtering problem. The stochastic system (3.3)–(3.4) can be viewed as an equation in the state module which is an algebraic vector bundle over Spec(R). Associated to the principal bundle of this state module are other bundles that are important in problems. For example, the estimation problem can be described as a symplectic algebraic vector bundle associated to the principal bundle. The symplectic structure can be defined pointwise over Spec(R) and it follows immediately that it can be pieced together algebraically. In general it is known [2] that a symplectic algebraic vector bundle is globally trivial. However, for the symplectic bundle for the filtering problem it is clear from the construction that it is globally trivial. In this geometric formulation it is important to recall the isomorphism between the convolution ring of Schwartz distributions and the ring of polynomials. This isomorphism implies that the geometric description provides a description in terms of functional analysis.

THEOREM 3.1. *Let* $(X(t), Y(t))_{t\in\mathbf{R}_+}$ *be the stochastic processes that satisfy the equations* (3.1)–(3.2) *where* (F, G) *is reachable and* (F, H) *is observable. The conditional mean,* $E[X(t)|Y(u), 0 \leqslant u \leqslant t]$, *satisfies the stochastic equation*

$$d\hat{X}(t;t) = \big(F\hat{X}(t;t) - [P(t,t;t)H^T] \qquad [H\hat{X}(t;t)]\big)dt + [P(t,t;t)H^T]dY(t) \tag{3.5}$$

where $\hat{X}(t;s) = E[X(t)|Y(r), 0 \leqslant r \leqslant s]$ *and*

$$\sigma_t\hat{X}(t;t) = \hat{X}(t-\tilde{\theta}_t;t),$$
$$P(t,t;t)\sigma_t = P(t,t-\tilde{\theta}_t;t),$$
$$\hat{X}(s;0) \equiv 0, \quad s \leqslant 0$$

where σ_t *is the operator corresponding to the delay* $\tilde{\theta}_t$ *and the brackets* $[\cdot]$ *are used to indicate where the delay operators act.*

The error covariance is P *and it satisfies the deterministic equations*

$$\frac{dP(t,t;t)}{dt} = FP + PF^T - [PH^T][HP] + GG^T, \tag{3.6}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \alpha}\right)P(t,t-\alpha;t) = FP(t,t-\alpha;t) - \qquad - [P(t,t;t)H^T][HP(t,t-\alpha;t)], \tag{3.7}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \alpha} + \frac{\partial}{\partial \beta}\right)P(t-\alpha,t-\beta;t) = -[P(t-\alpha,t;t)H^T][HP(t,t-\beta;t)] \tag{3.8}$$

where $P(0,0;0)=P_0$, $P(s,t;0)\equiv 0$ *for* $s<0$ *and* $t<0$, $\sigma_i P(t,s;r)=P(t-\tilde{\theta}_i,s;r)$ *and*

$$P(r,s;t)=E[\tilde{X}(r;t)\tilde{X}^T(s;t)],$$

$$\tilde{X}(s;t)=X(s)-\hat{X}(s;t).$$

Proof. For $t\in R_+$, $E|X_t|^2<\infty$ so the conditional mean $\hat{X}(t;t)$ and the error covariance $P(t,t;t)$ exist.

The conditional mean and the error covariance will be shown to satisfy equations that are obtained by geometric methods. Recall that for R, a ring of polynomials $R[\sigma_1,\ldots,\sigma_l]$, $\mathrm{Spec}(R)$ is affine space. For each point $x\in\mathrm{Spec}(R)=A^l$ localization at x gives $M \bmod x$ and $F(x)$, $G(x)$ and $H(x)$. The transfer function is also naturally localized. For this pointwise system we can use the usual results for filtering to obtain

$$\frac{dP_x}{dt}=F(x)P_x+P_xF^T(x)-[P_xH^T(x)]\,[H(x)P_x]+G(x)G^T(x). \tag{3.9}$$

It is clear that these pointwise Riccati equations can be pieced together locally algebraically. This local construction, for example at each maximal ideal, can be globalized to

$$\frac{dP}{dt}=FP+PF^T-[PH^T][HP]+GG^T. \tag{3.10}$$

Since pointwise this gives the error covariance, the optimal cost, it is clear that if there is one solution to (3.10) then it is the error covariance for the filtering solution. Again by a pointwise construction it follows that the conditional mean, the optimal estimate, satisfies the equation

$$d\hat{X}(t;t)=\big(F\hat{X}(t;t)-[P(t,t;t)H^T]\,[H\hat{X}(t;t)]\big)\,dt+[P(t,t;t)H^T]\,dY(t). \tag{3.11}$$

From the estimation problem it is clear that the error covariance exists. This is a simple argument of projection in the Hilbert space of random variables formed from the observations $(Y(t))$. Thus equation (3.11) defines the conditional mean because the error covariance P is known to exist and there is one and only one solution of (3.11) [13], [14]. To verify (3.10) it is only necessary to verify the differentiability of P, $\frac{dP}{dt}(t,t;t)$. Let $\tilde{X}(t;t)=X(t)-\hat{X}(t;t)$. The stochastic equation for $\tilde{X}$ is

$$d\tilde{X}(t;t)=F\tilde{X}\,dt+G\,dB(t)-[PH^T]\,[dY(t)-H\hat{X}\,dt]. \tag{3.12}$$

Since P is a continuous function, it can be shown that this equation has

a unique solution [14]. Recall that

$$P(t, t; t) = E[\tilde{X}(t; t)\tilde{X}^T(t; t)].$$

Integrating the stochastic equation for $\tilde{X}$, using the change of variables formula for products of ordinary and stochastic integrals [23] and interchanging expectation and the integrals obtained from the change of variables formula we have

$$P(t, t; t) = P_0 + \int_0^t FP\,ds + \int_0^t PF^T ds - \int_0^t [PH^T]\,[HP]\,ds + \int_0^t GG^T ds.$$

Thus $\dfrac{dP}{dt}$ exists and satisfies the formal Riccati equation (3.10).

Expressing the Riccati equation (3.10) as a delay differential equation it is clear that additional functions are necessary in order to solve this equation. These additional functions will be shown to satisfy certain partial differential equations.

Let $\tau > 0$ and consider $\hat{X}(\tau; t)$. From general results on Gaussian estimation it is known that

$$\hat{X}(\tau; t) = \int_0^t \varphi(\tau, s)\,d\bar{B}(s) \tag{3.13}$$

where $d\bar{B}(s) = dY(s) - H\hat{X}(s; s)\,ds$. The process $(\bar{B}(s))$ is sometimes called the innovations process. In addition it follows easily from the projection methods in the Hilbert space of random variables generated by the observations that

$$\varphi(\tau, s) = P(\tau, s; s)H^T. \tag{3.14}$$

Consider $P(r, s; t)$. Computing from the stochastic equation (3.1) it is easy to verify that the covariance of $X(r)$ is differentiable. For the moment consider X as scalar-valued. Expressing $X(r)$ in terms of an $L^2(P)$ basis of the observations and its orthogonal complement it is easy to verify that the covariance of $\hat{X}(r; t)$ is differentiable. The multidimensional generalization is trivial so that $P(r, s; t)$ has partial derivatives with respect to r and s. It follows from (3.12) and (3.13) and the change of variables formula for semimartingales that

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)P(t, t-a; t) = FP(t, t-a; t) - \\ - [P(t, t; t)H^T]\,[HP(t, t-a; t)] \tag{3.15}$$

and

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} + \frac{\partial}{\partial \beta}\right)P(t-a, t-\beta; t) \\ = -[P(t-a, t; t)H^T]\,[HP(t, t-\beta; t)]. \tag{3.16}$$

These equations (3.15)–(3.16) and equation (3.10) can also be formally derived using the method of augmenting the state as was done to solve the smoothing problem without delays (2.17). Specifically define the state as

$$\begin{bmatrix} X(t) \\ X(t-\alpha) \\ X(t-\beta) \end{bmatrix}.$$

To ensure that $X(t-\alpha)$ and $X(t-\beta)$ are "parameters", differentiation must be performed in directions that keep these constant. These directions correspond to the partial derivatives in (3.15)–(3.16).

4. Infinite time filtering for linear pure delay time stochastic systems

For linear systems without delays it is easy to show that the infinite time control problem is well posed assuming that the system is reachable. This follows from the fact that reachability is equivalent to the arbitrary placement of the poles of the system by state feedback [42], [48]. However, this equivalence is not true for linear pure delay time systems [4]. While there are a number of quite restrictive results for the equivalence of reachability and coefficient assignability of the characteristic polynomial by state feedback, to verify that the infinite time control problem is well posed only requires demonstrating one control for which the cost functional is finite. By duality the previous discussion could have been made for the infinite time filtering problem. Various results are available for these infinite time problems using functional analysis e.g. [15], [30], [46], [49]. However, when delays appear in the controls the conditions for applicability of these results are difficult to verify. The results that are given here use only the algebraic conditions of reachability and observability (1.3)–(1.4). These results were announced at IRIA in April 1979 [19] and at Oberwolfach in July 1979. Byrnes [6] has independently obtained a stabilizability result that could be used for these infinite time problems.

The steady state filtering problem arises when it is assumed that the infinite past of the observations, $\sigma\{Y(u), -\infty < u \leqslant t\}$, is available for the estimate of $X(t)$. While the finite time estimation problem always has a solution, it is well known that there are steady state estimation problems even for systems without delays that do not possess a solution. To solve the infinite time estimation problem for linear systems with pure delays it is necessary to establish a steady state property for the augmented Riccati equation (3.6)–(3.8). Specifically, it is necessary to show the convergence of a sequence of solutions defined on $[t_n, t]$ where

$t_n \downarrow -\infty$. Since it is more convenient to study a sequence of solutions on $[t, t_n]$ where $t_n \uparrow +\infty$, a deterministic optimal control problem for a linear pure delay time system will be introduced that will be naturally dua to the filtering problem described by the equations (3.1)–(3.2). This duality has been verified in [33]. However, it follows easily also by the geometric description of linear pure delay time systems. Using the duality between estimation and control for linear systems without delays [24] a pointwise duality over Spec(R) can be established for systems with delays. Since this pointwise result can be pieced together algebraically, global triviality follows as well as the duality for linear pure delay time systems.

The deterministic control problem is described by the equations

$$\frac{dx}{dt} = \sum [A_i x(t-\theta_i) + B_i u(t-\theta_i)], \tag{4.1}$$

$$y = Cx \tag{4.2}$$

with the cost functional

$$J(u; 0, T) = \int_0^T [\langle Cx(t), Cx(t)\rangle + \langle u(t), u(t)\rangle]dt + \langle Dx(T), x(T)\rangle \tag{4.3}$$

where $A_i = F_i^T$, $B_i = H_i^T$, $i = 1, \ldots, k$, $C = G^T$ and $D = P_0$. The linear pure delay time equation (4.1) can be expressed over the ring of polynomials R as

$$\frac{dx}{dt} = Ax + Bu \tag{4.4}$$

where $A \in \mathrm{Hom}_R(N, N)$, $B \in \mathrm{Hom}_R(R^p, N)$, $N = M^T$ and $R = R[\sigma_1, \ldots, \sigma_l]$ with transcendence degree l. Recall that (A, B) is reachable and (A, C) is observable. The family of admissible controls are all functions in $L^2([0, T], R^p)$. Using some elementary Hilbert space techniques it is easy to show that there is a unique admissible control that achieves the optimal cost.

The important feature of the duality between estimation and control that will be used here is that the optimal cost is computed from the solution of an augmented Riccati equation.

PROPOSITION 4.1. *For the deterministic control problem described by the equations* (4.1)–(4.3) *the solution of the following equations determines the*

optimal cost:

$$\frac{-dS(t,t;t)}{dt} = A^TS + SA - [SB][B^TS] + C^TC, \tag{4.5}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) S(t, t-a; t) = A^TS(t, t-a; t) - \tag{4.6}$$

$$-[S(t,t;t)B][B^TS(t,t-a;t)],$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a} + \frac{\partial}{\partial \beta}\right) S(t-a, t-\beta; t) \tag{4.7}$$

$$= [S(t-a, t; t)B][B^TS(t, t-\beta; t)]$$

where $S(T,T;T) = D$ and the elements of the ring R act on S in the same manner as in (3.6)–(3.8). Specifically, if $x = (x_1, x_2) \in \boldsymbol{R}^n \times \times L^2([-\tau, 0], \boldsymbol{R}^n)$ is the initial condition for (4.1), then the optimal cost J^* is

$$J^* = \langle S(0,0;0)x_1, x_1\rangle + 2\left\langle \int_{-\tau}^{0} S(0, -a; 0)x_2(a)da, x_1\right\rangle + \tag{4.8}$$

$$+ \int_{-\tau}^{0}\int_{-\tau}^{0} \langle S(-a, -\beta; 0)x_2(a), x_2(\beta)\rangle da\, d\beta.$$

When there are no delays in the control input, this result has been established by Delfour–Mitter [12]. For the control problem (4.1)–(4.3) this result can be verified by geometric methods. Pointwise over Spec(R) the optimal control results for systems without delays can be used. These equations piece together locally algebraically. Since there is a solution to these equations from the filtering results (3.6)–(3.8) it follows that these pointwise sufficient conditions for optimality provide the global sufficient conditions. This is an application of the Hamilton–Jacobi equation for sufficiency for an optimal control. The optimal cost (4.8) follows from these results.

For the infinite time optimal control problem to be well posed it must be demonstrated that there is a control such that $J(u; 0, +\infty)$ is finite. This fact will be demonstrated in the proof of the following result.

THEOREM 4.2. *The infinite time optimal control* (4.1)–(4.3) *where* $T = +\infty$, (A, B) *is reachable,* (A, C) *is observable and* $D \equiv 0$, *is well posed.*

A steady state solution to the equations (4.5)–(4.7) *exists as the solution of*

$$0 = A^T K(0,0) + K(0,0)A - [K(0,0)B]\,[B^T K(0,0)] + C^T C, \tag{4.9}$$

$$\frac{d}{da}K(0;a) = A^T K(0,a) - [K(0,0)B]\,[B^T K(0,a)], \tag{4.10}$$

$$-\left(\frac{\partial}{\partial a} + \frac{\partial}{\partial \beta}\right)K(a,\beta) = [K(a,0)B]\,[B^T K(0,\beta)] \tag{4.11}$$

where the elements of the ring R act on K dual to the action in (4.5)–(4.7). *The optimal cost for the infinite time control problem with the initial condition $x = (x_1, x_2)$ is*

$$J^* = \langle K(0,0)x_1, x_1\rangle + 2\left\langle \int_{-\tau}^{0} K(0,a)x_2(a)\,da, x_1\right\rangle + \tag{4.12}$$

$$+ \int_{-\tau}^{0}\int_{-\tau}^{0} \langle K(a,\beta)x_2(a), x_2(\beta)\rangle\,da\,d\beta.$$

Proof. Initially it will be shown that since (A, B) is reachable over R, there is a control such that the cost (4.3) with $T = +\infty$ is finite. A control will be constructed using the geometric description of the system. The usual projective embedding in linear system theory, sometimes called the Kalman embedding [9], and the method of pole placement or coefficient assignability for linear systems without delays shows that the coefficient assignability result varies locally algebraically with parameters. Thus locally it it possible to assign arbitrarily the coefficients of the characteristic polynomial by state feedback. Cover Spec(R) by Zariski open sets determined by the prime ideals such that the coefficient assignability result is satisfied in each of these sets. Since Spec(R) is quasi-compact, there is a finite covering by these open sets. Denote these open sets as $V_1, \ldots, V_k$, the corresponding prime ideals as $p_1, \ldots, p_k$. Form the local rings $R_1, \ldots, R_k$ where $R_j = R_{p_j}$. For each $j \in \{1, \ldots, k\}$ the linear system (4.4) localized by the local ring R_j, $j = 1, 2, \ldots, k$ has a control such that the infinite time problem has finite cost. This property follows because using state feedback locally the coefficients of the characteristic polynomial can be arbitrarily assigned. To solve the localized system using the transfer function over the local ring it suffices to show that it is a locally integrable function but this property is determined by the characteristic polynomial over the localized convolution ring of Schwartz distribution.

To define a global control that has finite cost for the given initial data it suffices to use

$$u = \sum_{j=1}^{k} u_j$$

where u_j is a control on V_j that gives finite local cost as described above. This pointwise description of a control over Spec(R) clearly gives a measurable, square integrable function. Elementary estimates show that the resulting cost is finite.

Let $S(t, t; t)$ be the solution of (4.5). Since the final time will be varied, it is necessary to be more explicit about the function S. Specifically, the time interval of the control problem will be indicated. Thus a solution of (4.5) will be described as $S(t, t; t, 0, T)$. Since the equations (4.5)–(4.7) are solved in reverse time, $S(t, t; t, 0, T)$ can be naturally identified with $S(t, t; t, t, T)$. Since the linear system is translation invariant by a translation of t, $S(t, t; t, t, T)$ can be computed as $S(0, 0; 0, 0, T-t)$. Since $\lim_{T\to\infty} S(0, 0; 0, 0, T-t)$ exists, so does $\lim_{T\to\infty} S(t, t; t, t, T)$.

Let $\mathscr{S}$ be the quadratic form on $\boldsymbol{R}^n \times L^2([-\tau, 0], \boldsymbol{R}^n)$ that gives the optimal cost (4.3) as a function of the initial condition. Let

$$\mathscr{S} = \begin{bmatrix} \mathscr{S}_{11} & \mathscr{S}_{12} \\ \mathscr{S}_{21} & \mathscr{S}_{22} \end{bmatrix}$$

where $\mathscr{S}$ is represented in block form according to the product $\boldsymbol{R}^n \times L^2([-\tau, 0], \boldsymbol{R}^n)$. Given $\varepsilon > 0$ there is a $T(\varepsilon)$ such that if $T < T_1 < T_2$ then

$$|\langle \mathscr{S}_{22}(t, T_1)x, y\rangle - \langle \mathscr{S}_{22}(t, T_2)x, y\rangle| \leqslant K|x|\,|y|\,\varepsilon \tag{4.13}$$

where K can be chosen independent of x, y and ε. From this inequality it follows immediately that $S(-\alpha, -\beta; 0, 0, T-t)$ converges weakly to a function k. Writing $\mathscr{S}_{22}(t, T)$ as an integral with kernel $S(-\alpha, -\beta; 0, 0, T-t)$ the inequality (4.13) shows that the partial derivatives with respect to α and β of this integral converge uniformly in T. This uniform convergence allows the interchange of these partial derivatives and the limit as $T \to +\infty$. Thus

$$\lim_{T\to\infty} S(-\alpha, -\beta; 0, 0, T-t) = k(\alpha, \beta) \text{ a.e.}$$

Again by the inequality (4.13) it is straightforward as above to show that $\lim_{T\to\infty} S(-\alpha, -\beta; 0, 0, T-t)$ is a continuous function of α and β. Thus k can be defined everywhere as this continuous limit. From these computations it also follows that $\lim_{T\to\infty} S(0, -\alpha; 0, 0, T-t)$ exists and is continuous. Thus as $T \to \infty$ the right-hand sides of (4.5)–(4.7) converge and therefore also the left-hand sides of these equations. Consider equation (4.6) and fix α. To show that the partial derivative with respect to t and the limit as $T \to \infty$ can be interchanged we can proceed as follows. Consider

the derivative in the direction $(1, 1)$, that is, given by $\frac{\partial}{\partial t}+\frac{\partial}{\partial a}$. A uniform estimate in T of the derivatives along this line allows the mean value theorem to give a uniform estimate to show that the differentiation and limit can be interchanged. A similar approach is used for equation (4.7). Thus we obtain the equations (4.9)–(4.11).

COROLLARY 4.2. *For the reachable and observable stochastic filtering equations* (3.1)–(3.2) *the steady state covariance equations exist and are obtained from the solution of the following equations*:

$$0 = FL(0,0)+L(0,0)F^T-[L(0,0)H^T][HL(0,0)]+GG^T, \tag{4.14}$$

$$-\frac{d}{da}L(0,a) = FL(0,a)-[L(0,0)H^T][HL(0,a)], \tag{4.15}$$

$$\left(\frac{\partial}{\partial a}+\frac{\partial}{\partial \beta}\right)L(a,\beta) = [L(a,0)H^T][HL(0,\beta)]. \tag{4.16}$$

A brief discussion will be provided now for the geometric approach to the stochastic optimal control of linear pure delay time systems. The details of the solution will be omitted because the methods are similar to the solution of the filtering problem. Consider the following linear pure delay time stochastic differential equation:

$$dX(t) = FX(t)dt+GdB(t)+CU(t)dt \tag{4.17}$$

where F and G are the same as in (3.1) and $C \in \mathrm{Hom}_R(R^p, M)$ and the cost functional is

$$J(u) = \tfrac{1}{2}E\Big[\int_0^T \langle QX(t), X(t)\rangle+\langle U(t), U(t)\rangle dt+\langle AX(t), X(T)\rangle\Big]. \tag{4.18}$$

It is assumed that (F, C) is reachable. To make this an interesting optimal control problem it is desirable to allow a large family of controls. The family of controls are required to be measurable with respect to the past information, but in general the control will affect the σ-algebras of the past of the state. One method to avoid these difficulties is to allow the control to be only a (suitably integrable) linear functional of the state. For these controls the solution of (4.17) exists and is unique and it can be easily shown that the σ-algebras generated by the state process do not vary as the controls are varied in this family. However, it is possible to consider a more general family of controls. This approach is in the spirit of some general necessary and sufficient conditions for stochastic optimal controls [11], [16], [40]. Let $(\mathcal{F}_t)$ be a fixed augmented increasing family of σ-algebras on a probability space $(\Omega, \mathcal{F}, P)$. A control U is admissible if it is $(\mathcal{F}_t)$-predictable and there is a unique law for the stochastic equation

(4.17) given an $(\mathscr{F}_t)$-Brownian motion. A sufficient condition for optimality is given by a smooth solution of the Hamilton–Jacobi (or dynamic programming) equation. Assuming that this solution is quadratic the equation can be reduced to the Hamilton–Jacobi equation of a deterministic optimal control problem which has a solution (4.5)–(4.7). Using these techniques it is also possible to obtain the solution to the stochastic control problem (4.17)–(4.18) with noisy partial observations.

References

[1] V. I. Arnol'd, *Characteristic classes entering in quantization conditions*, Functional Anal. Appl. 1 (1967), 1–13.

[2] H. Bass, *Unitary algebraic K-theory*, in: *Algebraic K-theory* III, Lecture Notes in Math. 343 (1972), Springer-Verlag, 57–265.

[3] R. W. Brockett, *Some geometric questions in the theory of linear systems*, IEEE Trans. Automatic Control AC-21 (1976), 444–455.

[4] R. Bumby, E. Sontag, H. Sussmann and W. Vasconcelos, *Remarks on the pole-shifting problem over rings*, J. Pure Appl. Algebra, 20 (1981), 13–18.

[5] C. I. Byrnes, *On the control of certain deterministic infinite-dimensional systems by algebro-geometric techniques*, Amer. J. Math. 100 (1978), 1333–1381.

[6] —, *On the stabilizability of linear control systems depending on parameters*, in: Proc. Eighteenth Conf. on Decision and Control, December 1979.

[7] C. Byrnes and N. Hurt, *On the moduli of linear dynamical systems*, Adv. in Math. Studies in Analysis 4 (1979), 83–122.

[8] C. I. Byrnes and T. E. Duncan, *Topological invariants arising in control theory*, to appear.

[9] —, —, *A note on the topology of spaces of Hamiltonian transfer functions*, Lectures in Appl. Math. 18 (1980), Amer. Math. Soc., 7–26.

[10] C. Carathéodory, *Variationsrechnung und partielle Differentialgleichungen erster Ordnung*, Teubner, Leipzig 1935.

[11] M. H. A. Davis and P. Varaiya, *Dynamic programming conditions for partially observable stochastic systems*, SIAM J. Control 11 (1973), 226–261.

[12] M. C. Delfour and S. K. Mitter, *Controllability, observability and optimal feedback control of affine hereditary differential systems*, ibid. 10 (1972), 298–328.

[13] —, —, *Hereditary differential systems with constant delays* I. *General case*, J. Differential Equations 12 (1972), 213–235.

[14] —, —, *Hereditary systems with constant delays* II. *A class of affine systems and the adjoint problem*, ibid. 18 (1975), 18–28.

[15] M. C. Delfour, C. McCalla and S. K. Mitter, *Stability and the infinite-time quadratic cost problem for linear hereditary differential systems*, SIAM J. Control 13 (1975), 48–88.

[16] T. E. Duncan, *Dynamic programming optimality criteria for stochastic systems in Riemannian manifolds*, Appl. Math. Optimization 3 (1977), 191–308.

[17] —, *A geometric approach to linear control and estimation*, Lecture Notes in Control and Information Sci. 16 (1979), Springer-Verlag, 118–141.

[18] —, *An algebro-geometric approach to estimation and stochastic control for linear pure delay time systems*, in: *Stochastic Control Theory and Stochastic Differential Equations*, Lecture Notes in Control and Information Sciences 16 (1979), Springer-Verlag, 332–343.

[19] T. E. Duncan, *On the steady state filtering problem for linear pure delay time systems*, Séminaires IRIA, 1979, 25–42.
[20] D. C. Fraser, *A new technique for optimal smoothing of data*, thesis, MIT, 1967.
[21] J. K. Hale, *Functional Differential Equations*, Springer-Verlag, 1971.
[22] R. Hermann and C. Martin, *Applications of algebraic geometry to systems theory: the McMillan degree and Kronecker indices of transfer functions as topological and holomorphic system invariants*, SIAM J. Control Optimization 16 (1978), 743–755.
[23] K. Itô, *On a formula concerning stochastic differentials*, Nagoya Math J. 3 (1951), 55–65.
[24] R. E. Kalman, *Contributions to the theory of optimal control*, Bol. Soc. Mat. Méxicana 1960, 102–119.
[25] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, J. Basic Engrg. Trans. ASME 83 (1961), 95–108.
[26] E. W. Kamen, *On an algebraic theory of systems defined by convolution operators*, Math. Systems Theory 9 (1975), 57–74.
[27] —, *An operator theory of linear functional differential equations*, J. Differential Equations 27 (1978), 274–297.
[28] F. Kappel, *Degenerate difference-differential equations: algebraic theory*, ibid. 24 (1977), 99–126.
[29] H. Kwakernaak, *Optimal filtering in linear systems with time delays*, IEEE Trans. Automatic Control AC-12 (1967), 169–173.
[30] R. H. Kwong and A. S. Willsky, *Estimation and filter stability of stochastic delay systems*, SIAM J. Control Optimization 16 (1978), 660–681.
[31] D. G. Lainiotis, *Optimal linear smoothing; continuous data case*, Internat. J. Control 17 (1973), 921–930.
[32] E. B. Lee and A. W. Olbrot, *On reachability over polynomial rings and a related genericity problem*, Proc. Fourteenth Conf. on Information Sci. and Systems, Princeton 1980.
[33] A. Lindquist, *A theorem on duality between estimation and control for linear stochastic systems with time delay*, J. Math. Anal. Appl. 37 (1972), 516–536.
[34] A. Lindquist, G. Picci and G. Ruckebusch, *On minimal splitting subspaces and Markovian representations*, preprint.
[35] L. Ljung and T. Kailath, *A unified approach to smoothing formulas*, Automatica 12 (1976), 147–157.
[36] D. Q. Mayne, *A solution of the smoothing problem for linear dynamic systems*, ibid. 4 (1966), 73–92.
[37] S. K. Mitter and R. B. Vinter, *Filtering for linear stochastic hereditary differential systems*, in: Intern. Symp. Control Theory, Numerical Methods, and Computer Systems Modelling, IRIA, Rocquencourt, France, June 1974.
[38] D. Quillen, *Projective modules over polynomial rings*, Invent. Math. 36 (1976), 167–171.
[39] R. Redheffer, *On the relation of transmission line theory to scattering and transfer*, J. Math. Phys. 41 (1962), 1–41.
[40] R. Rishel, *Necessary and sufficient dynamic programming conditions for continuous-time stochastic optimal control*, SIAM J. Control 8 (1970), 559–571.
[41] J.-P. Serre, *Modules projectifs et espaces fibrés à fibre vectorielle*, Sém. Dubreil–Pisot, no. 23, 1957–58.
[42] J. M. Simon and S. K. Mitter, *A theory of modal control*, Information and Control 13 (1968), 316–353.
[43] E. Sontag, *Linear systems over commutative rings: A survey*, Ricerche Automatica 7 (1976), 1–34.

[44] E. Sontag, *On split realizations of response maps*, Information and Control 37 (1978), 23–33.

[45] A. A. Suslin, *Projective modules over a polynomial ring are free*, Dokl. Akad. Nauk SSSR 229 (1976), 1063–1066 (Sov. Math. Dokl. 17 (1976), 1160–1164).

[46] R. B. Vinter, *Filter stability for stochastic evolution equations*, SIAM J. Control Optimization 15 (1977), 465–485.

[47] N. S. Williams and V. Zakian, *A ring of delay operators with applications to delay-differential systems*, ibid. 15 (1977), 247–255.

[48] W. M. Wonham, *On pole assignment in multi-input controllable linear systems*, IEEE Trans. Automatic Control AC-12 (1967), 660–665.

[49] J. Zabczyk, *Remarks on the algebraic Riccati equation in Hilbert space*, Appl. Math. Optimization 2 (1975/76), 251–258.

[50] L. Zachrisson, *On optimal smoothing of continuous-time Kalman processes*, Information Sci. 1 (1969), 143–172.