FILTERING FOR LINEAR STOCHASTIC HEREDITARY DIFFERENTIAL SYSTEMS*

by

Sanjoy K. Mitter(1) and Richard B. Vinter(2)

---

(1) Electrical Engineering Department and Electronic Systems Laboratory, M.I.T., Cambridge, Mass. 02139, U.S.A. and Consiglio Nazionale delle Ricerche, Laboratorio per Ricerche di Dinamica dei Sistemi e di Elettronica Biomedica, Italy.

(2) Electronic Systems Laboratory, M.I.T., Cambridge, Mass. 02139,USA.

* The research of the first author was supported by NSF Grant GK-41647, AFOSR Grant 72-2273 and NASA Grant NGL-22-009-124, all at the Electronic Systems Laboratory, M.I.T., Cambridge, Mass. 02139 and by the Consiglio Nazionale delle Ricerche, Italy. The research of the second author was supported by the Commonwealth Fund (Harkness Fellowship).

§1. Introduction

The theorem on the separation of control and filtering (Wonham 1) for linear optimal stochastic control problems with Gaussian noise processes and a quadratic cost function constitutes one of the central results of stochastic control theory. If the linear system is autonomous, then under appropriate stabilizability, observability, detectability and reachability hypotheses it can be shown that the cascade combination of the Kalman filter and the regulator defines an asymptotically stable closed-loop system.

In previous papers (cf. Delfour-Mitter 1 , Delfour-McCalla-Mitter 1) it has been shown that the infinite-time quadratic cost problem for a general class of hereditary systems can be satisfactory solved. In this paper the filtering problem for infinite dimensional systems described by integral equations involving evolution operators is first studied. It is then shown how these results may be specialized to solve the filtering problem for linear stochastic hereditary differential systems.

This paper may be considered to be an application of the work of (Bensoussan 1) to a class of problems somewhat larger than that considered by him in his book. The unbounded linear operator involved in hereditary systems does not satisfy any coercivity conditions nor is it a generator of a contraction semi-group. Nevertheless, it is possible to exploit the structure of the operator (in particular its spectral properties) to obtain reasonably complete results for the filtering problem. For details regarding the properties of this operator the reader is referred to (Delfour 1 and Vinter 4).

The point of view taken in this paper is that for control purposes it is necessary to estimate the "state" of the system.

For hereditary systems, this means that it is first necessary to set up the stochastic evolution equation corresponding to the stochastic functional differential equation describing the evolution of the state of the system, study its properties and prove that the stochastic evolution equation is an equivalent description of the system. This is a key-step and requires the use of the detailed structure of the evolution operator, its generator and the adjoint of the generator. Once this result is at hand, it is possible to use the general theory developed in earlier sections to obtain the Kalman filter.

The question of filter stability is then studied. It is shown that under an appropriate detectability hypothesis (which is verifiable) the covariance of the estimate, is bounded. In the general case it has not been possible (so far) to give reasonably weak conditions under which the operator defining the Kalman filter is asymptotically stable. Nevertheless if the original system is exponentially stable and the "forcing terms" in the filter are not too large then the asymptotic stability of the filter is assured. This result is significant in generating at least a non-trivial class of stable filters.

The filtering problem for hereditary systems has previously been considered by(Kwakernaak 1 and Lindquist). The approach used here (in particular the manner in which the dual control problem is used) and the emphasis on filter stability appear to be new. Moreover in this framework using the ideas of(Bensoussan-Viot 1) the separation theorem for hereditary systems can be proved (cf. also Lindquist 3).

Thus the linear-quadratic-gaussian problem for hereditary systems is almost as complete as that for linear ordinary differential equations.

§2 Some Preliminary Definitions

Take $X, \mathcal{U}$ real separable Hilbert spaces, $(\Omega, \mathcal{S}, \mu)$ a complete probability space.

§§2.1 Separable Hilbert space-valued random variables

The reader is referred to (Bensoussan 1, ch.3), (Grenander 1, ch.6) or (Barucha-Reid 1, ch.1) for more detailed exposition of this material.

$x : \Omega \to X$ is called an X-valued random variable (r.v.) if it is a (weakly) measurable map. The linear space of X-valued r.v.'s is denoted $Mes(\Omega, \mu; X)$.

An X-valued stochastic process is a map $x(\cdot) : \mathbb{R}^+ \to Mes(\Omega, \mu; X)$. $x(\cdot)$ is a measurable process if the map $(t,\omega) \longmapsto x(t,\omega)$ is measurable w.r.t. $\overline{\mu_{\mathcal{L}} \times \mu}$ ($\mu_{\mathcal{L}}$ denotes Lebesgue measure on $\mathbb{R}^+$).

$x \in Mes(\Omega, \mu; X)$ is first order if $x \in L^1[\Omega, \mu; X]$ and second order if $x \in L^2[\Omega, \mu; X]$. For a first order r.v. $x(\omega)$ we define the mean $E\{x(\omega)\}$, $(\bar{x})$

$$E\{x(\omega)\} = \int_\Omega x(\omega)\, d\mu \qquad \text{(Bochner Integral)}$$

For a second order r.v. $x(\omega)$, $(h, \bar{h}) \mapsto E\{\langle x(\omega) - E\{x(\omega)\}, h\rangle \langle x(\omega) - E\{x(\omega)\}, \bar{h}\rangle\}$ is a continuous, symmetric bilinear form which has unique representation through $Q \in \mathcal{L}(X)$, $Q \geq 0$, $Q^* = Q$ as $(h, \bar{h}) \mapsto \langle Qh, \bar{h}\rangle$. Q is the covariance of $x(\omega)$. The covariance of a second order random variable $x(\omega)$ is necessarily nuclear (Grenander 1 p. 129).

Given two X-valued second order r.v.'s $x(\omega), y(\omega)$, $(h, \bar{h}) \mapsto E\{\langle x(\omega) - \bar{x}, h\rangle \langle y(\omega) - \bar{y}, \bar{h}\rangle\}$ has unique representation $(h, \bar{h}) \mapsto \langle Rh, \bar{h}\rangle$, $R \in \mathcal{L}(X)$. R is called the covariance of $x(\omega), y(\omega)$ and is written $\operatorname{cov}\{x(\omega), y(\omega)\}$.

$x(\omega), y(\omega) \in Mes(\Omega, \mu; X)$ are independent if $\langle h, x(\omega)\rangle, \langle \bar{h}, y(\omega)\rangle$ are independent for all $h, \bar{h} \in X$. $x \in L^2[\Omega, \mu; X]$ is Gaussian if $\langle x(\omega), h\rangle$ is normally distributed for each $h \in X$.

§§2.2 Wiener Processes

The $\mathcal{U}$-valued stochastic process $W(t,\omega)$ is a Wiener process if (i) for finite collections $\{t_i\} \in \mathbb{R}^+, \{e_j\} \in \mathcal{U}$, $(W(t_i,\omega), e_j)$ is a family of real-valued gaussian r.v.'s (ii) $W(t,\omega)$ is second order for each $t \geq 0$ and there exists some nuclear $Q \in \mathcal{L}(X)$ s.t.

$$E\{\langle W(t_1,\omega), h\rangle \langle W(t_2,\omega), \bar{h}\rangle\} = \langle Qh, \bar{h}\rangle \min\{t_1, t_2\}$$

each $t_1, t_2 \geq 0$, $h, \bar{h} \in \mathcal{U}$ (iii) $E\{W(t,\omega)\} = 0$ each $t \geq 0$. See (Bensoussan 1, p. 167 et seq.) for properties of $W(t,\omega)$.

Notice that since Q is nuclear, $Q \geq 0$, $Q^* = Q$

$$Q(\cdot) = \sum \lambda_i e_i \langle e_i, \cdot\rangle$$

for some $\{\lambda_i\}, \lambda_i \geq 0$ with $\sum \lambda_i < \infty$ *, some orthonormal sequence $\{e_i\}$ in $\mathcal{U}$. We shall make use of the property (Curtain 1) that the Wiener process $W(t,\omega)$ has unique representation

$$W(t,\omega) = \lim_{N\to\infty} \sum_{i=1}^{N} \beta_i(t,\omega) e_i \qquad (\text{limit in } L^2[\Omega, \mu; X])$$

with the β_i's independent real valued Wiener processes.

* (Gelfand and Vilenkin 1)

§§2.3 The Wiener Integral

Suppose $b : \mathbb{R}^+ \to X$ is locally essentially bounded, measurable and that $\beta(t,\omega)$ is a real-valued Wiener process. Then the Wiener Integral

$$\int_0^T b(t)\, d\beta(t,\omega)$$

is defined in the usual manner as a limit in $L^2[\Omega, \mu; X]$ through a sequence of simple functions approximating $b(t)$ in $L^2[0,T; X]$. Now suppose that $B(\cdot) : \mathbb{R}^+ \to \mathcal{L}(\mathcal{U}, X)$ satisfies (i) $\|B(\cdot)\|$ is locally essentially bounded, measurable (ii) $t \mapsto B(t)x$ is measurable for each $x \in X$. The Wiener Integral

$$\int_0^T B(t)\, dW(t,\omega)$$

is defined in this case as

$$\lim_{N\to\infty} \sum_{i=1}^{N} \int_0^T B(t) e_i\, d\beta(t,\omega) \qquad (\text{limit in } L^2[0,T; X])$$

where each element in the sequence is evaluated as above. $(e_i, \beta_i(t,\omega)$ $i = 1, 2, ..$ as in §§2.2). For $B(\cdot)$ measurable w.r.t. the uniform operator topology this definition* coincides essentially with that in (Bensoussan 1, p. 180 et seq.). Notice that the Wiener Integral is defined modulo null-functions in $L^2[\Omega, \mu; X]$.

§§2.4 Properties of the Wiener Integral

Suppose $B(\cdot) : \mathbb{R}^+ \to \mathcal{L}(\mathcal{U}, X)$ is such that $\|B(\cdot)\|$ is measurable and locally essentially bounded and such that $B(\cdot)x$ is measurable for each $x \in X$. We have the following easily derived properties

(i) $\langle x, \int_0^T B(t)\, dW(t,\omega)\rangle = \int_0^T \langle B^*(t)x, \cdot\rangle\, dW(t,\omega)$ w.p. 1

(ii) $E\{\|\int_0^T B(t)\, dW(t,\omega)\|^2\} \leq \operatorname{tr}\{Q\} \int_0^T \|B(t)\|^2 dt$

(iii) $E\{\int_0^T \langle b_1(t), \cdot\rangle\, dW(t,\omega) \int_0^T \langle b_2(t), \cdot\rangle\, dW(t,\omega)\} = 0$

for $W(t,\omega), V(t,\omega)$ independent Wiener processes, $b_1(\cdot), b_2(\cdot) \in L^\infty_{loc}[\mathbb{R}^+; \mathcal{U}]$.

(iv) $E\{\int_0^{t_1} \langle b_1(t), \cdot\rangle\, dW(t,\omega) \int_0^{t_2} \langle b_2(t), \cdot\rangle\, dV(t,\omega)\} = \int_0^{\min\{t_1,t_2\}} \langle b_1(t), Q\, b_2(t)\rangle\, dt$

(b_1, b_2 as above).

Finally suppose that $\Phi(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathcal{L}(\mathcal{U}, X)$ is locally essentially bounded and such that $\Phi(\cdot,\cdot)x$ is $\overline{\mu_{\mathcal{L}} \times \mu_{\mathcal{L}}}$ - measurable for each $x \in \mathcal{U}$. Then**

(v) there exists a measurable version of

$$z(s) = \int_0^T \Phi(s,t)\, dW(t,\omega)$$

and (vi) (integration by parts)

$$\int_0^T \left[\int_0^T \Phi(s,t)\, ds\right] dW(t,\omega) = \int_0^T \left[\int_0^T \Phi(s,t)\, dW(t,\omega)\right] ds \qquad \text{w.p. 1}$$

(a measurable version of the integrand being chosen on the right).

§§ 2.5 Perturbed Evolution Equations

Take X a real Hilbert space, $\mathcal{P} = \{(t,s) \in \mathbb{R}^2 \mid t \geq s \geq 0\}$. We say the map $T(\cdot,\cdot) : \mathcal{P} \to \mathcal{L}(X)$ is a mild evolution operator if (a) $T(\cdot,\cdot) : \mathcal{P} \to \mathcal{L}_s(X)$ is continuous (subscript s denotes "w.r.t. the strong operator topology")

*Suppose $T(\cdot)$ is a C_0 semigroup then $T(\cdot)$ is not strongly measurable w.r.t. the uniform operator topology (except in the trivial situation where $T(\cdot)$ is uniformly continuous). It is precisely because we wish to attach meaning to such integrals as $\int_0^T T(t)\, dW(t,\omega)$ that we must adopt the definition of Wiener integral given here.

** See (Vinter 1)

and (b) $T(t,\tau)T(\tau,s) = T(t,s)$, $T(t,t) = I$ (identify in $\mathcal{L}(X)$) $t \geq \tau \geq s$.

Definition 2.1 Let $T(\cdot,\cdot)$ be a mild evolution operator and suppose that $B: \mathbb{R}^+ \to \mathcal{L}(X)$ is essentially locally bounded and has the property: given any $f \in C[\mathbb{R}^+_s; X_s]$, $t \mapsto B(t)f(t)$ is strongly measurable. Then the B-perturbed mild evolution operator $T_B(\cdot,\cdot)$ (corresponding to $T(\cdot,\cdot)$) is that unique mild evolution operator $\mathcal{U}(\cdot,\cdot)$ such that $\mathcal{U}(t,s)x_0 = T(t,s)x_0 + \int_s^t T(t,\sigma)B(\sigma)\mathcal{U}(\sigma,s)x_0 d\sigma$ all $t \geq s$ for each $s \in \mathbb{R}^+$, each $x_0 \in X$.

It is shown in e.g. (Vinter 2) that $T_B(\cdot,\cdot)$ is well-defined.

§3. System Description

Let the X-valued and $\mathbb{R}^k$-valued processes $\{x(t) | t \geq 0\}$, $\{z(t) | t \geq 0\}$ be defined respectively by

$$x(t) = T(t,0)x_0 + \int_0^t T(t,\sigma)B(\sigma)dw(\sigma) \qquad \text{(process)} \qquad (3.1)$$

$$z(t) = \int_0^t C(\tau)x(\tau)d\tau + \int_0^t F(\tau)dv(\tau) \qquad \text{(observation)} \qquad (3.2)$$

Here,

$X, \mathcal{U}$ are real separable Hilbert spaces .

$\{T(t,s) \in \mathcal{L}(X) | (t,s) \in \mathcal{P}\}$ is a mild evolution operator.

$\{w(t) | t \geq 0\}$, $\mathcal{U}$-valued Wiener process with covariance operator W, W some nuclear operator.

$\{z(t) | t \geq 0\}$, $\mathbb{R}^k$-valued Wiener process with covariance operator V . We assume (strongly) sample continuous versions of $w(t)$, $v(t)$ to have been chosen. This is possible (see Bensoussan 1 , p. 179).

$B(\cdot) \in L^\infty_{loc}[\mathbb{R}^+; \mathcal{L}(\mathcal{U},X)]$, $C(\cdot) \in L^\infty_{loc}[\mathbb{R}^+; \mathcal{L}(X,\mathbb{R}^k)]$, $F(\cdot) \in L^\infty_{loc}[\mathbb{R}^+; \mathcal{L}(\mathbb{R}^k,\mathbb{R}^k)]$.

x_0 is a X-valued Gaussian random variable with zero mean and covariance P_0 .

All random variables are defined w.r.t. the same complete probability space. $dw(t)$, $dv(t)$, x_0 are assumed independent.

In (3.1) $\int_0^t T(t,\sigma)B(\sigma)dW(\sigma)$ is a Wiener integral. It is known that there exists a measurable version of $x(\cdot)$ with summable sample paths; such a version is used in evaluating the observation $z(\cdot)$.

Assumption 3.1 There exists $\nu > 0$ such that

$$\langle F(t)VF^*(t)h,h\rangle \geq \nu \|h\|^2 \quad h \in \mathbb{R}^k \qquad \text{(uniformly in } t \in \mathbb{R}^+ \text{)} \blacksquare$$

Assumption 3.2 The map $T^*(\cdot,\cdot): \mathcal{P} \to \mathcal{L}(X)$ is strongly continuous. ▌

We remark that Assumption (3.2) is immediately satisfied if $T(\cdot,\cdot)$ is "time-invariant". ▌

From the properties of the stochastic integral, the local essential boundedness of $\|B(\cdot)\|$ and the independence of x_0 and $w(t)$ we deduce the following properties of the process

(i) $x(t)$ is a gaussian r.v. for each $t \geq 0$

(ii) $E\{x(t)\} = 0$

(iii) writing $\Lambda(t,s) = \mathrm{cov}\{x(t),x(s)\}$ we have

$$\langle \Lambda(t,s)h,\bar{h}\rangle = \langle T(t,0)P_0T^*(t,0)h,\bar{h}\rangle + \int_0^{\min\{t,s\}} \langle B^*(\tau)T^*(s,\tau)h, WB^*(\tau)T^*(t,\tau)\bar{h}\rangle d\tau,$$

$h,\bar{h} \in X$. We have $\Lambda(t,s) \in C_{loc}[\mathbb{R}^+, \mathcal{L}_s(X)]$, $\Lambda^*(t,s) = \Lambda(s,t)$ and $\|\Lambda(t,s)\|$ is locally bounded on $\mathcal{P}$.

§4. The Filtering Problem

We remark that for every $k \in L^2[0,t; \mathbb{R}^k]$

$$\int_0^t \langle k(s),\cdot\rangle dz(s,\omega) \triangleq \int_0^t \langle k(s), C(s)x(s,\omega)\rangle ds + \int_0^t \langle k(s), F(s)(\cdot)\rangle dv(s,\omega)$$

is a well-defined second order random variable. Define $Y_t \triangleq \{y \in L^2[\Omega,\mathbb{R}] | y = \int_0^t k(s)dz(s,\omega)$ for some $k \in L^2[0,t; \mathbb{R}^k]\}$. For each $t \geq 0$, the mapping

$X \to L^2[0,t; \mathbb{R}^k]$, $h \mapsto k_h(t,\cdot) \in L^2[0,t; \mathbb{R}^k]$

is called a mild solution to the filtering problem (at time t) if, writing

$$\hat{x}_h(t,\omega) = \int_0^t \langle k_h(t,s),\cdot\rangle dz(s,\omega) \qquad \text{each } h \in X \quad , \qquad (4.1)$$

we have

$$E\{ |\langle h, x(t,\omega)\rangle - \hat{x}_h(t,\omega)|^2 \} \leq E\{ |\langle h, x(t,\omega)\rangle - y|^2 \} \qquad (4.2)$$

all $y \in Y_t$.

Suppose $k_h(t,\cdot)$ is a mild solution to the filtering problem. Then $\hat{x}_h(t,\omega)$ defined by (4.1) is called an optimal linear estimate of $\langle x(t,\omega),h\rangle$. Further, if $k_h(t,\cdot)$ has representation

$$k_h(t,\cdot) = K^*(t,\cdot)h \qquad \text{all } h \in X \quad \text{a.e. on } [0,t]$$

where $K(t,\cdot) \in L^2[0,t; \mathcal{L}(\mathbb{R}^k; X)]$, $K(t,\cdot)$ is called a (strong) solution to the filtering problem and the X-valued random variable $\hat{x}(t,\omega) = \int_0^t K(t,s)dz(s,\omega)$ is called an optimal linear estimate of $x(t,\omega)$. We remark that if $\hat{x}(t,\omega)$ is an optimal linear estimate of $x(t,\omega)$ then, in particular,

$$E\{ \|x(t,\omega) - \hat{x}(t,\omega)\|^2 \} \leq E\{ \|x(t,\omega) - \int_0^t K(t,s)dz(s,\omega)\|^2 \}$$

for all $K(t,\cdot) \in L^2[0,t; \mathcal{L}(\mathbb{R}^k, X)]$.

§5. The Mild Wiener-Hopf Equation

Lemma 5.1 Given $k_h(t,\cdot) \in L^2[0,t; \mathbb{R}^k]$, $h \in X$ define

$$\hat{x}_h(t) = \int_0^t \langle k_h(t,s),\cdot\rangle dz(s), \quad \tilde{x}_h(t) = \langle x(t),h\rangle - \hat{x}_h(t)$$

Then,

$\hat{x}_h(t)$ is an optimal linear estimate of $\langle x(t),h\rangle \Leftrightarrow E\{\tilde{x}_h(t)\langle z(\tau),p\rangle\}=0$, all $\tau\in[0,t], p\in\mathbb{R}^k$ ∎

Proof This is a simple consequence of the projection theorem. See (Vinter and Mitter 1) for details. c.f. Theorem 2.1 of (Curtain 2). ∎

Mild solutions to the filtering problem will be characterized through the Mild Wiener-Hopf Equation :

$$\int_0^t \mathcal{C}(\sigma)\Lambda(\sigma,s)\mathcal{C}^*(s)k_h(t,s)ds + \mathcal{F}(\sigma)N\mathcal{F}^*(\sigma)k_h(t,\sigma) = \mathcal{C}(\sigma)\Lambda(\sigma,t)h \quad \text{a.e. } \sigma\in[0,t] \tag{5.1}$$
$$k_h(t,\cdot)\in L^2[0,t;\mathbb{R}^k],\ h\in X.$$

Lemma 5.2 Within the class of $L^2[0,t;\mathbb{R}^k]$ functions, the mild Wiener Hopf equation (5.1) has a unique solution $k_h(t,\cdot)$. We have

$$\text{ess. sup}\{|k_h(t,\cdot)|\,|\,[0,t]\} \leq \text{constant } \|h\| \tag{5.2}$$

(constant independent of h). ∎

Proof Taking note of Assumption 3.1, it is easily shown that the map $L^2[0,t;\mathbb{R}^k]\to L^2[0,t;\mathbb{R}^k]$ defined by $k_h(t,\cdot)\mapsto$ {L.H.S. of Wiener-Hopf equation} has a bounded inverse. This gives uniqueness.

We also find that this map, when restricted to L^∞ takes values in L^∞ , and the restriction viewed as a map $L^\infty\to L^\infty$ has a bounded inverse. (5.2) readily follows. See Vinter and Mitter 1 for details. ∎

Lemma 5.3 For given $t\geq 0, h\in X$, let $k_h(t,\cdot)\in L^2[0,t;\mathbb{R}^k]$ satisfy the Mild Wiener-Hopf equation (5.1). Then $k_h(t,\cdot)$ is a mild solution to the filtering problem. ∎

Proof Let $k_h(t,\cdot)\in L^2[0,t;\mathbb{R}^k]$ satisfy the mild Wiener-Hopf equation. Defining

$$\tilde{x}_h(t) = \langle x(t),h\rangle - \int_0^t \langle k_h(t,s),\cdot\rangle\, dz(s)$$

we show by direct expansion that

$$E\{\tilde{x}_h(t)\langle z(\tau),p\rangle\} = 0 \qquad \text{each } \tau\in[0,t],\ \text{all } p\in\mathbb{R}^k$$

(see Vinter and Mitter 1 for details.) The assertion now follows from Lemma 5.1. ∎

Proposition 5.1 For each $t\geq 0, h\in X$ let $k_h(t,\cdot)$ be the unique solution to the mild Wiener-Hopf equation (5.1). Let $K(t,\cdot)$ be the unique element in $L^\infty[0,t;\mathcal{L}(\mathbb{R}^k,X)]$ such that

$$K^*(t,\cdot)h = k_h(t,\cdot) \quad \text{a.e. on } [0,t],\ \text{every } h\in X \tag{5.3}$$

Then $K(t,\cdot)$ is the unique solution to the filtering problem. ∎

Proof Again see (Vinter and Mitter 1) for details. That (5.3) well-defines $K(t,\cdot)$ as a $L^\infty[0,t;\mathcal{L}(\mathbb{R}^k,X)]$ element is easily deduced from (5.2). In view of Lemmas (5.2), (5.3) it remains to show that the filtering problem has a unique mild solution. But the optimal linear estimate $\hat{x}_h(t)$ of $\langle h,x(t)\rangle$ is unique being a projection. We have merely to show therefore that $\hat{x}_h(t)$ has a unique representation

$$\hat{x}_h(t) = \int_0^t \langle k(s),\cdot\rangle\, dz(s) \qquad k(\cdot)\in L^2[0,t;\mathbb{R}^k]$$

But this follows simply from the coercivity of $\mathcal{F}N\mathcal{F}^*$ hypothesized in Assumption (3.1). ∎

§6. Global Optimality of the Filter

Take a fixed time interval $[0,T]$. Then it can be shown that $\tilde{z}:\Omega\to$ {maps from $[0,T]$ into $\mathbb{R}^k$} defined by

$$\tilde{z}(\omega) = \int_0^{(\cdot)} \mathcal{C}(\tau)x(\tau,\omega)d\tau + \mathcal{F}(\cdot)v(\cdot,\omega)$$

takes values almost surely in $L^2[0,T;\mathbb{R}^k]$ and $\omega\mapsto\tilde{z}(\omega)$ defines an $L^2[0,T;\mathbb{R}^k]$-valued gaussian random variable.

Define $\mathcal{H}=L^2[0,T;\mathbb{R}^k]$ and let σ be the measure induced on the Borel fields of $\mathcal{H}$ under $\tilde{z}$. Define further the closed subspace $\tilde{\mathcal{H}}$ of $L^2[\Omega,\mu;X]$ as

$$\tilde{\mathcal{H}} = \{\varphi\in L^2[\Omega,\mu;X]\,|\,\varphi(\omega)=g(\tilde{z}(\omega)),\ \text{some } g\in L^2[\mathcal{H},\sigma;X]\}$$

Any element in $\tilde{\mathcal{H}}$ is called an estimate of $x(T,\omega)$. The projection of $x(T,\omega)$ onto $\tilde{\mathcal{H}}$ is called the optimal non-linear estimate of $x(T,\omega)$.

Now in (Benssonssan 1 , Ch. 6), where filtering of systems for which the generator of $T(\cdot,\cdot)$ satisfies certain coercivity conditions is treated, it is shown that the optimal linear estimate $\hat{x}(T,\omega)$ of §4 is the optimal non-linear estimate of $x(T,\omega)$ also. The argument carries over almost unaltered to assure that $\hat{x}(T,\omega)$ is still the optimal non-linear estimate of $x(T,\omega)$ in the more general setting under consideration here.*
In outline :

(a) Let $x_1(\omega), x_2(\omega)$ be X_1, X_2 -valued zero-mean gaussian random variables respectively (X_1, X_2 real, separable Hilbert spaces), let $\bar{\sigma}$ be the measure induced on the Borel sets of X_2 under $x_2(\omega)$, let $G=L^2[X_2,\bar{\sigma};X_1]$ and

$$\tilde{G} = \{x\in L^2[\Omega,\mu;X_1]\,|\,x(\omega)=g(x_2(\omega)),\ \text{some } g\in G\}$$

Then a familiar finite-dimensional argument can be patterned (see Bensoussan p. 117-119) to show that if $\tilde{\Lambda}\in\mathcal{L}(X_2,X_1)$ satisfies

$$E\{\langle x_1(\omega)-\tilde{\Lambda}x_2(\omega),h_1\rangle\langle x_2(\omega),h_2\rangle\} = 0\ ,\quad \text{every } h_1\in X_1,\ h_2\in X_2$$

then $\tilde{\Lambda}x_2(\omega)$ is the projection of $x_1(\omega)$ onto $\tilde{G}$, i.e. the optimal linear estimate (if it exists) coincides with the optimal non-linear estimate.

(b) $\qquad \omega\mapsto\int_0 \Gamma(t)\,dz(t,\omega)$

defines an element in $\tilde{\mathcal{H}}$ for any $\Gamma\in L^\infty[0,T;\mathcal{L}(\mathbb{R}^k,X)]$ and in particular for $\Gamma(t)=K(T,t)$ ($K(T,t)$ as in Proposition 5.1) (Bensoussan 1, pp. 230-231). Thus the estimate $\hat{x}(T,\omega)$ of §4 is a linear estimate of $x(T,\omega)$ given $\tilde{z}(\omega)$.

(c) Finally, for arbitrary $h\in X$, we have that

$$E\{\langle x(T,\omega)-\int_0^T K(T,t)dz(t,\omega),h\rangle \int_0^T\langle z(t,\omega),f(t)\rangle dt\} = 0$$

for $f(t)$ a simple $\mathbb{R}^k$-valued function by Lemma 5.1. A limiting argument gives the relation for general $f\in L^2[0,T;\mathbb{R}^k]$. Since $x(T,\omega)$ and $\tilde{z}(\omega)$ are zero-mean gaussian r.v.'s we deduce that $\hat{x}(T,\omega)$ is the optimal nonlinear estimate from (a) and (b).

*Further, for any bounded linear map $Q:X\to Y$, Y a separable, real Hilbert space, $Q\hat{x}(T,\omega)$ is the optimal non-linear estimate of $Qx(T,\omega)$.

§7. Duality of Estimation and Control

We introduce

Control Problem For given $t \geq 0,\ h \in X$

$$\begin{cases} \text{minimize } J_t(u,h) \\ \text{subject to } u \in L^2[0,t;\mathbb{R}^k] \text{ and} \\ p(\tau) = T^*(t,\tau)h + \int_\tau^t T^*(s,\tau)\mathcal{C}^*(s)u(s)\,ds \end{cases}$$

Here,

$$J_t(u,h) = \int_0^t [\langle p(\tau), B(\tau)W B^*(\tau)p(\tau)\rangle + \langle u(\tau), F(\tau)V F^*(\tau)u(\tau)\rangle]\,d\tau + \langle p(0), P_0\, p(0)\rangle$$

A minimizing u is called an optimal control (for given $t \geq 0, h \in X$).

Now it has been observed (Bensoussan 1, p. 165) that there is in some sense a "duality" between such a control problem and the filtering problem. In the present framework it is convenient to change the emphasis somewhat; rather than merely view this duality as a noteworthy structural property, we exploit it in the actual dynamical filter construction in a fundamental way :

Theorem 7.1 (Duality of Estimation and Control).
For given $t \geq 0,\ h \in X,\ k_h(t,\cdot) \in L^2[0,t;\mathbb{R}^k]$ is a mild solution to the filtering problem if and only if $-k_h(t,\cdot)$ is an optimal control. ▮

Proof We refer to Vinter and Mitter 1 for details. An outline is the following. For given $t \geq 0,\ h \in X$ the "cost function" $J_t(u,h)$ is expressed as

$$J_t(u,h) = \pi(u,u) - 2\lambda(u) + \{\text{terms independent of } u\}$$

where $\pi(\cdot,\cdot): L^2 \times L^2 \to \mathbb{R}$ is continuous, bilinear, symmetric, coercive and $\lambda(\cdot): L^2 \to \mathbb{R}$ is bounded, linear.

By standard results concerning minimization of quadratic forms (see e.g. Lions 1), there is a unique optimal control u uniquely characterized through the variational inequality

$$\pi(u,v) = \lambda(v), \quad \text{all } v \in L^2 \tag{7.1}$$

Next it is shown that if $k_h(t,\cdot) \in L^2$ is a mild solution to the filtering problem (for given $h \in X, t \geq 0$) then $u(\cdot) = -k_h(t,\cdot)$ satisfies (7.1) ; the duality theorem follows from the unique characterization of u through (7.1) and the uniqueness of $k_h(t,\cdot)$ established in §5. ▮

§8. Solution to the Control Problem

For convenience we write $N(\tau) = B(\tau)WB^*(\tau),\ R(\tau) = F(\tau)VF^*(\tau)$. The control problem becomes

$$\begin{cases} \text{Minimize } J_t(u,h) \\ \text{subject to } p(\tau) = T^*(t,\tau)h + \int_\tau^t T^*(s,\tau)\mathcal{C}^*(s)u(s)\,ds \end{cases}$$

(for given $t \geq 0,\ h \in X$), with

$$J_t(u,h) = \int_0^t [\langle p(\tau), N(\tau)p(\tau)\rangle + \langle u(\tau), R(\tau)u(\tau)\rangle]\,d\tau + \langle p, P_0\, p(0)\rangle$$

Let us define the family of maps $\{T_t^{\dagger}(r,s) \mid (r,s) \in \mathcal{P}[0,t]\}$ $(\mathcal{P}[0,t] = \{(r,s) \in \mathbb{R}^2 \mid t \geq r \geq s \geq 0\})$ by

$$T_t^{\dagger}(r,s) = T^*(t-s, t-r)$$

It is a straigthforward matter to show the following (see Vinter and Mitter 1 for details)

Lemma 8.1 Under the assumption that $T^*(\cdot,\cdot)$ is strongly continuous on $\mathcal{P}[0,t]$ then $\{T_t^{\dagger}(r,s) \mid (r,s) \in \mathcal{P}[0,t]\}$ is a mild evolution operator. ▮

We now change variables $\tau \to t - \bar{\tau}$ to give

$$\begin{cases} \text{Minimize } \bar{J}_t(\bar{u},h) \\ \text{subject to } \bar{p}(\bar{\tau}) = T_t^{\dagger}(\bar{\tau},0)h + \int_0^{\bar{\tau}} T_t^{\dagger}(\bar{\tau},s)\bar{\mathcal{C}}^*(s)\bar{u}(s)\,ds \end{cases}$$

with

$$\bar{J}_t(\bar{u},h) = \int_0^t [\langle \bar{p}(\bar{\tau}), \bar{N}(\bar{\tau})\bar{p}(\bar{\tau})\rangle + \langle \bar{u}(\bar{\tau}), \bar{R}(\bar{\tau})\bar{u}(\bar{\tau})\rangle]\,d\bar{\tau} + \langle \bar{p}(t), P_0\, \bar{p}(t)\rangle$$

Here, $p(\tau) = \bar{p}(t-\tau)$ etc.

But this is precisely the control problem studied in (Bensoussan, Delfour and Mitter 1). Indeed by Lemma 8.1, $T_t^{\dagger}(\cdot,\cdot)$ is a mild evolution operator, $\bar{N}(\cdot) \in L^\infty[0,t;\mathcal{L}(X)],\ \bar{R}(\cdot) \in L^\infty[0,t;\mathcal{L}(\mathbb{R}^k)],\ P_0 \in \mathcal{L}(X),\ \mathcal{C}^*(\cdot) \in L^\infty[0,t;\mathcal{L}(\mathbb{R}^k,X)]$; $\bar{N}(t),\ \bar{R}(t)$ are self-adjoint, non-negative for each t and $\bar{R}(t)$ is coercive.

Given $\bar{P} \in C[0,t;\mathcal{L}_s(X)]$, $\mathcal{Y}_{\bar{P}}^{\dagger}(r,s)$ is taken to be the $(-\bar{\mathcal{C}}^*\bar{R}^{-1}\bar{\mathcal{C}}\bar{P})$-perturbed mild evolution operator corresponding to $T_t^{\dagger}(r,s)$ (see §2).

We know* (Bensoussan, Delfour and Mitter 1) that there exists a unique $\bar{P} \in C[0,t;\mathcal{L}_s(X)]$ such that

$$\begin{aligned} \langle \bar{P}(\tau)h,\bar{h}\rangle = {} & \langle \mathcal{Y}_{\bar{P}}^{\dagger}(t,\tau)h, P_0\mathcal{Y}_{\bar{P}}^{\dagger}(t,\tau)\bar{h}\rangle \\ & + \int_\tau^t \langle \mathcal{Y}_{\bar{P}}^{\dagger}(\sigma,\tau)h, [\bar{R}(\sigma) + \bar{P}(\sigma)\bar{\mathcal{C}}^*(\sigma)\bar{R}^{-1}(\sigma)\bar{\mathcal{C}}(\sigma)]\mathcal{Y}_{\bar{P}}^{\dagger}(\sigma,\tau)\bar{h}\rangle\,d\sigma \end{aligned} \tag{8.1}$$

and the unique optimal control u is given by

$$\bar{u}(\tau) = -\bar{R}^{-1}(\tau)\bar{\mathcal{C}}(\tau)\bar{P}(\tau)\mathcal{Y}_{\bar{P}}^{\dagger}(\tau,0)h \tag{8.2}$$

§9. The Kalman Bucy Filter

Given $P \in C[0,t;\mathcal{L}_s(X)]$, let us define $\mathcal{Y}_P(\cdot,\cdot)$ to be the $(-P\mathcal{C}^*R^{-1}\mathcal{C})$-perturbed evolution operator corresponding to $T(\cdot,\cdot)$. We shall require the following technical result, proved in(Vinter and Mitter 1):

Lemma 9.1 For each $t \geq \tau \geq \sigma \geq 0$, $P \in C[0,t;\mathcal{L}_s(X)]$, we have

$$[\mathcal{Y}_{\bar{P}}^{\dagger}(t-\sigma, t-\tau)]^* = \mathcal{Y}_P(\tau,\sigma) \;▮$$

---

* Actually in this reference it is only shown that $\bar{P}$ is weakly continuous. However it is simple matter to exploit the property that the adjoint of $T_t^{\dagger}(\tau,s)$ is strongly continuous to show that $\bar{P}$ is strongly continuous also.

<u>Proposition 9.1</u> (Integral Riccati Equation)

There exists a unique $P(\cdot) \in C[0,t\,;\mathcal{L}_s(X)]$ such that, all $h, \bar{h} \in X$

$$\langle P(\tau)h, \bar{h} \rangle = \langle \mathcal{Y}_p^*(\tau,0)h, P_0 \mathcal{Y}_p^*(\tau,0)\bar{h} \rangle$$

$$+ \int_0^\tau \langle \mathcal{Y}_p^*(\tau,\sigma)h, [B(\sigma)WB^*(\sigma) + P(\sigma)\mathcal{C}^*(\sigma)[\mathcal{F}(\sigma)V\mathcal{F}^*(\sigma)]^{-1}\mathcal{C}(\sigma)P(\sigma)]\mathcal{Y}_p^*(\tau,\sigma)\bar{h} \rangle d\sigma \quad (9.1) ▌$$

<u>Proof</u> It is clear from Lemma (9.1) that $P(\cdot)$ is a solution to (9.1) if and only if $\bar{P}(\cdot) = P(t-(\cdot))$ is a solution to 8.1. Existence and uniqueness of solutions to (9.1) follow from existence and uniqueness of solutions to (8.1).

The field is now set for construction of the dynamical filter:

<u>Theorem 9.1</u> (Kalman-Bucy filter)

Let $P(\cdot) \in C[0,t\,;\mathcal{L}_s(X)]$ be the unique solution to (9.1). Then for each $t \geq 0, h \in X$ the optimal non-linear estimate of $x(t)$ is given by

$$\hat{x}(t,\omega) = \int_0^t \mathcal{Y}_p(t,\tau)P(\tau)\mathcal{C}^*(\tau)[\mathcal{F}(\tau)V\mathcal{F}^*(\tau)]^{-1}\,dz(t,\omega) \quad (9.2)$$

and the process $\hat{x}(t,\omega)$ satisfies

$$\hat{x}(t,\omega) = \int_0^t T(t,\tau)P(\tau)\mathcal{C}^*(\tau)[\mathcal{F}(\tau)V\mathcal{F}^*(\tau)]^{-1}[dz(\tau,\omega) - \mathcal{C}(\tau)\hat{x}(\tau,\omega)\,d\tau] \quad ▌ \quad (9.3)$$

<u>Proof</u> By (8.2) and Theorem (7.1) we have, for each $h \in X$

$$\hat{x}(t,\omega) = \int_0^t \langle k_h(t,s), \cdot \rangle\, dz(s,\omega)$$

is a mild solution to the filtering problem where

$$k_h(t,s) = [\mathcal{F}(s)V\mathcal{F}^*(s)]^{-1}\mathcal{C}(s)\bar{P}(t-s)\mathcal{Y}_{\bar{p}}^{\dagger}(t-\tau,0)h$$

and $\bar{P}(\cdot) \in C[0,t\,;\mathcal{L}_s(X)]$ solves (8.1). We may write

$$\hat{x}_h(t,\omega) = \int_0^t \langle h, \mathcal{Y}_p(t,s)P(s)\mathcal{C}^*(s)[\mathcal{F}(s)V\mathcal{F}^*(s)]^{-1}(\cdot) \rangle\, dz(s,\omega) \quad (9.4)$$

where $P(\cdot)$ is the unique solution to (9.1) in $C[0,t\,;\mathcal{L}_s(X)]$; we have used Lemma 9.1. In view of Proposition 5.1, we conclude (9.2). (See Vinter and Mitter 1 for details). Finally, (9.3) is deduced from (9.4) by using the defining property for perturbed evolution operators :

$$\mathcal{Y}_p(t,s)\bar{h} = T(t,s)\bar{h} - \int_s^t T(t,\sigma)P(\sigma)\mathcal{C}^*(\sigma)[\mathcal{F}(\sigma)V\mathcal{F}^*(\sigma)]^{-1}\mathcal{C}(\sigma)\mathcal{Y}_p(\sigma,s)\bar{h}\,d\sigma$$

for each $\bar{h} \in X$.

Again we refer to (Vinter and Mitter 1). ▌

§10. Filter Stability

The Generalized Kalman filter for the filtering problem of §4 is taken to be the map $\mathcal{K}$ from $\mathbb{R}^k$-valued measurable processes $\{z(t,\omega) \mid t \geq 0\}$ into X-valued measurable processes $\{\hat{x}(t,\omega) \mid t \geq 0\}$ defined by

$$\hat{x}(t,\omega) = \int_0^t \mathcal{Y}_p(t,s)K(s)\,dz(s,\omega) \qquad \text{each } t \geq 0 \quad (10.1)$$

$(K(s) = P(s)\mathcal{C}^*(s)[\mathcal{F}(s)V\mathcal{F}^*(s)]^{-1})$, $\mathcal{Y}_p$, P as in §9. The domain of $\mathcal{K}$ comprises all processes $z(t,\omega)$ for which the R.H.S. of (10.1) is defined (for each t) and corresponding to which a measurable version of $\int_0^t \mathcal{Y}_p(t,s)K(s)\,dz(s,\omega)$

exists.

We note in particular that the process $z^*(t,\omega)$ lies in the domain of $\mathcal{K}$ where

$$\left.\begin{aligned} x^*(t,\omega) &= T(t,0)x_0^*(\omega) + \int_0^t T(t,s)B(s)\,dW(s,\omega) \\ z^*(t,\omega) &= \int_0^t \mathcal{C}(\tau)x^*(\tau,\omega)\,d\tau + \int_0^t \mathcal{F}(\tau)\,dv(\tau,\omega) \end{aligned}\right\} \quad (10.2)$$

(10.2) is identical to (3.1) <u>except</u> that $x_0^*(\omega)$ is now allowed to be an arbitrary X-valued random variable independent of

Of course the process $z^*(t,\omega)$ is interpreted as

$$z^*(t,\omega) = \int_0^t \mathcal{C}(\tau)T(\tau,0)x_0^*(\omega)\,d\tau + \int_0^t \mathcal{C}(\tau)\int_0^\tau T(\tau,\sigma)B(\sigma)\,dW(\sigma,\omega)\,d\tau + \int_0^t \mathcal{F}(\tau)\,dv(t,\omega)$$

Because $x_0^*(\omega)$ is no longer assumed Gaussian, $\hat{x}^*(t,\omega)$ given by

$$\hat{x}^*(t,\omega) = \int_0^t \mathcal{Y}_p(t,s)K(s)\,dz^*(s,\omega)$$

need not be gaussian or even second order.

Now a highly desirable property of the filter is that the error process $\hat{x}(t,\omega) - x(t,\omega)$ be insensitive to errors in the modelling of $x_0(\omega)$. We make this precise in the following definition :

<u>Definition 10.1</u> The filter $\mathcal{K}$ is <u>stable</u> if there exists some measure μ_∞ on the Borel sets of X, such that writing μ_t^* for the measure induced on the Borel sets of X by the X-valued r.v. $\hat{x}^*(t,\omega) - x^*(t,\omega)$ we have

$$\mu_t^* \to \mu_\infty \text{ (weakly)}$$

for x_0^* an <u>arbitrary</u> X-valued r.v. ▌

Here by weak convergence of μ_t^* we mean the following: for every continuous function $f: X \to \mathbb{R}$ taking values in some bounded set we have

$$\int_X f(x)\,d\mu_t^*(x) \to \int_X f(x)\,d\mu_\infty(x) \quad , \quad t \to \infty$$

(Gikhman and Skorokhod 1).

Let us first take notice of the following representation of the "error" process

<u>Proposition 10.1</u> Let the processes $x^*(t,\omega), z^*(t,\omega)$ be as defined in (10.2). Take

$$\hat{x}^*(t,\omega) = \int_0^t \mathcal{Y}_p(t,s)K(s)\,dz^*(s,\omega)$$

$(K(s) = P(s)\mathcal{C}^*(s)[\mathcal{F}(s)V\mathcal{F}^*(s)]^{-1})$. Then

$$\hat{x}^*(t,\omega) - x^*(t,\omega) =$$

$$\mathcal{Y}_p(t,0)x_0^*(\omega) + \int_0^t \mathcal{Y}_p(t,\sigma)[B(\sigma)\,dW(\sigma,\omega) + K(\sigma)\mathcal{F}(\sigma)\,dv(\sigma,\omega)] \quad (10.3)$$

(Here $P(\cdot)$ and $\mathcal{Y}_p(\cdot,\cdot)$ are taken as in Theorem 9.1) ▌

<u>Proof</u> See (Vinter and Mitter 1) ▌

The following corollary, interpreting $P(t)$ as the "error covariance" in the case when $x_0^* = x_0$ is almost immediate from the results in §2 :

<u>Corollary 10.1</u> Take the processes $\hat{x}(t,\omega)$, $x(t,\omega)$ as in Theorem 9.1. Then

$$\operatorname{cov}\{\hat{x}(t,\omega) - x(t,\omega)\} = P(t) \qquad \text{each } t \geq 0 \;▌$$

It would appear that minimal assumptions ensuring filter stability

are the following
Assumption 10.1
(i) $\| \mathcal{Y}_p(t,0) \| \to 0$ as $t \to \infty$
(ii) There exists some nuclear operator $P_\infty \in \mathcal{L}(X)$ such that

$$\lim_{t\to\infty} \int_0^t \langle \mathcal{Y}_p^*(t,\sigma)h, [B(\sigma)WB^*(\sigma) + K(\sigma)\mathcal{C}(\sigma)P(\sigma)]\mathcal{Y}_p^*(t,\sigma)\bar{h} \rangle d\sigma = \langle P_\infty h, \bar{h} \rangle$$

for each $h, \bar{h} \in X$. ▮

Assumption 10.1 is of an unsatisfactory ad hoc nature; the following lemma supplies sufficient conditions for Assumption 10.1 to be met which, although excessively strong, are at least open to direct verification. The basic idea is that if the process $x(t,\omega)$ is exponentially stable and the maps defining the process and observations are sufficiently "small" then $\mathcal{Y}_p(t,0)$ is exponentially stable.

Lemma 10.1 Suppose that there exist $\omega_0 > 0, M \geq$ such that $\|T(t,\tau)\| < Me^{-\omega_0(t-\tau)}$ and that

$$\{ \|P_0\| + \sup_{t\geq 0} [B(t)WB^*(t)] \} \sup_{t \geq 0} \{ \| \mathcal{C}^*(t)[F(t)VF^*(t)]^{-1}\mathcal{C}(t) \| \} < \omega_0 / M^3$$

then there exists some $\varepsilon_0 > 0$ such that

$$\| \mathcal{Y}_p^*(t,\sigma) \| < Me^{-\varepsilon_0(t-\sigma)} \tag{11.1}$$

and Assumption 10.1 is satisfied. ▮

Proof See (Vinter and Mitter 1) . We note part (i) of Assumption 10.1 is immediate from 11.1. Part (ii) follows essentially from properties (ii) and (iv) of the Wiener Integral given in §§2.4 and the completeness of the trace class operators* in the trace norm. ▮

Proposition 10.2 Suppose that Assumption 10.1 holds (for example, if the conditions of Lemma 10.1 are met). Then the filter $\mathcal{K}$ above is stable and the limiting measure μ_∞ is gaussian with zero mean and covariance operator P_∞ . We have in addition
(i) if $x_0^*(\omega)$ is first order, then mean $\{\mu_t^*\} \to$ mean $\{\mu_\infty\}$ (strongly)
(ii) if $x_0^*(\omega)$ is second order, then

$$\operatorname{cov}\{\mu_t^*\} \to \operatorname{cov}\{\mu_\infty\} \quad \text{(in the weak } \mathcal{L}(X) \text{ topology)} ▮$$

Proof This makes use of results in (Grenander 1) and patterns arguments in (Vinter). ▮

What does not appear obtainable at the present level of generality are conditions for filter stability in terms of (appropriately defined) "controllability" and "detectability" of the system under consideration.

We are able to conclude boundedness of the error covariance however from the appropriate detectability assumption; this is significant because, in applications to delay systems, the detectability hypothesis is directly verifiable.

For simplicity we specialize to time invariant systems.

Definition 10.2 Suppose that (3.1) and (3.2) are time-invariant, i.e. $T(t,\tau) = T(t-\tau)$, $B(t) = B$ etc. Then the system (3.1) - (3.2) is detectable if there exists some bounded linear map $D : X \to \mathbb{R}^k$ such that the semigroup $T_{A^*+\mathcal{C}^*D}(t)$ generated by $A^*+\mathcal{C}^*D$ (A the generator of $T(\cdot)$) is L^2-stable. ▮

Proposition 10.3 Suppose that the system (3.1)-(3.2) is time invariant and detectable. Let $\hat{x}(t,\omega)$ be as given in Theorem 9.1. Write μ_t for the measure induced on the Borel sets of X by $\hat{x}(t,\omega) - x(t,\omega)$ and write

*See (Gelfand and Vilenkin 1)

for $\operatorname{cov}\{\hat{x}(t,\omega) - x(t,\omega)\}$. Then
(i) $\{P(t)\}_{t\geq 0}$ is bounded (in trace class) and $\{\mu_t\}_{t \geq 0}$ is weakly compact*
(ii) when additionally we assume $P_0 = 0$, there exists some gaussian measure μ_∞ with zero mean and covariance P_∞ where P_∞ satisfies

$$\langle P_\infty A^* h, \bar{h} \rangle + \langle h, P_\infty A^* \bar{h} \rangle + \langle BWB^* h, \bar{h} \rangle - \langle P_\infty \mathcal{C}^* [FVF^*]^{-1} \mathcal{C} P_\infty h, \bar{h} \rangle = 0$$

for all $h, \bar{h} \in \mathcal{D}\{A^*\}$ such that

$$\mu_t \longrightarrow \mu_\infty \qquad \text{(weakly) and}$$
$$P(t) \to P_\infty \qquad \text{(in trace class).} ▮$$

Proof See (Vinter and Mitter 1). ▮

Notice that Proposition 10.3 makes no assertion about filter stability in the sense of Definition 10.1. Indeed the conclusions only apply for $x_0^*(\omega)$ the initial condition for which the filter was constructed.

§11. Specialization to Stochastic Delay Systems

We now turn attention to the stochastic delay system :

$$\begin{cases} dx(t,\omega) = \mathcal{L}x_t(\omega)dt + B\,dw(t,\omega) & \text{(process)} \\ x(0,\omega) = [h(\omega)](0) & t \geq 0 \end{cases} \tag{11.1}$$

$$\begin{cases} dz(t,\omega) = \mathcal{C}x_t(\omega)\,dt + F\,dv(t,\omega) & \text{(observations)} \\ z(0,\omega) = 0 \end{cases} \tag{11.2}$$

Here

$$\mathcal{L}x_t = \sum_{i=0}^{m} A_i \begin{cases} x(t+\theta_i) & t+\theta_i \geq 0 \\ h(t+\theta_i) & t+\theta_i < 0 \end{cases} + \int_{-b}^{0} A(\theta) \begin{cases} x(t+\theta) & t+\theta \geq 0 \\ h(t+\theta) & t+\theta < 0 \end{cases} d\theta$$

and

$$\mathcal{C}x_t = C_0 x(t) + \int_{-b}^{0} C(\theta) \begin{cases} x(t+\theta) & t+\theta \geq 0 \\ h(t+\theta) & t+\theta < 0 \end{cases} d\theta$$

We suppose
$\mathcal{U}, X$ real, separable, Hilbert spaces
$-b < \theta_m < \theta_{m-1} < \ldots < \theta_0 = 0$; $A_0, A_1, \ldots, A_m \in \mathcal{L}(X)$
$A \in L^\infty[-b,0;\mathcal{L}(X)]$, $B \in \mathcal{L}(\mathcal{U},X)$, $C_0 \in \mathcal{L}(X;\mathbb{R}^k)$
$C \in L^\infty[-b,0;\mathcal{L}(X,\mathbb{R}^k)]$, $F \in \mathcal{L}(\mathbb{R}^k)$

$w(t,\omega)$ is a measurable, almost surely strongly sample continuous $\mathcal{U}$-valued Wiener process with nuclear covariance W . $v(t,\omega)$ is a measurable, almost surely strongly sample continuous $\mathbb{R}^k$-valued Wiener process with covariance $V \in \mathcal{L}(\mathbb{R}^k, \mathbb{R}^k)$. $h(\omega)$ is an M^2-valued** gaussian r.v.

* (Gikhman and Skorokhod 1)
** M^2 denotes the Hilbert space $L^2[-b,0;X] \times X$. Elements in M^2 are written $(h(\theta)\ -b \leq \theta < 0 ; h(0))$

with zero mean and covariance P_0 .

It is assumed that $dw(t,\omega), dv(t,\omega), h(\omega)$ are independent and that $FVF^* > 0$.

The process (11.1) is interpreted as follows : the X-valued stochastic process $\{x(t,\omega) \mid t \geq 0\}$ is a solution to (11.1) if (a) it is almost surely (strongly) sample continuous, and (b) for each $t \geq 0$,

$$x(t,\omega) = [h(\omega)](0) + \int_0^t [\mathcal{L} x_s(\omega)]\, ds + B W(t,\omega)$$

A standard Picard iteration argument gives the solution as unique, if it exists (c.f. Lindquist 1)

Now we take the stochastic abstract evolution equation corresponding to (11.1) to be

$$\tilde{x}(t,\omega) = T(t) h(\omega) + \int_0^t T(t-s) \tilde{B}\, dW(s,\omega) \qquad (11.3)$$

where $\tilde{X} = M^2$, $\{T(t) \in \mathcal{L}(\tilde{X}) \mid t \geq 0\}$ is the semigroup describing evolution of "trajectory segments" of the homogeneous equation (Delfour and Mitter 1) and $\tilde{B} : \mathcal{U} \to \tilde{X}$ is defined as

$$[\tilde{B} u](\theta) = \begin{cases} Bu & \theta = 0 \\ 0 & \theta < 0 \end{cases} \qquad (\, B \text{ as above})$$

We take the abstract observation process $\tilde{z}(t,\omega)$ to be

$$\tilde{z}(t,\omega) = \mathcal{C} \tilde{x}(t,\omega) + F dv(t,\omega)$$

Now for the theory developed in the previous sections to be applicable here, we require that (11.3) and (11.1) be equivalent in an appropriate sense. The necessary result is embodied in the following proposition :

Proposition 11.1 There exists an almost surely (strongly) sample continuous version of (11.3) which we also write as $\tilde{x}(t,\omega)$. Defining the projection $H : \tilde{X} \to X$ by $Hh = h(0)$ we have

$$x(t,\omega) \triangleq H \tilde{x}(t,\omega) \quad \text{all } t \geq 0,\ \text{each } \omega \in \Omega$$

is the unique solution of (11.1) (to within uniform stochastic equivalence) and for each $t \geq 0$

$$[\tilde{x}(t,\omega)](\theta) = \begin{matrix} x(t+\theta,\omega) & t+\theta \geq 0 \\ [h(\omega)](t+\theta) & t+\theta < 0 \end{matrix}$$

for almost all $\theta \in [-b, 0]$, w.p. 1. ▮

Proof The somewhat lenghtly proof in (Vinter 1) is not reproduced here. The result is crucial in the present development however and the essential steps are therefore outlined.

(a) Let $\tilde{A}$ be the infinitesimal generator of the C_0 semi-group $T(t)$ above. We first need to show that $q(t,\omega) = \int_0^t T(t-s)\tilde{B}\, dW(s,\omega)$ (and thence $\tilde{x}(t,\omega) = T(t)h(\omega) + \int_0^t T(t-s)\tilde{B}\, dW(s,\omega)$) has an almost surely (strongly) sample continuous version. A basic difficultly* is that range$\{\tilde{B}\} \not\subset \mathcal{D}\{\tilde{A}\}$, and

* This difficulty seems to preclude development of a satisfactory general theory of stochastic processes described by $dx(t,\omega) = Ax(t,\omega)dt + B dW(t,\omega)$, A the generator of a C_0 semigroup and B a general bounded linear map. The difficulty can be circumvented by relaxing the underlying probability measure to be merely finitely additive and defining a "Wiener process" with absolutely continuous sample paths (Balakrishnan 1).

it does not appear that for $T(t)$ a general C_0 semigroup, $\tilde{B}$ a general bounded linear map, a sample continuous version of $q(t,\omega)$ necessarily exists. However, exploiting the very special structure of $\tilde{B}$ and $T(t)$ under consideration, we can show that $q(t,\omega)$ satisfies some generalized Kolmogorov condition and then make use of a result in (Bensoussan 1 , p. 175) to establish the desired sample continuity.

(b) We next show that the (almost surely) strongly sample continuous version of $\tilde{x}(t,\omega)$ uniquely satisfies

$$\langle \tilde{x}(t,\omega), y \rangle = \langle x_0(\omega), y \rangle + \int_0^t \langle \tilde{A}^* y, x(s,\omega) \rangle\, ds + \langle y, \tilde{B} W(t,\omega) \rangle \quad \text{w.p. 1} \qquad (11.4)$$

for each $y \in \mathcal{D}\{\tilde{A}^*\}$, each $t \geq 0$

within the class of (almost surely) weakly sample continuous processes.

(c) Finally by consideration of the detailed structure of $\tilde{A}^*$ given in (Vinter 4), we show that all sample functions $\tilde{x}(t,\omega)$ which are strongly continuous and satisfy (11.4) for all $t \geq 0$ (such sample functions have full measure) have representation as "segments" of some strongly continuous function $\varphi : [-b, \infty) \to X$ with $\varphi(\theta) = h(\theta)$ for $-b \leq \theta \leq 0$ which satisfies (11.1).

It is clear from Proposition 11.1 that the " X-part" of $\tilde{x}(t,\omega)$ is precisely the solution of (11.1), the " L^2-part" of $\tilde{x}(t,\omega)$ describes evolution of "trajectory segments" of the stochastic delay equation (11.1) and that the abstract observation process $\tilde{z}(t,\omega)$ is precisely $z(t,\omega)$. ▮

Comment $\mathcal{C}$ defines a bounded linear operator $M^2 \to \mathbb{R}^k$. Note that in the present framework we cannot treat point delays in the observations, i.e.

$$\mathcal{C} x_t = \sum_{i=0}^{m} C_i\, x(t+\theta_i) + \int_{-b}^0 C(\theta) x(t+\theta)\, d\theta$$

for in this case $\mathcal{C}$ becomes an unbounded operator and the theory above is no longer applicable. Of course point delays in the observations can be approximated by "distributed" delays in an obvious manner. ▮

Now let H be the projection operator introduced in Proposition 11.1. Noting that assumptions (3.1) and (3.2) are met we have from previous sections :

Proposition 11.2 There exists a unique $P(\cdot) \in C[0,t\, ; \mathcal{L}_s(\tilde{X})]$ such that

$$\langle P(\tau)h, \bar{h} \rangle = \langle \mathcal{Y}_p^*(\tau,0)h, P_0 \mathcal{Y}_p^*(\tau,0)\bar{h} \rangle$$
$$+ \int_0^\tau \langle \mathcal{Y}_p^*(\tau,\sigma)h, [\tilde{B}W\tilde{B}^* + P(\sigma)\mathcal{C}^*[FVF^*]^{-1}\mathcal{C}P(\sigma)]\, \mathcal{Y}_p^*(\tau,\sigma)\bar{h} \rangle\, d\sigma \qquad (11.5)$$

and the optimal non-linear estimate $\hat{x}(t,\omega)$ of $x(t,\omega)$ is $H\hat{\tilde{x}}(t,\omega)$ where $\hat{\tilde{x}}(t,\omega)$ is given by

$$\hat{\tilde{x}}(t,\omega) = \int_0^t \mathcal{Y}_p(t,s) P(s) \mathcal{C}^* [FVF^*]^{-1} dz(s,\omega)$$

(the evolution operator $\mathcal{Y}_p(t,s)$ interpreted as in §9) and satisfies

$$\hat{\tilde{x}}(t,\omega) = \int_0^t T(t-\tau) P(\tau) \mathcal{C}^* [FVF^*]^{-1} [dz(\tau,\omega) - \mathcal{C}\hat{\tilde{x}}(\tau,\omega) d\tau] \quad ▮$$

We can exploit the fact that $T(\cdot)$ has generator $\tilde{A}$ to further characterize the covariance operator $P(t)$.

Proposition 11.3 Let $P(t) \in C[0,t\, ; \mathcal{L}_s(X)]$ be the unique solution to (11.5) Then $\langle P(t)h, \bar{h} \rangle$ is absolutely continuous and

$$\begin{cases} d/dt \langle P(t)h,\bar{h}\rangle = \langle P(t)h, A^*\bar{h}\rangle + \langle A^*h, P(t)h\rangle \\ \qquad\qquad + \langle h, BWB^*\bar{h}\rangle - \langle h, P\mathscr{C}^*[FVF^*]^{-1}\mathscr{C}P\bar{h}\rangle \\ P(t_0) = P_0 \end{cases} \tag{11.6}$$

for every $h,\bar{h} \in \mathcal{D}\{A^*\}$ ∎

Proposition 11.4 Suppose that (i) $C_0 = 0$ (ii) $C(\cdot) \in W^{1,2}$ (i.e. $\mathscr{C}x_t = \int_{-b}^{0} C(\theta)x(t+\theta)d\theta$ with $C(\cdot)$ sufficiently smooth) then $P(t)$ is the unique solution to (11.6) within the class of maps $\varphi \in C[0,\infty;\mathcal{L}_w(\tilde{X})]$ such that $\langle \varphi h,\bar{h}\rangle$ is absolutely continuous for each $h,\bar{h} \in \mathcal{D}\{A^*\}$. ∎

Proposition 11.3 is proved in (Vinter 2). Proposition 11.4 also follows from results in (Vinter 2) noting that the additional hypotheses on $\mathscr{C}$ ensure that $\text{range}\{\mathscr{C}^*\} \subset \mathcal{D}\{A^*\}$ (see Vinter 4).∎

We now turn to specialization of the filter stability results of §10.

Even with the additional structure here present it does not appear possible to obtain satisfactory stability conditions in terms of suitably defined controllability and detectability of appropriate operator pairs.

We remark though that some results are now available characterizing the growth of $T(t)$ in terms of the maps $A_0, A_1, ..$ so that the hypotheses of Lemma 11.1 may be verified. We can therefore generate at least a class of non-trivial delay systems to which correspond stable optimal filters.

It is possible however to give a fairly simple characterization of detectable delay systems in the sense of Definition 11.2 and hence (in view of Proposition 10.3) of systems for which the error covariance is bounded (in trace norm). First take note of :

Lemma 11.1 Take $\tilde{A}$ the generator of $T(t)$. Then $\tilde{A}^*$ has compact resolvent and the generalized eigenmanifold $\mathfrak{M}^{\oplus}$ corresponding to the spectral set comprising all poles of $\tilde{A}^*$ with non-negative real parts is finite dimensional. ∎

Proof This follows simply from the property that $\tilde{A}$ has compact resolvent implies $\tilde{A}^*$ has compact resolvent and the known spectral properties of $\tilde{A}$. (See Vinter 4). ∎

Using notions in (Hale 1 , p. 98 et seq.) concerning decomposition of the state space of time-invariant delay systems into generalized eigenmanifolds of the generator it is not difficult to show

Proposition 11.5 Let $\mathfrak{M}^{\oplus}$ be as in Lemma 11.1. Then the delay system (11.1), (11.2) is detectable if and only if

$$\mathfrak{M}^{\oplus} \subset \text{range}\{\mathscr{C}^*\} \quad ∎$$

§12. A comment on time varying stochastic delay systems

Attention has been restricted to time-invariant delay systems in §11. Now consider the time-varying situation. Write $T(t,s)$ for the evolution operator describing evolution of trajectory segments of the homogeneous time-varying delay system and write $\tilde{A}(t)$ for its generator (we make suitable assumptions on $A_i(t)$, $A(t,\theta)$ as time functions, see Vinter 4). The following serious difficulties arise :

(i) it is not clear in this case that $T^*(\cdot,\cdot)$ is strongly continuous, a crucial hypothesis in development of the filter.

(ii) it is not apparent how we would prove "equivalence" of $[\tilde{x}(t,\omega)](0)$ and $x(t,\omega)$ in this situation; without this crucial step the whole development is of questionable value.

(iii) there is no hope in general of the error covariance $P(t)$ being uniquely characterized through a differential Riccati equation.

Difficilties (ii) and (iii) arise in connection with the following awkward property of the adjoint $\tilde{A}^*(t)$ of the generator $\tilde{A}(t)$: in general*

$$\bigcap_{t\geq 0} \mathcal{D}\{\tilde{A}^*(t)\} \quad \text{is not dense in } \tilde{X}$$

See (Vinter 4).

* Indeed for the scalar system $dx/dt = t\,x(t-1)$,

$$\bigcap_{t\geq 0} \mathcal{D}\{A^*(t)\} \subset \{h \in M^2 \mid h(0) = 0\}$$

References

A.V. Balakrishnan 1 , to appear

A.T. Bharucha-Reid 1 , "Random Integral Equations"
Academic Press, New York, 1972

A. Bensoussan 1 , "Filtrage Optimal des Systèmes Lineaires"
Dunod, Paris, 1971

A. Bensoussan, M.C. Delfour and S.K. Mitter 1 , "Notes on Infinite Dimensional Systems"
Monograph (to appear)

A. Bensoussan and M. Viot , "Optimal Control of Stochastic Linear Distributed Parameter Systems"
IRIA Technical Report, 1974

R.T. Curtain 1 , "Stochastic Differential Equations in a Hilbert Space"
J. Diff. Eqns., 10, 1971, pp. 412-431

R.T. Curtain 2 , "Infinite Dimensional Filtering"
SIAM J. Control (to appear)

M.C. Delfour , "State Theory of Linear Hereditary Differential Systems"
to appear

M.C. Delfour and S.K. Mitter 1 , "Controllability, Observability and Optimal Feedback Control of Hereditary Differential Systems"
SIAM J. Control, 10, 1972, pp. 298-328

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, vol. 13, 1, 1975 (to appear)

I.M. Gelfand and N.Y. Vilenkin 1 , "Generalized Functions" vol. 4
Academic Press, New York, 1964

I.I. Gikhman and A.V. Skorokhod 1 , "Introduction to the Theory of Random Processes"
Saunders, London, 1969

U. Grenander 1 , "Probabilities on Algebraic Structures"
Wiley, New York, 1966

J. Hale 1 , "Functional Differential Equations"
Springer Verlag, New York, 1971

H. Kwakernaak 1 , "Optimal Filtering in Linear Systems with Time Delay"
IEEE Trans. Aut. Con., 12,2,1967, pp. 169-173

A. Lindquist 1 , "A Theorem on Duality between Estimation and Control for Linear Stochastic Systems with Time Delay"
J. Math. Analysis and Appl., 37,2,1972 pp. 516-536

A. Lindquist 2 , "Optimal Control of Linear Stochastic Systems with Applications to Time Lag Systems"
Information Sciences, 5,1973 pp. 81-126

A. Lindquist , "On Feedback Control of Linear Stochastic Systems"
SIAM J. of Control, vol. 11,2,May 1973 pp. 323-343

J.L. Lions 1 , "Control Optimal de Systems Gouvernés par des Equations aux Dérivées Partielles"
Dunod, Paris, 1968

R.B. Vinter 1 , "Stochastic Delay Equations Formulated as Stochastic Evolution Equations" MIT Technical Report 1974

R.B. Vinter 2 , "Some Results Concerning Perturbed Evolution Equations with Applications to Delay Systems"
Technical Report Electronic Systems Laboratory, M.I.T.1974

R.B. Vinter 3 , "Invariant Measures Induced by Stochastic Evolution Equations"
Technical Report Electronic Systems Laboratory, MIT 1974

R.B. Vinter 4 , "On the Evolution of the State of Linear Differential Delay Equations in M^2 : Properties of the Generator"
Technical Report Electronic Systems Laboratory, MIT 1974

R.B. Vinter and S.K. Mitter 1 , "Filtering of Stochastic Evolution Equations"
Technical Report Electronic Systems Laboratory, MIT,1974

W.M. Wonham , "Random Differential Equations in Control Theory" in "Probabilistic Methods in Applied Mathematics" Editor Bharucha-Reid, Academic Press, New York, 1970